CALDER

URBAN INSTITUTE

NATIONAL CENTER *for* ANALYSIS *of* LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the Urban Institute with Duke University, Stanford University, University of Florida, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

Value Added of Teachers in High-Poverty Schools and Lower-Poverty Schools

TIM SASS, JANE HANNAWAY, ZEYU XU, DAVID FIGLIO, AND LI FENG

WORKING PAPER 52 • NOVEMBER 2010 (UPDATED VERSION OF WORKING PAPER 41)

Value Added of Teachers in High-Poverty Schools and Lower-Poverty Schools

Tim R. Sass
Florida State University

Jane Hannaway
The Urban Institute

Zeyu Xu
The Urban Institute

David N. Figlio
Northwestern University

Li Feng
Texas State University

Contents

Acknowledgements

This research was supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER) funded through Grant R305A060018 to the Urban Institute from the Institute of Education Sciences, U.S. Department of Education. The authors also gratefully acknowledge support from the U. S. Department of Education, Contract No. ED-04-CO-0036/0002 to RTI, UI Subcontract No. 15-312-0209234. This paper greatly benefitted from comments on an earlier draft made by U. S. Department of Education staff, including Andy Abrams, David Goodwin and Stephanie Stullich.

CALDER • The Urban Institute
2100 M Street N.W., Washington, D.C. 20037
202-261-5739 • www.caldercenter.org

Value Added of Teachers in High–Poverty Schools and Lower–Poverty Schools

Tim R. Sass, Jane Hannaway, Zeyu Xu, David N. Figlio, and Li Feng

CALDER Working Paper No. 52
November 2010

Abstract

This paper examines whether teachers in schools serving students from high-poverty backgrounds are as effective as teachers in schools with more advantaged students. The question is important. Teachers are recognized as the most important school factor affecting student achievement, and the achievement gap between disadvantaged students and their better off peers is large and persistent.

Using student-level microdata from 2000-2001 to 2004-2005 from Florida and North Carolina, the authors compare the effectiveness of teachers in high-poverty elementary schools (>70% FRL students) with that of teachers in lower-poverty elementary schools (<70% FRL students). The results show that the average effectiveness of teachers in high-poverty schools is in general less than teachers in other schools, but only slightly, and not in all comparisons. The authors also find differences in within-school-type variation in teacher effectiveness in nearly every comparison. These differences are largely driven by the longer tail at the bottom of the teacher effectiveness distribution in high-poverty schools. Teachers at the top of the effectiveness distribution are very similar across school settings.

The observed differences in teacher quality between high-poverty and lower-poverty schools are not due to differences in the observed characteristics of teachers, such as experience, certification status and educational attainment. Rather, they appear to arise from differences in the marginal return or payoff from increases in a characteristic. In particular, the gain in productivity from increased experience is much stronger in lower-poverty schools. The lower return to experience in high-poverty schools does not appear to be a result of differences in the quality of teachers who leave teaching or who switch schools. Rather, it may be the case that the effect of experience on teacher productivity may depend on the setting in which the experience is acquired. If there are positive spillovers among teachers that depend on teacher quality (ie. teacher "peer effects") or if exposure to challenging student populations lessens the future productivity of teachers (i.e. leads to "burn out"), teachers in schools serving large proportions of low-income students may simply not improve much as time goes by.

These findings suggest that solutions to the achievement gap between high and lower-poverty schools may be complex. Changing the quality of new recruits or importing teachers with good credentials into high-poverty schools may not be sufficient. Rather, the findings suggest that measures that induce highly effective teachers to move to high-poverty schools and which promote an environment in which teachers' skills will improve over time are more likely to be successful.

Introduction

Whether measured by student achievement or educational attainment, the divergence in student performance between schools serving primarily low-income populations and those enrolling students from more affluent families is stark. In 2009, only 14 percent of 4th grade students from high-poverty schools scored at or above the "proficient" level in reading on the National Assessment of Educational Progress (NAEP), whereas half the students in low-poverty schools met or exceeded the threshold for proficiency. Similarly, in math just 17 percent of students from high-poverty schools scored at the proficient level or above on the NAEP while 60 percent of students from low-poverty schools performed at the proficient level or better.

The differences are even larger for students in large cities, where the proportion of students who are eligible for free and reduced-price lunch (FRL) who earn a proficient score in either math or reading is lower than for FRL students in the rest of the nation (Uzzell et al. 2010). Twelfth-grade students in high-poverty schools are also less likely to earn a high school diploma than their counterparts from low-poverty schools (68 versus 91 percent) and are less likely to attend a four-year college (28 percent versus 52 percent) (Aud et al. 2010). Correspondingly, nearly 90 percent of so-called "dropout factories," where fewer than 60 percent of high-school freshman are still attending the same school in grade 12, are schools serving large numbers of low-income students. The majority of these schools with low graduation rates are located in the nation's cites (Balfanz and Legters 2005). In half of the 100 largest cities in the United States 50 percent or more of high school students attend "dropout factories" (Balfanz and Legters 2004).

While student, parental and neighborhood factors undoubtedly contribute to the observed performance differential between students in low and high-poverty schools, it is hard to deny that systematic differences in school quality are partly to blame for the observed gaps in achievement and educational attainment. The source of quality differentials between schools serving primarily low-income students and those serving more affluent students, and hence the appropriate policies to ameliorate differences in school quality, are less clear, however.

Differences in teacher quality would appear to be the most likely reason for disparities in the quality of high-poverty and lower-poverty schools. Recent empirical evidence finds teachers to be the most important schooling factor affecting student achievement. Rockoff (2004), Rivkin, Hanushek, and Kain (2005) and Aaronson, Barrow, and Sander (2007) show that persistent measures of teachers' contributions to student achievement or "value added" vary tremendously across teachers, and that the within-school variation in value added is at least as large as between-school variation. Similarly, Figlio and Lucas (2004) demonstrate that within-school variation in grading standards (and presumably other teacher behaviors as well) is nearly the same as the overall variation in grading standards.

Previous research has also highlighted disparities in the qualifications of teachers in schools serving primarily disadvantaged and minority students versus teachers in schools with more advantaged student bodies (Clotfelter, Ladd, and Vigdor 2005; Goldhaber, Choi, and Cramer 2007; Lankford, Loeb, and Wyckoff 2002). However, while observed teacher characteristics (e.g. educational attainment, certification status, years of experience beyond the first few years, etc.) vary across schools, these differences are only weakly related to teacher performance (Harris and Sass 2007; Clotfelter, Ladd, and Vigdor 2007).

Several recent studies also clearly establish that schools serving disadvantaged students have more difficulty hiring and retaining teachers. Teachers in general appear to prefer schools that serve students from higher income families and who are higher-achieving, and teachers working in schools with more highly disadvantaged students are more likely to leave their school district or transfer to a lower-needs school within their district (Lankford, Loeb, and Wyckoff 2002; Hanushek, Kain, and Rivkin 2004; Boyd, Lankford, Loeb, and Wyckoff 2005; Clotfelter, Ladd, and Vigdor 2005; Imazeki 2005; Scafidi, Sjoquist, and Stinebrickner 2007; Feng 2009). Accountability pressures can exacerbate the problems that schools serving low-achieving student populations face in retaining high-quality teachers (Feng, Figlio, and Sass 2010). Teachers with better earning opportunities are also more likely to leave teaching (Dolton and van der Klaauw 1999), which might make it even harder for schools in economically disadvantaged area to hold on to good teachers. It may also be harder for these schools to recruit teachers to begin with, as potential teachers tend to prefer to work in schools near where they grew up (Boyd et al. 2005).

The combination of evidence on the importance of teacher quality, differences in observable qualifications of teachers across schools, and the mobility patterns of teachers has led many observers to conclude that the quality of teachers in high-poverty schools is generally inferior to that of teachers in lower-poverty schools. This view has fueled policy initiatives designed to encourage promising new teachers to teach in high-poverty schools ("Teach for America") and provide incentives for existing teachers to move from lower-poverty schools to high-poverty schools. For example, the recent "Race-to-the-Top" competition graded applicants in part on whether they had plans to provide financial incentives to teach in high-poverty schools.[1] Similarly, the U.S. Department of Education has funded a set of experiments in seven school districts throughout the country, known as the "Talent Transfer Initiative" that provide differential compensation to highly effective teachers who agree to teach in high-poverty schools.

Despite the circumstantial evidence, in fact little is known about the relative productivity of teachers in schools serving economically disadvantaged student populations and those enrolling students from more affluent families. In this paper we seek to fill this void and inform the debate on teacher labor market policies by addressing four related research questions:

1. How does the average contribution of teachers in high-poverty elementary schools compare to that of teachers in lower-poverty elementary schools in terms of student achievement gains in mathematics and in reading/language arts?
2. Are there differences in the variation of teacher effectiveness within schools serving largely students from low-income families vis-à-vis the set of schools serving more affluent populations?
3. To what extent do observed teacher characteristics (e.g., certification, experience, education) in high-poverty elementary schools and in schools with lower-poverty levels explain differences in teacher contributions to student learning?

[1] The federal Race to the Top guidelines required states to include plans for ensuring an "equitable distribution of effective teachers and principals." Applicants were graded in part on "The extent to which the State, in collaboration with its participating LEAs … has a high-quality plan … to ensure that students in high-poverty and/or high minority schools … have equitable access to highly effective teachers … Plans … may include, but are not limited to, the implementation of incentives and strategies in such areas as recruitment [and] compensation."

4. To what extent does teacher mobility contribute to differences in teacher value added across high and lower-poverty schools?

Our analysis has important implications for public policy. Given the large variation in teacher effectiveness, major improvements in student outcomes could be realized if schools were to identify the most successful teachers and deploy them in the settings where they could make the most difference. Having improved information about the likely contribution of teachers in high-poverty schools and in other schools, and the degree to which these differences can be explained by the types of factors that are observable ex ante, may help frame the magnitude of the potential problem of staffing schools serving disadvantaged students. And understanding the variation of measured teacher effectiveness within and across high-poverty schools and other schools can offer insight into the potential scope for and design of teacher compensation policies aimed at attracting and retaining highly effective teachers in the most challenging schools. It may also have implications for the design of performance accountability systems, in particular the balance targeted to the school level and individual teacher level.

Data and Sample

In this study we use student-level microdata from two states, Florida and North Carolina. Currently, these are the only states in which teachers and students can be linked to specific classrooms across all schools in the state for several years. The two states' data systems have some differences (e.g., different tests), but are generally comparable in terms of characteristics of students and teachers. We coordinate the analysis to ensure reasonable comparability in results across the two states.[2]

Sources

The primary source of data for Florida is the Florida Department of Education's K-20 Education Data Warehouse (FL-EDW). The FL-EDW is a longitudinal data system that includes individual records for students

[2] Specific variable definitions for the two states are provided in Appendices A and B.

and school personnel, as well as information on courses and schools. Both students and teachers can be linked to specific classrooms with a unique course offering identification variable. Available student-level information includes basic demographics (age, race/ethnicity, gender), program participation (special education, limited English proficiency, gifted, free/reduced-price lunch) and test scores. Teacher data includes demographics, experience, educational attainment, certification status and certification exam scores. Basic demographic and experience information is also available for principals.

North Carolina data used in this analysis were collected by the North Carolina Department of Instruction (NCDPI) and contain detailed administrative records on students, teachers, and schools. Each student enrolled in North Carolina public schools is assigned a unique randomized identifier which allows us to track individual students over time. Student-level information includes the school attended, ethnicity, sex, free/reduced price lunch status, English proficiency, special education status, level of parental education, and state assessment scores (end-of-grade (EOG) test scores in reading and math from grades 3 through 8). School level data include information on student membership, locale, grade span, and AYP status. Data are compiled each year by the North Carolina Education Research Data Center (NCERDC) at Duke University. We formed a student-level longitudinal file by merging annual datasets. School and teacher information was then linked to the student-level longitudinal data file.

Achievement Measures

Florida

Annual testing in all grades 3-10 began in Florida in the 1999-00 school year with the administration of the Florida Comprehensive Achievement Test – Norm Reference Test (FCAT-NRT). The FCAT-NRT is a version of the Stanford Achievement Test and, as the name implies, is tied to national norms. The FCAT-NRT is essentially a “no-stakes” test in Florida; it is not used for accountability purposes. In contrast, the FCAT Sunshine State Standards exam (FCAT-SSS) is used to assign school grades and determines individual student retention and secondary school graduation. The FCAT-SSS is a criterion-reference test based on the curriculum standards established in Florida. The FCAT-SSS was first administered in all grades 3-10 in the 2000-01 school year. We use the FCAT-SSS in this analysis for grades 3 through 5. Scores for the FCAT-SSS are normalized (with a mean of zero and a standard deviation of one) by grade, year and subject in order to control for any variation in the test over time. Differences in student scores from year to year are used to show student progress in terms of their performance relative to their peers' performance.

North Carolina

The state of North Carolina has required schools to administer math and reading end-of-grade exams for 3rd-8th graders since 1994 and subject-specific end-of-course exams in secondary school since 1996. These tests are typically administered during the last two weeks of the school year. At the elementary school level, math and reading scores are normalized by year and grade (with a mean of zero and standard deviation of one) so that these test scores are comparable across year and grade. Using these normalized scores, we create gain scores by taking the difference between the current year's score and the score from the previous year, and use these as our dependent variable in our elementary school analysis. We focus on grades 3 through 5 in elementary schools.

Samples

The study period covers school years 2000-01 through 2004-05 in both Florida and North Carolina. In both states we focus on elementary schools, grades 3-5. In order to ensure that we correctly attribute student achievement gains to individual teachers, we include only students in self-contained classrooms.

It should be noted that although teachers are directly linked to students in Florida, this is not the case for North Carolina data. In North Carolina, individual student test scores are linked to test proctors, who may not necessarily be the students' class instructors. Therefore, in North Carolina we take additional steps to verify if a test proctor is the actual teacher in a classroom. This is done by first linking test classes with instruction classes by unique proctor and teacher IDs and then comparing the aggregate student demographic characteristics (class size, number of White students, and number of male students) of a test classroom with those of an instruction classroom. If the test classroom sufficiently resembles the instruction classroom, we have reasonable confidence that the test proctor is also the instructor and therefore establish links between that teacher and his students. This strategy has been successfully applied to North Carolina data in earlier research,[3] and in this study we keep only those teachers who can be verified to be instructors in classrooms under study.

Second, we restrict the samples to reduce heterogeneity at the school level by excluding charter schools. The restriction has a relatively small effect on the number of schools included. In our first study year (2000/01) there were no charter schools in North Carolina and only 30 in Florida.

Third, in order to reduce noise and generate more reliable teacher value-added estimates, we limit analysis to classes with 10-40 students. Classes with more than 40 students are more likely to be test classes rather than instruction classes. Therefore including these large classes increases the probability of inappropriately attributing student performance to test proctors instead of instructors. On the other hand, small classes offer us too few student observations for each teacher, reducing the precision of teacher performance estimates for those teachers. Classes with more than 40 students or fewer than 10 students account for about four percent of all classes under study.

Table 1 summarizes the size of our analysis sample relative to the population of teachers in each state. We compare schools with poverty levels (defined by the percentage of students who are eligible for free/reduced-price lunch) above 70 percent with schools below 70 percent. We also compare schools at greater extremes: those above 70 percent poverty with those less than 30 percent poverty.[4] After applying the sample restrictions we are able to provide reliable value-added estimates for 14,052 unique elementary school teachers in Florida (5,232 in schools with over 70 percent FRL students; 6,975 in schools with between 30-70 percent FRL students; 3,084 in schools with less than 30 percent FRL students). Due to the difficulties in matching test proctors and teachers in North Carolina about 70 percent of teachers can be reliably linked to

[3] See Xu, Hannaway, and Taylor (2008) for details.

[4] To some extent these breaks are arbitrary, but have some rationale. Seventy percent is about the average poverty level in Title I schools in both states (NC=63 percent; FL=72 percent). And 30 percent is about the average poverty level in non-Title I schools in both states (NC=31 percent; FL=30 percent).

individual students in the Tar Heel state. This yields a sample of 9,212 elementary school teachers in North Carolina (2,316 in schools with over 70 percent FRL students; 5,945 in schools with between 30-70 percent FRL students; 2,207 in schools with less than 30 percent FRL students).

Table 2 presents the descriptive statistics for our samples from Florida and North Carolina. As expected, higher and lower-poverty schools serve different student populations. Compared to lower- poverty schools (<70% FRL and <30% FRL), high-poverty schools (>70% FRL) in both states have a larger share of African-American and Hispanic students. Indeed, the average percent of Black students in high- poverty schools is more than twice the percent in lower-poverty schools (<70% FRL) and more than four times the percent in the lowest-poverty schools (<30% FRL). And in both states student performance levels[5] are considerably lower in low-poverty schools than in high-poverty schools.[6]

Interestingly, in Florida the average performance *gains* in math are significantly higher in high- poverty schools (>70% FRL) than other schools. This is different from North Carolina, where the average performance gains in math in high-poverty schools are either not different or lower than gains in lower- poverty schools. In both states, the average gains in reading in high-poverty schools are lower than those seen in lower-poverty schools, but the differences are small.

Teacher qualifications are also different in high-poverty and lower-poverty schools in both states. Although Florida tends to have less experienced teachers in general, in both Florida and North Carolina high-poverty schools have a larger percentage of first-year teachers than lower-poverty schools. In Florida, the rate of inexperienced teachers (those teachers with two or fewer years of experience) in high-poverty schools is nearly double that in lower-poverty schools while in North Carolina the differences are much more modest. Teachers in high-poverty schools are also less likely to have a graduate degree, to hold a regular license and to be National Board certified than teachers in other schools. In Florida, principals in high-poverty schools are more likely to be new to the school than principals in lower-poverty schools. In North Carolina, the PRAXIS

[5] Note these are normalized performance measures, as discussed earlier.

[6] Indeed, in North Carolina students in schools with 0-70 percent FRL and in schools with 0-30 percent FRL outperform their counterparts in schools with more than 70% FRL students, on average, by about 0.5 and 0.7 standard deviations respectively in both math and reading. These findings are comparable to the Florida results.

scores of teachers in high-poverty schools are also significantly lower (by 0.03 to 0.04 standard deviations) than the Praxis scores of teachers in lower-poverty schools.

Analytic Strategy

Our basic strategy involves four steps. First, we estimate the determinants of individual student achievement in a "value-added" framework that takes into account prior student test performance and other factors (family background, peer characteristics, school management) that might contribute to student performance. The average student achievement of a teacher's students beyond these factors indicates the 'teacher effect'.[7] More specifically, the teacher effect is the average achievement of a teacher's students over and above what would be expected given the current and past values of students' own performance, family background, peer academic performance and school-level inputs. We distinguish between schools serving primarily students from low-income families (70 percent or more FRL) and those serving fewer students in poverty (less than 70 percent FRL).[8] Separate teacher-effect estimates are produced for each teacher unique teacher/school-type combination.

In the second step of our analysis, we separate the estimated teacher-school type effects into sub-samples of teachers in higher and lower-poverty schools. We determine how the mean 'teacher effect' and the dispersion of teacher effects vary across schools serving different proportions of students in poverty. For each sub-sample we then determine the relationship between teacher credentials and effectiveness by

---

[7] A basic value-added model can be written as $A_{it} = \lambda A_{it-1} + \boldsymbol{\beta}_1 \mathbf{X}_{it} + \boldsymbol{\beta}_2 \mathbf{P}_{-ijmt} + \boldsymbol{\beta}_3 \mathbf{S}_{mt} + \delta_k + \nu_{it}$, where *A* represents student achievement, **X** is a vector of student/family characteristics, **P** is a vector of classroom peer characteristics and **S** is a vector of school-level characteristics. The subscripts i, j, k and m denote individual students, classrooms, teachers and schools respectively. The variable δ_k is an indicator for teacher k and ν_{it} is a random error term. The estimated value of δ_k represents the effect of an individual teacher on student performance, holding constant other factors that affect test scores, include innate student ability and parental support, peer influences and school-level inputs like physical facilities and school leadership. Prior-year test scores are included in the model to account for the cumulative effects of past peer, teacher and school inputs. See Rivkin (2007) for discussion of value-added models.

[8] We only present estimates broken down by high-poverty (70+% FRL) and lower-poverty school (<70% FRL) types. However, we also estimated teacher/school-type effects for high-poverty (70+% FRL) and low-poverty schools (<30% FRL). Results using this breakdown of school types are similar to those presented in the paper and are included in an Appendix, which is available upon request.

regressing the teacher effect estimates on a set of observable teacher characteristics, including experience indicators, and indicator for post-baccalaureate degrees, and indicators for certification status.

Third, we decompose the difference in teacher effectiveness across schools serving students from different poverty levels into three components: differences due to differences in observable teacher characteristics or endowments, differences due to differences in the marginal effect of teacher characteristics on student achievement and differences due to the interaction of the differences in endowments and marginal effects. Thus, for example, we can determine if differences in average teacher quality are due to differences in observable qualifications or to differences in the "payoff" from qualifications (e.g. the impact of additional experience on student achievement) across school types.

Finally, we explore how differences in teacher mobility may explain differences in the experience-effectiveness profile of teachers in high-poverty and lower-poverty schools. We compare the effectiveness or value added of "stayers" and "leavers" for each school type.

Methodological Challenges

Non-random Sorting of Teachers and Students

One of the key challenges in estimating a teacher's contribution to student performance is the non-random sorting between teachers and students both across and within schools. Evidence has shown a matching between observed teacher qualifications (such as years of experience) and student achievement, possibly as a result of teacher preference and parent pressure (Clotfelter, Ladd, and Vigdor 2007). When both teacher quality and student performance are systematically related to student ability and motivation, the relationship between teacher and student performance cannot be reliably estimated.

To mitigate such potential biases resulting from non-random matching of teachers to students, we estimate models with student fixed effects in addition to basic value-added models with student covariates. Models with student fixed effects take advantage of repeated student performance measures over time, and identify teacher effects using within-student variation of teacher inputs. These student indicators absorb any time-invariant unobservable attributes of a student, such as student innate ability, race and gender. Put differently, student fixed effects models compare the performance of a student to her own average performance over time as she moves across different classrooms, containing different peers and teachers.[9]

Distinguishing Teacher and School-Setting Effects

A second key challenge to estimating teacher value added is distinguishing teacher effects from school effects. Our model includes school-level characteristics, where available, to control for observed characteristics of the school. In particular, in Florida we account for observed traits of the principal, including administrative experience, the square of experience, and whether the principal is new to the school. However, to the extent that there are important unmeasured school characteristics that influence student achievement gains, they will become part of the estimated teacher effect. Often value-added models include school "fixed effects" in order to capture the effects of unobserved time-invariant school-level factors. However, when school fixed effects are employed, individual teacher effects are measured relative to other teachers at the same school. Such a strategy is not relevant in the present study since our goal is to make comparisons of teacher quality across schools.

In order to determine the extent to which teacher estimates are picking up unmeasured school-level influences we conduct an analysis with teacher-school type effects, where school-type refers to the school's poverty level. Each unique "teacher-school type" combination is therefore treated as a separate teacher effect. Thus teachers who teach in both high-poverty (>70% FRL) and lower-poverty schools (<70% FRL) would generate two teacher effects estimates, one for years in which they are teaching in a high-poverty school and another for years in which they are teaching in a lower-poverty school.[10] For those teachers who switch school

[9] With student fixed effects and the assumption of non-decay of prior inputs the model becomes: $A_{it} - A_{it-1} = \boldsymbol{\beta}_1 \mathbf{X}_{it} + \boldsymbol{\beta}_2 \mathbf{P}_{-ijmt} + \boldsymbol{\beta}_3 \mathbf{S}_{mt} + \gamma_i + \delta_k + \nu_{it}$, where γ_i represents the individual student-specific effect. The student covariate model replaces student fixed effects with student time-invariant (and quasi-time-invariant) characteristics.

[10] We do not conduct an analysis with the <30% poverty because of an inadequate number of "switchers" in this smaller set of schools.

types (and thus generate two teacher effect estimates) we compare their estimated effects across the two school types. If the teacher effects for these "switchers" are not significantly different across school types, it suggests that our estimates of teacher quality are not significantly biased by unmeasured school-level characteristics. If, however, there are significant differences in the within-teacher effects across school types, there are two possible conclusions. It could be that there are significant school quality differences between school types or that teachers perform better in some school environments than in others.

Noise in Teacher Effect Estimates

As noted earlier, the estimated teacher effects are essentially the average gain of a teacher's students, conditional on student, peer and school characteristics that are beyond the control of the teacher. Student test scores tend to be "noisy," that is the same student will not achieve the same score on an exam each time they take it due to random factors like whether they got a good night's sleep, whether they are feeling ill on exam day, and whether or not they get lucky when guessing between a couple of possible multiple-choice answers. Such random fluctuations tend to cancel out when averaging over a large number of students. We therefore impose a restriction of a minimum of 10 students per teacher when reporting estimated teacher effects.

In addition to placing a restriction on the minimum number of students per teacher, it is also possible to mitigate dispersion in estimated teacher effects through the use of Empirical Bayes (EB) or "shrinkage" estimators. The EB estimates of teacher productivity are essentially a weighted average of the individual teacher effect estimates and the average teacher effect estimate, with greater weight given to the individual estimates the smaller is their standard error.[11] As noted by McCaffrey et al. (2010), standard fixed-effects software routines compute fixed effects relative to some arbitrary hold-out unit (e.g. an omitted teacher), which can produce incorrect standard errors and thus inappropriate Empirical Bayes estimates. Therefore, to estimate the teacher effects and their standard errors we employ the Stata routine *felsdvregdm*, developed by

[11] For details on the computation of Empirical Bayes estimates, see Morris (1983) and Jacob and Lefgren (2005).

Mihaly et al. (2010), which imposes a sum-to-zero constraint on the teacher estimated teacher effects and produces the appropriate standard errors for computing the Empirical Bayes (shrunken) estimates of teacher value added. For each state, our estimates of teacher effectiveness can therefore be interpreted as the impact of a teacher on student learning gains relative to the average teacher in our sample.

Results

Differences in Mean Teacher Value Added Across High/Lower-Poverty Schools

Table 3 compares value-added estimates of teachers in high- and lower-poverty elementary schools by state and subject. Two sets of estimates are shown in the table. The first is based on student fixed-effects models and the second is based on models that replace student fixed-effects with student covariates. As discussed above, there is an inherent trade-off between the student covariate and student-fixed effects analyses. In the student covariate models we are essentially comparing the effects of teachers on the achievement of students with the same observable characteristics. If students differ in unobserved ways and those unobserved characteristics are correlated with teacher assignments, estimates of teacher effectiveness will be biased. Thus, what appear to be differences in teacher quality could, in fact, be due to systematic differences in unobserved student characteristics across teachers. In contrast, the student-fixed-effects estimates essentially compare the impact of different teachers on annual achievement gains for the same student. This has the advantage of controlling for unobserved time-invariant student characteristics, thereby mitigating the bias associated with non-random assignment of students to teachers. The downside of student fixed-effects estimates are they tend to be quite noisy (i.e. the gains a specific student makes in any one year tend to bounce around a lot over time), more so than cross-student estimates from student covariate models. This contributes to noise in the teacher effect estimates, which would tend to make differences in teacher quality across schools appear to be statistically insignificant.

The estimates in Table 3 show that in three of four comparisons high-poverty-school (>70% FRL) teachers in North Carolina are less effective than lower-poverty-school (<70% FRL) teachers. The difference across school types is not significant for the fixed-effect estimates in reading, which is not surprising given the inherent noise in of student-fixed-effects models. In each comparison the differences are fairly small, on the order of 0.02 to 0.04 of a standard deviation. In Florida, the results are more mixed. The mean value added of math teachers in high-poverty schools is slightly higher than the mean for teachers in lower-poverty schools (<70% FRL). In reading, the findings from the student-covariate models were more like North Carolina; teachers in lower-poverty schools tend to outperform those in high-poverty schools. Once again, the differences are statistically insignificant for value-added estimates in reading based on the student-fixed-effect model. As with North Carolina, observed differences in mean teacher effectiveness are small in magnitude. Taken together, results from the two states cast doubt on the conventional wisdom that teacher quality in high-poverty schools is uniformly worse than in lower-poverty schools.

As discussed above, one challenge to our findings is that the teacher value-added estimates are absorbing school effects. If this is the case, school conditions affecting student performance would be attributed to teachers. Including school level controls – % FRL, % LEP, new principal indicator, principal experience – deals with this problem to some extent. But to further examine this issue we also compare the effectiveness of teachers who taught in both high and lower-poverty schools (<70% FRL). For these teachers we have two estimates of their effectiveness, one in a high-poverty school and one in a lower- poverty school. The results of this comparison, presented in Table 4, suggest that our estimates of differences in teacher quality across school types are not biased by omission of unobserved school characteristics. In both Florida and North Carolina there are no significant differences in measured value added associated with switching school types.

Differences in the Variability of Teacher Value Added Across High/Lower-Poverty Schools

In addition to mean value added, we also compare the within-school-type variation of teacher effectiveness across high-poverty and lower-poverty schools in Table 3. The overall variation in productivity among teachers is striking, as has been shown in earlier research (e.g., Rivkin, Hanushek, and Kain 2005). For both high-poverty and lower-poverty schools the difference between teachers' effectiveness at the 10 percentile and the 90th percentile in math is one-third to one-half a standard deviation in both states; in reading it is about a quarter of a standard deviation.

The standard deviations associated with high-poverty schools (>70% FRL) are significantly larger than those associated with lower-poverty schools (<70% FRL) for every comparison in both Florida and North Carolina, except for the estimates based on the student fixed effect model of math achievement in North Carolina. Thus while there are not large and consistent differences in average teacher quality across school types, there does appear to be much greater heterogeneity in teacher quality within the group of schools serving primarily students from low-income households.

Estimates in Table 5 and the corresponding graphical presentation in Figures 1 and 2 provide a more detailed look at the variation in teacher quality within high-poverty and lower-poverty schools. For both states and in both subjects, the least effective teachers in high-poverty schools contribute less to student achievement gains than do the least effective teachers in lower-poverty schools. Across both subjects and states we find that the advantage in teacher effectiveness experienced by lower-poverty schools diminishes as one moves up the quality distribution. Specifics of the teacher quality gradient do vary by subject and state, however. For reading in Florida, the gap in teacher quality across school types becomes insignificant as one moves up the quality scale whereas the top teachers in lower-poverty schools still outperform their counterparts in high-poverty schools in North Carolina by a very small amount. On the math side, the best teachers in lower-poverty schools hold a slight edge in effectiveness over the best teachers in high-poverty schools. In Florida, the relative positions actually reverse; the best math teachers in high-poverty schools are somewhat more effective than the top-tier math teachers in lower-poverty schools. Put simply, the worst teachers in high-poverty schools are less effective than the teachers at the bottom of the distribution in lower-poverty schools. The gap narrows as one moves up the quality distribution, however. The best teachers in high-poverty schools are nearly as good or in one case slightly better than the best teachers in lower-poverty schools.

The differential at the bottom end of the teacher quality distribution between high and low-poverty schools poses important policy questions. Is the difference due to recruiting lower quality teachers in the first place in high-poverty schools, or is it due to having more inexperienced teachers in those schools? If inexperienced teachers in different school types are of equal quality, then one should be focusing on salary and other mechanisms for retaining teachers. If high-poverty schools get worse draws in the first place, then targeted hiring incentives may be optimal.

To explore this issue, we restrict our sample of teachers to those with two or fewer years of experience, and compare their effectiveness between high and low-poverty schools in Table 6. In North Carolina, the effectiveness of inexperienced teachers (both math and reading) in high-poverty schools is generally not different from that in low-poverty schools. In fact, the best (the 75th percentile for math teachers and the 75th and the 90th percentile for reading teachers) inexperienced teachers in high-poverty schools are better than the best inexperienced teachers in low-poverty schools. In Florida, inexperienced math and reading teachers in low-poverty schools outperform their counterparts in high-poverty schools at the lower end of the effectiveness distribution, whereas the reverse is true at the higher end of the distribution. No difference is detected at the median. In short, in both states, inexperienced teachers are not discernibly different in terms of their effectiveness between high and low-poverty schools on average. However, there is more variation of teacher quality among inexperienced teachers in high-poverty schools than those in low-poverty schools. The greater variation is due in part to the best new teachers in high-poverty schools being more effective than the top inexperienced teachers in lower-poverty schools.

Our results thus far indicate that teacher quality is not uniformly lower in high-poverty schools. Rather, differences in average teacher quality appear to be driven by the relatively poor performance of the least effective teachers in high-poverty schools. It appears that the differential at the bottom end of the teacher quality distribution between high-and low-poverty schools is not due to the recruiting of low-quality teachers by high-poverty schools. In order to better understand the factors contributing to the differences in the distribution of teacher quality across school types we turn to an analysis of the observable characteristics of teachers.

Sources of Variation in Teacher Quality between High-Poverty and Lower-Poverty Schools

Table 7 presents results from a Oaxaca-Blinder decomposition of the variation in estimated teacher effectiveness.[12] Observable teacher characteristics include a vector of teacher experience categories and indicators for teacher educational attainment and licensure status. For both math and reading in North Carolina, differences in average teacher quality between high-poverty and lower-poverty schools are almost entirely (over 90 percent) due to differences in the marginal return to characteristics, rather than the levels of the characteristics themselves.[13] While a bit more muddled, the results for Florida also suggest that differences in observable characteristics are not at the heart of differences in teacher quality across high-poverty and lower-poverty schools. For math in Florida, average teacher quality is actually slightly higher in high-poverty schools; though, like in North Carolina, essentially all of the quality differential is due to differences in the marginal effects of characteristics, rather than differences in the levels of qualifications. Over half of the mean difference in teacher quality in reading in Florida is attributable to differences in the

[12] For regressions where the estimated teacher effect is the dependent variable (Tables 7 and 8) we employ the standard (non-shrunken) teacher effect estimates since the shrunken estimates would distort the estimated marginal effects of teacher characteristics on teacher productivity. Given that the dependent variable (the teacher effect) is an estimated value we employ feasible generalized least squares to account for estimation error.

[13] The reference group in the decomposition is teachers in high-poverty schools. Therefore the differences in teacher effectiveness due to differences in the marginal effects are evaluated at the mean characteristics of teachers in high-poverty schools.

marginal returns to characteristics and only about 16 is due to differences in the levels of characteristics themselves. The remaining 26 percent is due to the interaction of characteristics and marginal effects.

The differences in the marginal returns to specific characteristics are discernable in the regression results presented in Table 8. In lower-poverty schools in North Carolina additional experience pays dividends through the 6-12 year category. After that the difference in effectiveness between veterans and the least experienced (0-2 years of experience) teachers remains relatively constant as experience increases. In high-poverty schools the payoff for 3-5 years of experience is about equal to that in lower-poverty schools. However, beyond the five-year mark, the returns to experience appear to drop off in high-poverty schools. For North Carolina math teachers in high-poverty schools, teachers with 13 or more years of experience are equal in effectiveness to the least experienced teachers working in high-poverty schools. For reading the results are more variable, but the overall pattern suggests a much weaker link between experience and effectiveness among reading teachers in high-poverty schools. For Florida, there is a weaker relationship between experience and effectiveness than in North Carolina, regardless of school type. Among Florida teachers in high-poverty schools there are no discernable differences in teacher quality across experience categories in either math or reading. For reading teachers in lower-poverty schools, teachers with six or more years of experience out-perform those with five or fewer years of experience. In math, mid-career teachers (13-20 years of experience) are slightly more effective than the least experienced teachers, but otherwise the payoff to experience appears negligible.

Some interesting differences in the returns to graduate degrees and full certification appear as well. For both math and reading in North Carolina, full licensure is positively correlated with teacher effectiveness, while educational attainment has either no significant effect (math) or a negative correlation with value added (reading). In contrast, neither educational attainment or licensure status are significantly correlated with teacher value added in high-poverty schools in North Carolina. In Florida the differential effects of educational attainment and licensure status on teacher effectiveness are less dramatic. For both types of schools and in both subjects, Florida elementary school teachers holding a regular license are more effective than their colleagues with temporary licenses. The estimated marginal effects of full licensure do appear to be somewhat lower in high-poverty schools, however. In Florida the correlation between graduate degrees and value added is always insignificant, save for a very small positive effect in math.

To summarize, our findings thus far indicate small differences in average teacher quality between high-poverty and lower-poverty schools, but significantly greater variation in effectiveness among teachers working in high-poverty schools. The greater variability in quality among teachers working in high-poverty schools appears to be caused by lower quality of the least effective teachers. The best teachers in high-poverty schools are on par with the best teachers in lower-poverty schools but the least effective teachers in high-poverty schools are much less effective than their counterparts in lower-poverty schools. The observed differences in teacher quality across school types appear to be due to differences in the marginal return or "payoff" to teacher characteristics rather than the level of qualifications. In particular, the experience-productivity relationship appears to be much stronger in lower-poverty schools. We explore possible explanations for the differential returns to experience in the next section.

Why is the Return to Teacher Experience Lower in High-Poverty Schools?

We posit four possible explanations for the observed differences in the relationship between teacher experience and teacher productivity in different school settings. The first two possibilities relate to differences in teacher career paths. Recent work on teacher labor market decisions suggests that the relationship between teacher quality and teacher attrition/mobility may vary across school settings (Boyd, et al. 2007; Goldhaber, Gross, and Player 2007). If relatively low-quality new teachers in high-poverty schools are less likely to leave their initial schools than are their counterparts in schools serving more affluent student populations, then the experience-quality profile would be flatter in high-poverty schools. Differential attrition of low-quality teachers could be due to lower opportunity costs or less effective monitoring of teacher performance in high-poverty schools. Alternatively, differences in teacher mobility across school types could be driving observed differences in teacher quality-experience patterns. There might be a "Dance of the

Lemons" whereby low-productivity teachers with some experience in lower-poverty schools eventually migrate to high-poverty schools. This would tend to lower the average quality of experienced teachers in high-poverty schools (and raise the average the quality of experienced teachers in lower-poverty schools) which would make the slope of the experience-quality relationship appear flatter in high-poverty schools.

In Table 9 we present the average quality of early-career teachers, broken down by school type and mobility. We find only modest support for the notion that lower-poverty schools do a better job at culling out low-performing early-career teachers. In Florida, we find no difference in the first-year performance of stayers and leavers in math for either school type. Among reading teachers, leavers tend to be less effective in their first year of teaching than stayers, but the difference is about equal across school types. Interestingly, for both school types and both subjects Florida teachers who leave their initial school at the end of two years were more effective in their first year of teaching than teachers who stayed at their initial school more than two years.

There is somewhat stronger evidence of differential attrition in North Carolina. There we find that across both subjects, teachers in lower-poverty schools who stay on after the first year were more effective in their first year of teaching than those who leave at the end of the first year. In contrast, we find no significant difference in first-year value added between first-year stayers and leavers in high-poverty schools in North Carolina. These differences do not hold up for teachers who depart at the end of their second year. We find no significant difference in either first-year or second-year teacher performance between second-year stayers and leavers in either subject and in either school type in North Carolina.

In Table 10 we present evidence on the relationship between teacher mobility and teacher quality across school types. Consistent with evidence presented in Table 4, we find no significant within-teacher differences in productivity as they switch from lower-poverty to higher-poverty schools or vice-versa.[14] Further, we find no evidence in support of the "Dance-of-the-Lemons" hypothesis. There are no significant

[14] The quality differentials for switchers in Tables 4 and 10 are not directly comparable because the samples are different. In Table 4, teachers who switch schools types more than once are included as are "switches" that occurred because a school changed poverty status (even if a teacher remained in the same school). Table 10 only includes one-time switchers and teachers who actually changed schools.

differences in the average quality of teachers who switch from lower-poverty to higher-poverty schools vis-à-vis teachers who move from higher-poverty schools to lower-poverty schools.

In addition to differences in attrition/mobility, observed differences in the quality-experience profile across school types could result from direct effects of the school environment on changes in teacher quality over time. Teachers who stay in high-poverty schools may simply "burn-out" faster, resulting in smaller increases in productivity over time compared to teachers working in less stressful environments. While we have no direct evidence on the burn-out hypotheses, it is consistent with evidence that regular education teachers are more likely to leave schools with challenging student populations (Lankford, Loeb, and Wyckoff 2002; Hanushek, Kain, and Rivkin 2004; Boyd et al. 2005; Clotfelter, Ladd, and Vigdor 2005; Imazeki 2005; Scafidi, Sjoquist, and Stinebrickner 2007; Feng 2009) and that special education teachers often cite the stress of working with students with special needs and the lack of pupil progress relative to effort expended as reasons for switching from special to regular education (Billingsley and Cross 1991). Further, peer effects may play a role in the apparent diminished effect of experience on productivity in high-poverty schools. Feng and Sass (2008) find that teachers are more likely to leave their initial placement the greater is the gap between their own productivity and the average quality of other teachers at their school. The most effective teachers who transfer tend to go to schools whose faculties are in the top quartile of teacher quality. Jackson and Bruegmann (2009) show that a teacher's students have higher achievement gains the higher the value added of the teacher's colleagues.

Conclusions and Policy Implications

This study focuses on the effectiveness of teachers in high-poverty and lower-poverty schools. Large and persistent differences in student performance between schools serving large proportions of students from low-income households and those primarily students from more affluent households exist. Prior research indicates that teachers are the most important school factor affecting student achievement and thus policies designed to impact teacher quality are the logical starting point for reducing gaps in school performance. Prior federal efforts, primarily Title I, as well as state efforts target dollars to schools serving the most disadvantaged students, but they provide wide local latitude in determining what happens in these schools, including the assignment of teachers.

Our findings show that teachers in high-poverty schools are generally less effective than teachers in lower-poverty schools, though the differences are small and not consistent across states and subject areas. We do find consistent evidence, however, that the variation in effectiveness among teachers in high- poverty schools is greater than the variation among teachers in lower-poverty schools. Differences in the distribution of teacher quality appear to be driven by the relatively poor performance of the least effective teachers in high-poverty schools; the best teachers in high-poverty schools are on par with the best teachers in lower-poverty schools but the least effective teachers in high-poverty schools are much less effective than their counterparts in lower-poverty schools.

The observed differences in teacher quality across school types appear to be due to differences in the marginal return or "payoff" to teacher characteristics rather than the level of qualifications. In particular, the experience-productivity relationship appears to be much stronger in lower-poverty schools. The lower return to experience in high-poverty schools does not seem to be a result of differences in the quality of teachers who leave teaching or who switch schools. Rather, it may be the case that the effect of experience on teacher productivity depends on the setting in which the experience is acquired. If there are positive peer effects among teachers that depend on teacher quality or if exposure to challenging student populations produces "burn out" and lessens the future productivity of teachers, teachers in schools serving large proportions of low-income students may simply not improve much as time goes by.

Our results have implications for both school accountability and teacher labor-market policies. Because high-poverty schools also tend to have lower performance levels, they tend to be schools that are identified as not performing up to par and subject to accountability pressure. To the extent that school- level measures are used in accountability systems, they are likely to inadequately appreciate the contributions of the best teachers in these schools who are performing at least as well as the top teachers in more advantaged and higher performing schools. Our findings on the variability of teacher effectiveness call for accountability mechanisms that take into account not only school level performance measures, but also the individual level teacher contributions.

Our results suggest that solutions to the achievement gap between high-and lower-poverty schools may be complex. Changing the quality of new recruits or importing teacher with good credentials into high-poverty schools may not be sufficient. Rather, our findings indicate that measures that promote retention of the most-effective teachers already in high-poverty schools that induce highly effective experienced teachers in lower-poverty schools to move to high-poverty schools and which promote an environment in which teachers' skills will improve over time are more likely to be successful.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25(1): 95–135.

Aud, Susan, William Hussar, Michael Planty, Thomas Snyder, Kevin Blanco, Mary Ann Fox, Lauren Frohlich, Jana Kemp, and Lauren Drake. 2010. *The Condition of Education 2010*. (NCES 2010-028). Washington, D.C.: U.S. Department of Education, Institute of Education Sciences.

Balfanz, Robert, and Nettie Legters. 2004. *Locating the Dropout Crisis: Which High Schools Produce the Nation's Dropouts? Where are they Located? Who Attends Them?* Report No. 70. Baltimore, MD: Center for Research on the Education of Students Placed at Risk (CRESPAR), Johns Hopkins University.

———. 2005. "The Graduation Gap: Using Promoting Power to Examine the Number and Characteristics of High Schools with High and Low Graduation Rates in the Nation and Each State." CSOS Policy Brief. Baltimore, MD: Center for Social Organization of Schools, Johns Hopkins University.

Billingsley, Bonnie, and Lawrence H. Cross. 1991. "Teachers Decisions to Transfer from Special to General Education." *Journal of Special Education* 24(4):496–511.

Boyd, Donald J., Hamilton Lankford, Susanna Loeb, and James H. Wyckoff. 2005. "The Draw of Home: How Teachers' Preferences for Proximity Disadvantage Urban Schools." *Journal of Policy Analysis and Management* 24(1):113–32.

Boyd, Donald J., Pamela L. Grossman, Hamilton Lankford, Susanna Loeb, and James H. Wyckoff. 2007. "Who Leaves? Teacher Attrition and Student Achievement." CALDER Working Paper 23. Washington, D.C.: The Urban Institute.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2005. "Who Teaches Whom? Race and the Distribution of Novice Teachers." *Economics of Education Review* 24(4):377–92.

———. 2007. "How and Why Do Teacher Credentials Matter for Student Achievement?" CALDER Working Paper 2. Washington, D.C.: The Urban Institute.

Dolton, Peter, and Wilbert Van der Klaauw. 1999. "The Turnover of UK Teachers: A Competing Risks Explanation." *Review of Economics and Statistics* 81(3):543–52.

Feng, Li. 2009. "Opportunity Wages, Classroom Characteristics, and Teacher Mobility." *Southern Economic Journal* 75(4):1165–90.

Feng, Li, David N. Figlio, and Tim R. Sass. 2010. "School Accountability and Teacher Job Choice." NBER Working Paper 16070. Cambridge, MA: National Bureau of Economic Research, Inc.

Feng, Li, and Tim R. Sass. 2008. "Teacher Quality and Teacher Mobility." Mimeo. Florida State University.

Figlio, David N., and Maurice E. Lucas. 2004. "Do High Grading Standards Affect Student Performance?" *Journal of Public Economics* 88(9-10):1815–34.

Goldhaber, Dan, Hyung-Jai Choi, and Lauren Cramer. 2007. "A Descriptive Analysis of the Distribution of NBPTS-Certified Teachers in North Carolina." *Economics of Education Review* 26(2):160–72.

Goldhaber, Dan, Bethany Gross, and Daniel Player. 2007. "Are Public Schools Really Losing Their "Best"? Assessing the Career Transitions of Teachers and Their Implications for the Quality of the Teacher Workforce." CALDER Working Paper 12. Washington, D.C.: The Urban Institute.

Imazeki, Jennifer. 2005. "Teacher Salaries and Teacher Attrition." *Economics of Education Review* 24(4):431–49.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004. "Why Public Schools Lose Teachers." *Journal of Human Resources* Vol. 39(2): 326–54.

Harris, Douglas N., and Tim R. Sass. 2007. "Teacher Training, Teacher Quality and Student Achievement." CALDER Working Paper 3. Washington, D.C.: The Urban Institute.

Jackson, C. Kirabo, and Elias Bruegmann. 2009. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers." *American Economic Journal: Applied Economics,* 1(4): 85–108.

Jacob, Brian A., and Lars Lefgren. 2005. "Principals as Agents: Subjective Performance Measurement in Education." NBER Working Paper 11463. Cambridge, MA: National Bureau of Economic Research, Inc.

Lankford, Hamilton, Susanna Loeb, and James H. Wyckoff. 2002. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis." *Educational Evaluation and Policy Analysis* 24(1):37–62.

McCaffrey, Daniel F., J. R. Lockwood, Kata Mihaly, and Tim R. Sass. 2010. "A Review of Stata Routines for Fixed Effects Estimation in Normal Linear Models." Mimeo.

Mihaly, Kata, Daniel F. McCaffrey, J.R. Lockwood, and Tim R. Sass.2010. "Centering and Reference Groups for Estimates of Fixed Effects: Modifications to Felsdvreg." *Stata Journal* 10(1): 82-103.

Morris, Carl N. 1983. "Practical Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association* 78(381): 47–55.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73(2): 417–58.

Rivkin, Steven G. 2007. "Value-Added Analysis and Education Policy." CALDER Brief 1. Washington, D.C.: The Urban Institute.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94(2): 247–52.

Scafidi, Benjamin, David L. Sjoquist, and Todd R. Stinebrickner. 2007. "Race, Poverty, and Teacher Mobility." *Economics of Education Review* 26(2):145–59.

Uzzell, Renata, Candace Simon, Amanda Horwitz, Anne Hyslop, Sharon Lewis and Michael Casserly. 2010. *Beating the Odds: Analysis of Student Performance on State Assessments and NAEP*. Washington, DC: Council of the Great City Schools.

Xu, Zeyu, Jane Hannaway, and Colin Taylor. 2008. "Making a Difference? The Effects of Teach for America in High School." CALDER Working Paper 17. Washington, D.C.: The Urban Institute.

Tables and Figures

Table 1. Number of Teacher in State and Sub-samples, by State and School Poverty Category

State and poverty category	Teachers of relevant classes in state[1]	Teachers linked to students [2]	Teachers with reliable value-added estimates [3]
Florida			
Elementary school	26,525	26,525	14,052
0-30% FRL	8,418	8,418	3,084
30-70% FRL	17,028	17,028	6,975
70-100% FRL	13,620	13,620	5,232
North Carolina			
Elementary school	25,238	17,504	9,212
0-30% FRL	6,062	4,275	2,207
30-70% FRL	16,061	11,347	5,945
70-100% FRL	7,128	4,670	2,316

[1] Classes include self-contained classes in grades 4 and 5. Since teachers can switch between schools in different poverty categories, and since in some schools the percentage of FRL students is not available in all years, the sum of the teachers in school poverty categories do not add up to the total number of unique teachers independent of school poverty category.
[2] In North Carolina individual student test scores are linked to test proctors, who are not necessarily classroom instructors. We compare the student demographic characteristics of the test classrooms and the instructional classrooms and keep only those teachers who can be verified to be instructors. In Florida, there exist course offering codes that identify each unique classroom. Both students and teachers can be linked to specific course offerings.
[3] Our final analytic samples exclude teachers from charter schools. Additionally, we exclude classrooms with fewer than 10 students or more than 40 students in our estimation samples. Further, value-added estimates are only reported for teachers who taught at least 10 students with information on all of the achievement model variables.

Table 2. Student, Teacher and School Characteristics in Analytic Sample, by State and School Poverty Category

	Florida			North Carolina		
Student, teacher and school characteristic	*0-30% FRL*	*0-70% FRL*	*70-100% FRL*	*0-30% FRL*	*0-70% FRL*	*70-100% FRL*
Students						
Male (%)	50.72**	49.95**	48.87	50.77**	50.59	50.61
White (%)	73.81**	61.68**	23.68	78.96**	68.17**	25.66
Black (%)	9.30**	15.78**	42.13	13.28**	22.30**	58.36
Hispanic (%)	10.93**	17.18**	31.03	2.88**	4.82**	8.36
Other minorities (%)	5.95**	5.35**	3.16	4.88**	4.70**	7.62
Student performance (level scores)						
Math	0.49**	0.25**	-0.16	0.43**	0.15**	-0.32
Reading	0.50**	0.26**	-0.18	0.39**	0.14**	-0.34
Student performance (gains scores)						
Math	-0.02	-0.01**	0.02	0.02*	0.01	0.02
Reading	-0.01**	-0.01**	-0.01	0.02**	0.01**	0.00
Teachers/Principals						
Experience (% of teachers)						
0 years	7.43**	9.09**	12.01	2.64**	3.39**	5.40
1-2 years	14.11**	16.65**	26.43	14.18**	16.20	17.41
3-5 years	16.2**	16.01**	16.13	16.83**	16.01	15.76
6-12 years	25.48**	22.94**	19.41	26.04*	23.97	23.33
13-20 years	21.58**	18.16**	11.19	17.60**	16.25**	13.58
21-27 years	9.41**	9.79**	8.04	14.18**	14.47	13.80
28 or more years	5.81**	7.35**	6.59	8.53*	9.71	10.71
Graduate degree (%)	35.72**	31.46**	29.23	32.09*	29.65	28.80
Regular license (%)	94.81**	93.36**	88.66	97.96**	96.62**	94.42
NBPTS certified (%)	4.52**	3.58**	1.70	13.15**	10.67**	6.92
Praxis score	—	—	—	0.21**	0.13**	-0.20
Principal's experience	11.23**	11.48**	11.04	—	—	—
New principal indicator (%)	13.18**	13.99**	17.54	—	—	—
Schools						
Poverty level (% FRPL)	17.36**	40.40**	85.54	20.45**	42.84**	83.91
LEP students (%)	3.87**	6.17**	17.98	—	—	—
Special education students (%)	15.54**	17.16**	16.37	—	—	—

— Not available

Note: All statistics are based on the math analysis sample, except for reading achievement test scores (which are based on the reading analysis sample). * denotes that the estimate is statistically different from the corresponding estimate for schools with 70-100% FRL students at the 5% level, and ** denotes significance at the 1% level.

Table 3. Means and Standard Deviations of Teacher Value Added, by State, School Poverty Category, Subject and Model

	Florida			North Carolina		
Subject and model	*0-70% FRL*	*70-100% FRL*	*Difference*	*0-70% FRL*	*70-100% FRL*	*Difference*
Means						
Math						
FE estimates	-0.0782	-0.0709	-0.0073**	-0.1030	-0.1207	0.0177**
Student covariate estimates	-0.0112	-0.0022	-0.0090**	-0.0054	-0.0434	0.0380**
Reading						
FE estimates	-0.1637	-0.1631	-0.0006	0.1533	0.1541	-0.0008
Student covariate estimates	-0.0057	-0.0091	0.0034**	-0.0039	-0.0164	0.0125**
Standard deviations						
Math						
FE estimates	0.1043	0.1204	-0.0161**	0.0867	0.0880	-0.0013
Student covariate estimates	0.1216	0.1495	-0.0279**	0.2474	0.2749	-0.0275**
Reading						
FE estimates	0.0521	0.0587	-0.0007**	0.1096	0.1259	-0.0164**
Student covariate estimates	0.0482	0.0588	-0.0106**	0.0751	0.0860	-0.0109**

Note: Differences between Title I and Non-Title I teacher effects (t-test) / standard deviations (F-test) are * significant at 5%, ** significant at 1%

Table 4. Differences in Estimated Teacher Value Added for Teachers Who Taught in both 0-70% FRL Schools and 70-100% FRL Schools, by State, Subject and Model

	Florida		North Carolina	
Subject and model	*Mean differences*	*Standard error*	*Mean differences*	*Standard error*
Elementary school				
Math				
FE estimates	0.0001	0.0048	-0.0034	0.0112
Student covariate estimates	0.0014	0.0043	-0.0069	0.0072
Reading				
FE estimates	-0.0031	0.0029	-0.0057	0.0050
Student covariate estimates	-0.0043	0.0023	0.0013	0.0041

Note: Differences between high and low-poverty teacher effects (t-test) are * significant at 5%, ** significant at 1%

Table 5. Teacher Value Added at Various Percentiles, by State, Subject and School Poverty Category

	Florida			North Carolina		
Subject and teacher performance percentile	*0-70% FRL*	*70-100% FRL*	*Difference*	*0-70% FRL*	*70-100% FRL*	*Difference*
Math						
10	-0.1622	-0.1805	0.0181**	-0.2076	-0.2454	0.0378**
25	-0.0823	-0.0836	0.0013	-0.1081	-0.1318	0.0238**
50	0.0087	0.0156	-0.0071*	0.0035	-0.0269	0.0303**
75	0.1004	0.1236	-0.0231**	0.1196	0.0992	0.0205**
90	0.1895	0.2293	-0.0397**	0.2259	0.2078	0.0180*
Reading						
10	-0.0917	-0.1106	0.0189**	-0.1049	-0.1257	0.0208**
25	-0.0455	-0.0577	0.0122**	-0.0527	-0.0696	0.0168**
50	0.0084	0.0001	0.0083**	0.0056	-0.0089	0.0145**
75	0.0610	0.0582	0.0030	0.0627	0.0507	0.0121**
90	0.1104	0.1141	0.0036	0.1150	0.1038	0.0112**

Note: * denotes statistical significance at the 5% level, and ** denotes statistical significance at the 1% level

Figure 1A. Teacher Value Added at Various Percentiles for High-and Lower-Poverty Schools, Math Teachers in North Carolina

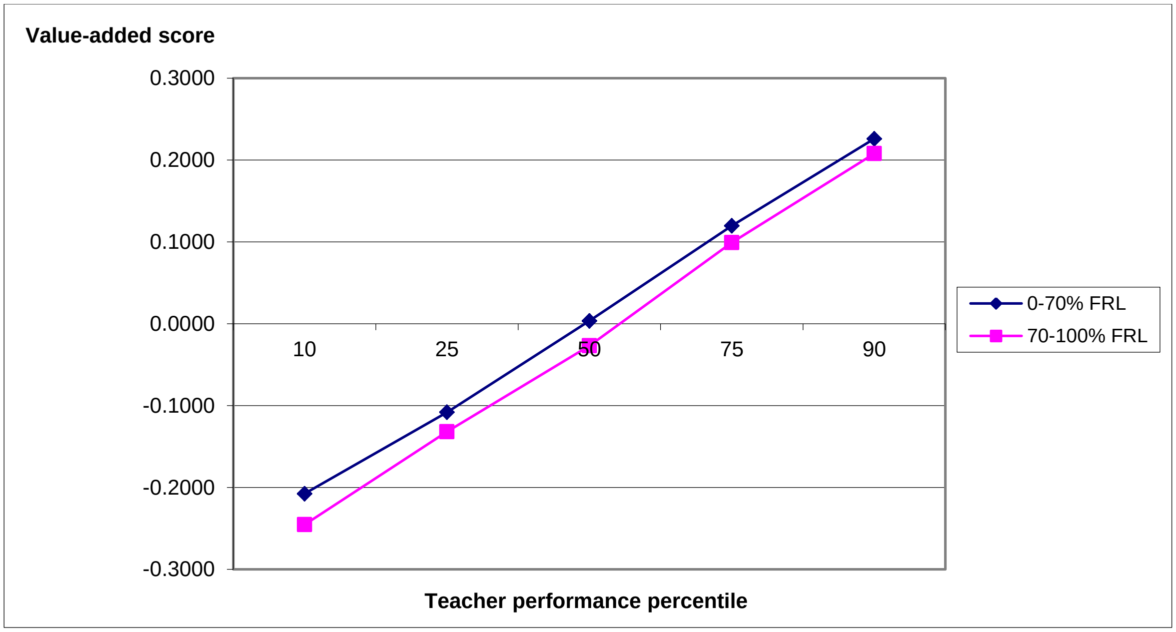


Figure 1B. Teacher Value Added at Various Percentiles for High- and Lower-Poverty Schools, Reading Teachers in North Carolina

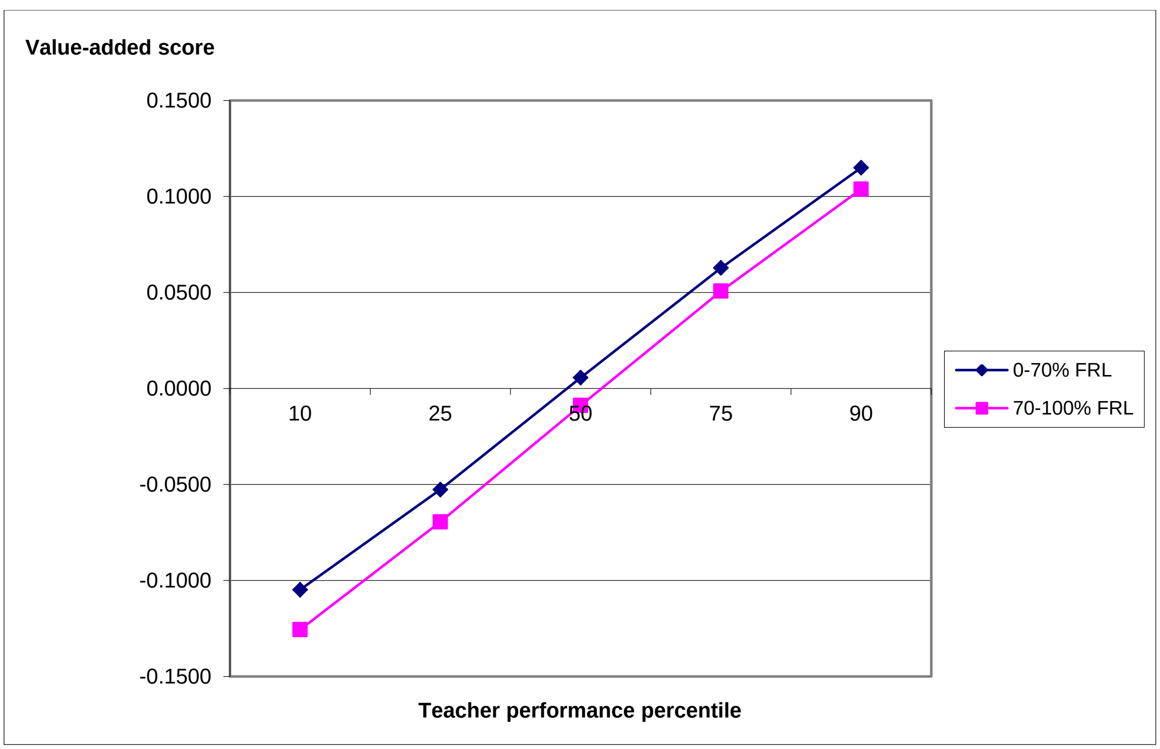


Figure 2A. Teacher Value Added at Various Percentiles for High- and Lower-Poverty Schools, Math Teachers in Florida

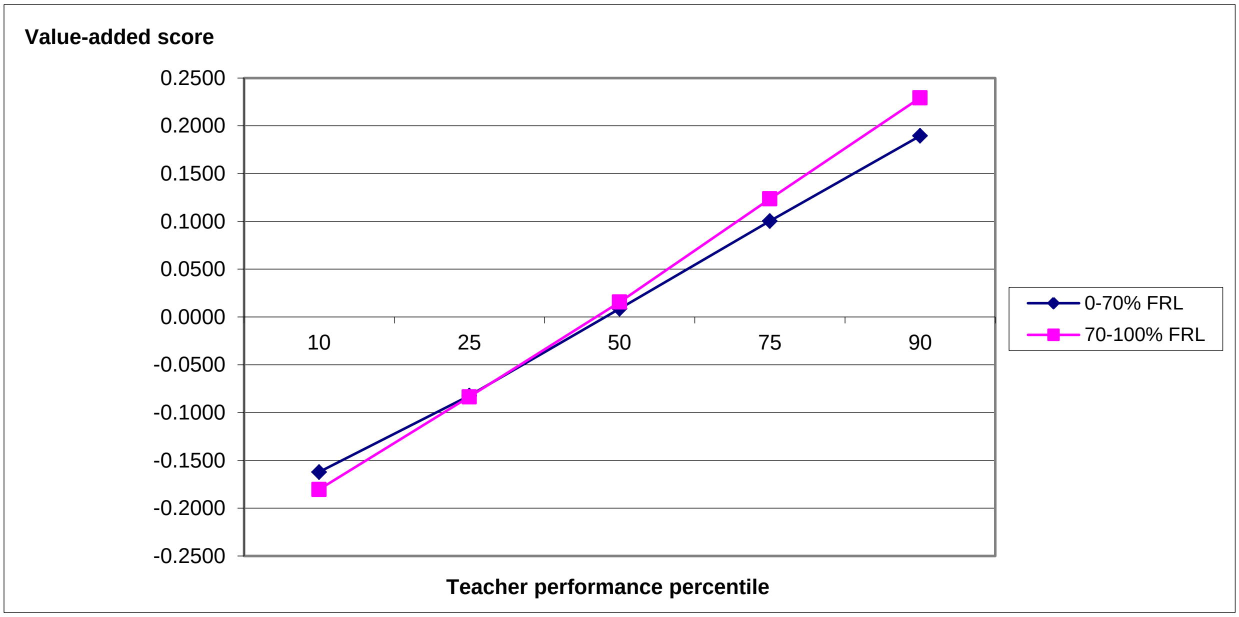


Figure 2B. Teacher Value Added at Various Percentiles for High- and Lower-Poverty Schools, Reading Teachers in Florida

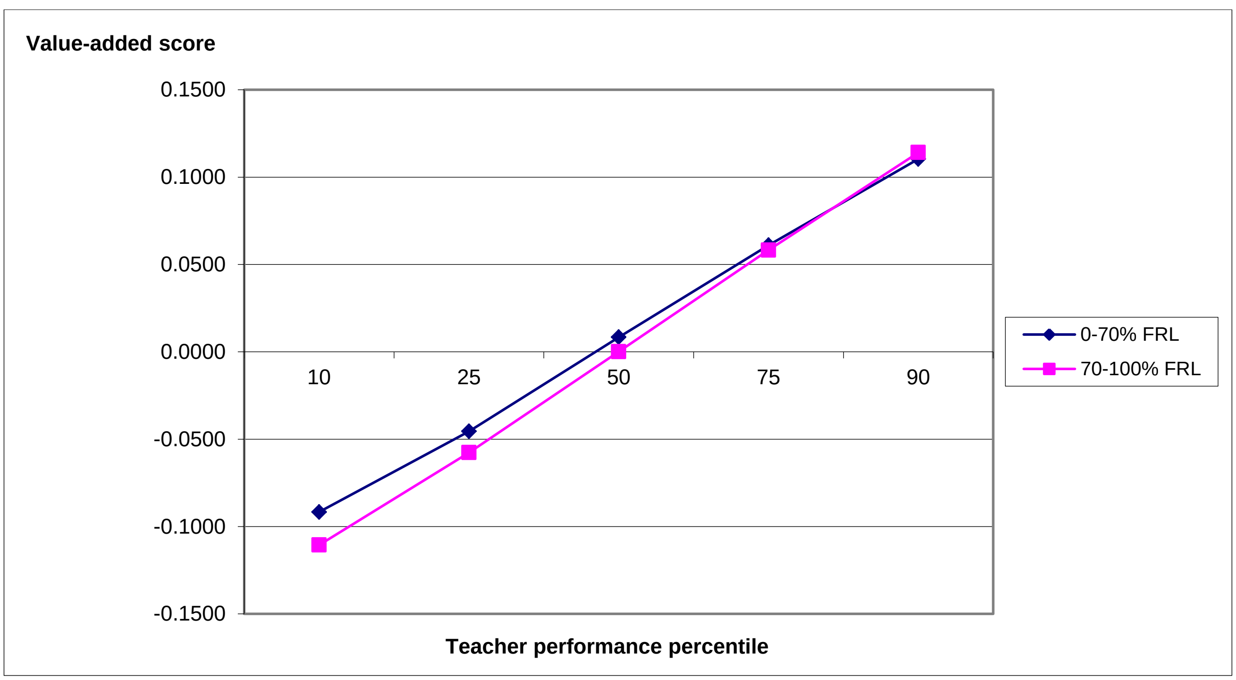


Table 6. Teacher Value Added at Various Percentiles, by state, Subject and School Poverty Category (Teachers with 2 or Fewer Years of Experience)

Subject and teacher performance percentile	Florida			North Carolina		
	0-70% FRL	*70-100% FRL*	*Difference*	*0-70% FRL*	*70-100% FRL*	*Difference*
Math						
10	-0.1790	-0.2100	0.0310**	-0.2203	-0.2066	-0.0137
25	-0.0925	-0.1024	0.0099**	-0.1179	-0.1031	-0.0148
50	0.0036	0.0102	-0.0066	-0.0083	-0.0134	0.0051
75	0.0965	0.1235	-0.0279**	0.1115	0.1256	-0.0142*
90	0.1838	0.2404	-0.0566**	0.2150	0.2242	-0.0092
Reading						
10	-0.1157	-0.1305	0.0148**	-0.1147	-0.1112	-0.0035
25	-0.0563	-0.0640	0.0077*	-0.0635	-0.0574	-0.0061
50	0.0076	0.0063	0.0013	0.0002	0.0101	-0.0099
75	0.0697	0.0794	-0.0097**	0.0551	0.0639	-0.0089**
90	0.1300	0.1586	-0.0286**	0.1071	0.1202	-0.0131*

Note: * denotes statistical significance at the 5% level, ** denotes statistical significance at the 1% level

Table 7. Sources of Predicted Mean Difference in FGLS Estimates of Teacher Effects, by State, Subject and School Poverty Category

Subject, predicted value added, overall difference and sources of difference	Florida		North Carolina	
	0-70% FRL	*70-100% FRL*	*0-70% FRL*	*70-100% FRL*
Math				
Mean predicted value added	0.0137**	0.0281**	0.0093**	-0.0230**
Overall Difference	-0.0144**		0.0323**	
Sources of difference				
Characteristics	0.0021		0.0014	
Marginal effect of characteristics	-0.0180**		0.0295**	
Interaction of char. and marginal effects	0.0015		0.0015	
Reading				
Mean predicted value added	0.0129**	0.0024	0.0088**	-0.0179**
Difference	0.0105**		0.0267**	
Sources of difference				
Characteristics	0.0017*		0.0015*	
Marginal effect of characteristics	0.0061*		0.0243**	
Int. of char. and marginal effects	0.0027**		0.0009	

Note: Difference is * significant at 5%, ** significant at 1%

Table 8. FGLS Estimates of the Determinants of Teacher Value Added, by state, subject and School Poverty Category

Subject and teacher characteristic	Florida		North Carolina	
	0-70% FRL	*70-100% FRL*	*0-70% FRL*	*70-100% FRL*
Math				
3-5 years	0.0058	0.0047	0.0335**	0.0352*
6-12 years	0.0106	0.0130	0.0420**	0.0371*
13-20 years	0.0146*	0.0029	0.0438**	0.0203
21-27 years	-0.0089	0.0098	0.0406**	0.0217
28 or more years	-0.0054	0.0027	0.0363**	0.0284
Graduate degree	0.0105*	-0.0094	-0.0039	0.0093
Regular license	0.0486**	0.0314*	0.0611**	0.0249
Reading				
3-5 years	0.0012	0.0073	0.0278**	0.0277*
6-12 years	0.0129**	0.0044	0.0404**	0.0194
13-20 years	0.0177**	0.0011	0.0434**	0.0293*
21-27 years	0.0205**	0.0055	0.0478**	0.0414**
28 or more years	0.0210**	-0.0032	0.0477**	0.0265
Graduate degree	0.0009	0.0013	-0.0120**	-0.0084
Regular license	0.0506**	0.0309**	0.0494**	0.0240

Note: Coefficient is * significant at 5%, ** significant at 1%

Table 9. Mean Teacher Value Added, by Teacher Mobility, State, Subject and School Poverty Category

	Florida		North Carolina	
Subject and teacher mobility	*0-70% FRL*	*70-100% FRL*	*0-70% FRL*	*70-100% FRL*
Math				
Leaves at the end of 1st year	-0.0910	-0.0231	-0.0876	-0.1135
Stays beyond 1st year	-0.0648	-0.0607	-0.0642	-0.0466
Difference	**-0.0262**	**0.0376**	**-0.0234**	**-0.0669***
Leaves at the end of 2nd year (year 1 efficiency)	0.1411	0.1583	-0.0467	-0.0168
Stays beyond 2nd year (year 1 efficiency)	-0.2382	-0.1096	-0.0596	-0.0414
Difference	**0.3793****	**0.2679****	**0.0129**	**0.0246**
Leaves at the end of 2nd year (year 2 efficiency)	0.0796	0.1243	-0.0340	0.0393
Stays beyond 2nd year (year 2 efficiency)	-0.0064	0.1125	-0.0257	-0.0238
Difference	**0.0860**	**0.0118**	**-0.0083**	**0.0631**
Reading				
Leaves at the end of 1st year	-0.0570	-0.0738	-0.0410	-0.0636
Stays beyond 1st year	0.0496	0.0284	-0.0309	-0.0197
Difference	**-0.1066****	**-0.1021****	**-0.0102**	**-0.0439****
Leaves at the end of 2nd year (year 1 efficiency)	0.0489	-0.0058	-0.0377	-0.0045
Stays beyond 2nd year (year 1 efficiency)	-0.1351	-0.1563	-0.0291	-0.0222
Difference	**0.1841****	**0.1504****	**-0.0086**	**0.0176**
Leaves at the end of 2nd year (year 2 efficiency)	-0.0295	-0.0058	-0.0086	0.0503
Stays beyond 2nd year (year 2 efficiency)	0.0152	0.0809	-0.0097	-0.0166
Difference	**-0.0446**	**-0.0867****	**0.0011**	**0.0669**

Note: * denotes statistical significance at the 10% level, ** denotes statistical significance at the 5% level

Table 10. Mean Teacher Value Added for Those Who Taught in both High- and Low-Poverty Settings, by State, Subject and the Direction of Switch

	Florida			North Carolina		
Subject and direction of school change	*Pre-switch performance*	*Post-switch performance*	*Difference*	*Pre-switch performance*	*Post-switch performance*	*Difference*
Math						
Switch from high to low-poverty schools	0.0198	0.0219	0.0022	0.0116	-0.0097	-0.0213
Switch from low to high-poverty schools	0.0258	0.0260	0.0002	0.0138	-0.0123	-0.0261
Difference (High-to-Low vs. Low-to-High)	-0.0060	-0.0040		-0.0022	0.0027	
All Switchers (high-poverty vs. low)			-0.0011			-0.0007
Reading						
Switch from high to low-poverty schools	0.0057	0.0106	0.0049	-0.0044	0.0009	0.0053
Switch from low to high-poverty schools	0.0059	0.0062	0.0003	-0.0229	-0.0068	0.0161
Difference (High-to-Low vs. Low-to-High)	-0.0002	0.0044		0.0185	0.0077	
All Switchers (high-poverty vs. low)			-0.0008			0.0046

Note: * denotes statistical significance at the 10% level, ** denotes statistical significance at 5% level.

Appendix

Appendix Table A1. Summary of Florida and North Carolina Variables Used in Value-Added Model

FL Variable	NC Variable	Description
First Stage Analysis – Value-Added Model		
Student Time-Varying Characteristics		
strucmove	struc_chg	Student attended a different school in the previous year and 30 percent or more of prior-year school mates made the same move
nonstrucmove	schlchange	Student attended a different school in the previous year and few than 30 percent of prior-year schoolmates made the same move
repeat	repeating	Student is repeating a grade [Note: interacted with grade level]
Classroom Peer Characteristics		
math_class_size	clsize	Class size
peer_nonwhite	pct_nonwt	Fraction of classmates who are not white
peer_male	pct_male	Fraction of classmates who are male
Student Time-Invariant (or Quasi-Time-Invariant) Characteristics		
male	male	Student is male
black	black	Student is black
hispanic	hisp	Student is Hispanic
othernw	other_race	Student is neither black, Hispanic or white
age_g3	age_g3	Student's age (in months) at first enrollment in grade 3 (as of March of relevant school year)
lep	limeng	Student is in a limited English proficiency program
gifted	gifted	Student is in a program for "gifted" students
speech_lang	speech_lang	Speech/language impaired students
learn_dis	learn_dis	Specific learning disabled students
mental_dis	mental_dis	Mentally disabled students
physical_dis	physical_dis	Physically disabled students
emotional_dis	emotional_dis	Emotionally disabled students
other_dis	other_dis	Other disabilities
ayfrlnch	schlunch	Student receives free/reduced-price lunch
School Time-Varying Variables		
p_exper	—	Administrative experience of principal
p_exper_sqd	—	Administrative experience of principal squared
new_princ_at_school	—	Indicator that principal is in first year as principal at current school

Second Stage Analysis - Teacher Variables

ever_nbpts_cert	everNBCT	Teacher who has or will have received NBPTS certification during the sample period
exper_01_02	exper1_2	Teacher with 1-2 years of experience
exper_03_05	exper3_5	Teacher with 3-5 years of experience
exper_06_12	exper6_12	Teacher with 6-12 years of experience
exper_13_20	exper13_20	Teacher with 13-20 years of experience
exper_21_27	exper21_27	Teacher with 21-27 years of experience
exper_28pls	exper28up	Teacher with 28 or more years of experience
adv_degree	advdegree	Teacher holds a post-baccalaureate degree
prof_cert	reg_lic	Teacher holds a regular (non-temporary) state teaching license
	praxis	Teacher test score. Praxis scores for NC. General knowledge exam scores for FL

Appendix Table A2. Means and Standard Deviations of Teacher Value Added, by State, School Poverty Category, Subject and Model

	Florida			North Carolina		
Subject and model	*0-30% FRL*	*70-100% FRL*	*Difference*	*0-30% FRL*	*70-100% FRL*	*Difference*
Means						
Math						
FE estimates	0.0337	0.0220	0.0117**	0.0063	-0.0078	0.0141*
Student covariate estimates	0.0148	0.0159	-0.0011	0.0277	-0.0198	0.0475**
Reading						
FE estimates	0.0012	-0.0030	0.0042	0.0045	0.0106	0.0151**
Student covariate estimates	0.0177	0.0008	0.0169**	0.0183	-0.0109	0.0292**
Standard deviations						
Math						
FE estimates	0.1433	0.1979	-0.0546**	0.2070	0.1996	0.0074
Student covariate estimates	0.1361	0.1723	-0.0362**	0.1684	0.1816	-0.0133**
Reading						
FE estimates	0.0890	0.1119	-0.0229**	0.0830	0.0918	-0.0088**
Student covariate estimates	0.0877	0.1015	-0.0138**	0.0826	0.0952	-0.0126**

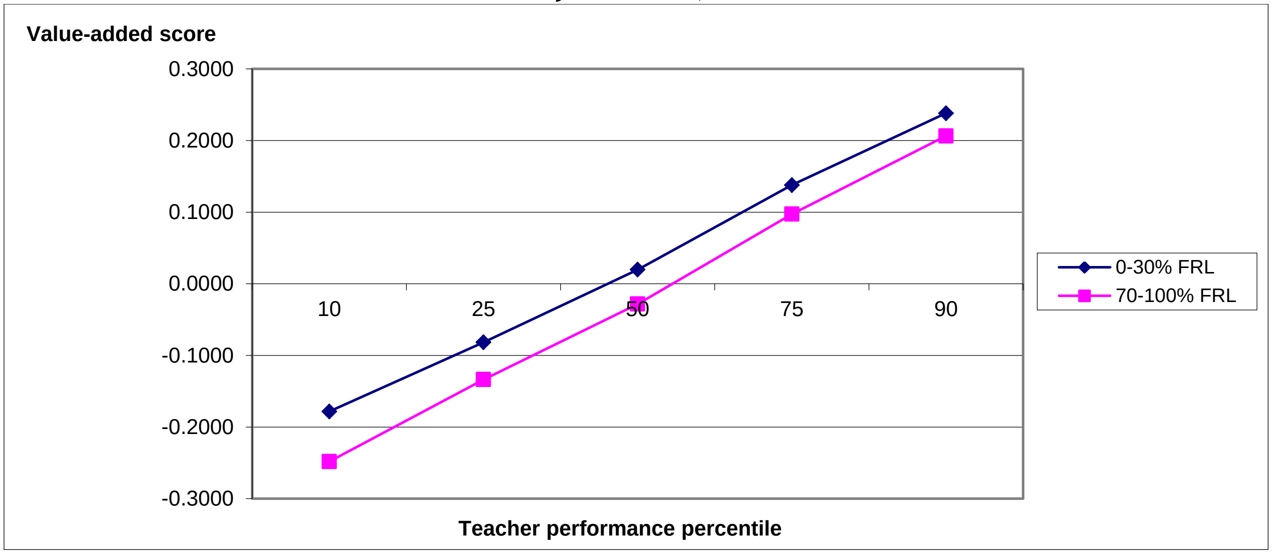
Note: Differences between Title I and Non-Title I teacher effects (t-test) / standard deviations (F-test) are * significant at 5%, ** significant at 1%

Appendix Table A3. Teacher Value Added at Various Percentiles, by State, Subject and School Poverty Category

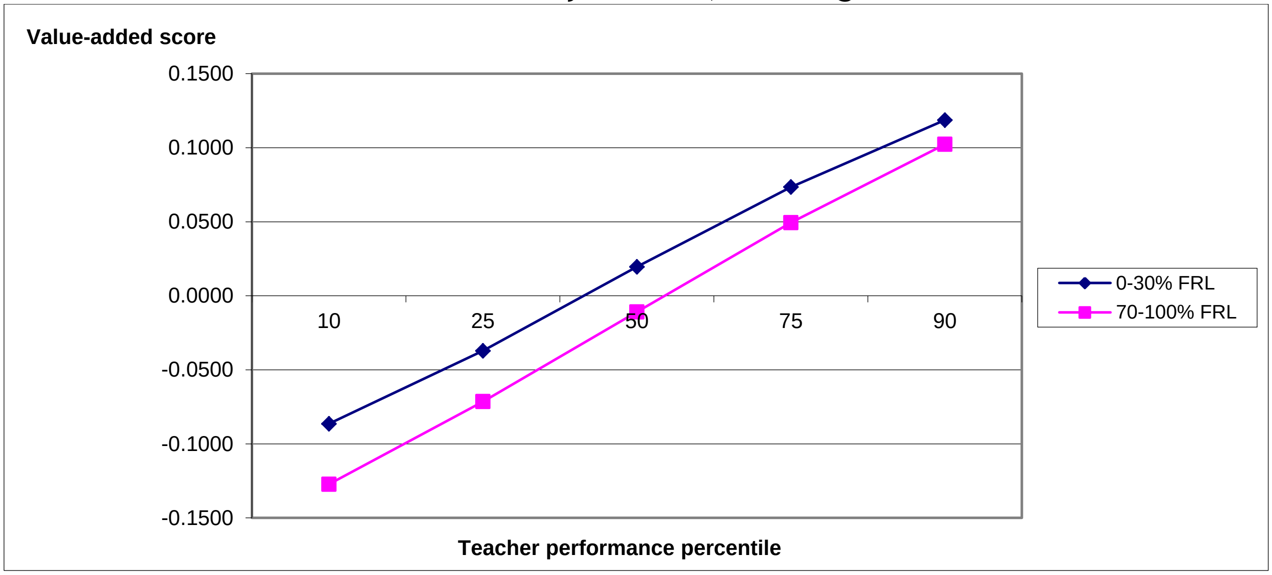
	Florida			North Carolina		
Subject and teacher performance percentile	*0-30% FRL*	*70-100% FRL*	*Difference*	*0-30% FRL*	*70-100% FRL*	*Difference*
Math						
10	-0.1546	-0.1935	0.0389**	-0.1782	-0.2481	0.0699**
25	-0.0770	-0.0921	0.0151**	-0.0817	-0.1337	0.0520**
50	0.0127	0.0095	0.0032	0.0198	-0.0283	0.0480**
75	0.1031	0.1238	-0.0207**	0.1378	0.0972	0.0406**
90	0.1922	0.2276	-0.0354**	0.2382	0.2065	0.0317**
Reading						
10	-0.0930	-0.1186	0.0256**	-0.0864	-0.1273	0.0409**
25	-0.0417	-0.0620	0.0203**	-0.0372	-0.0715	0.0343**
50	0.0179	-0.0007	0.0186**	0.0196	-0.0109	0.0305**
75	0.0775	0.0604	0.0171**	0.0735	0.0494	0.0241**
90	0.1300	0.1182	0.0118*	0.1187	0.1023	0.0164**

Note: * denotes statistical significance at the 5% level, and ** denotes statistical significance at the 1% level

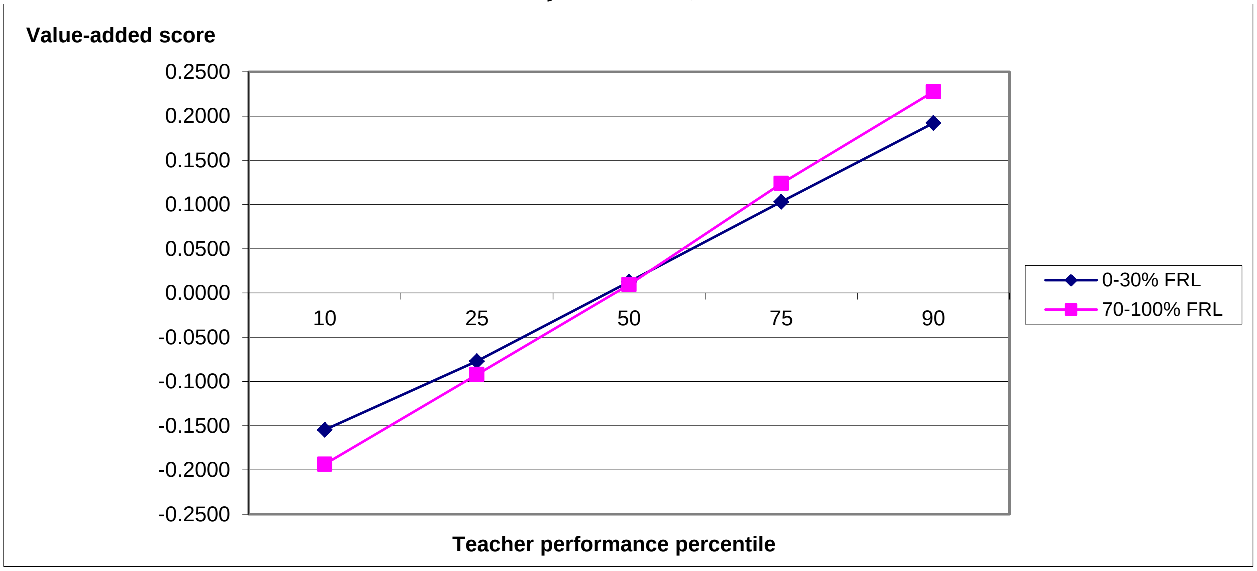
Appendix Figure A1A. Teacher Value Added at Various Percentiles for High-and Lower-Poverty Schools, Math Teachers in North Carolina



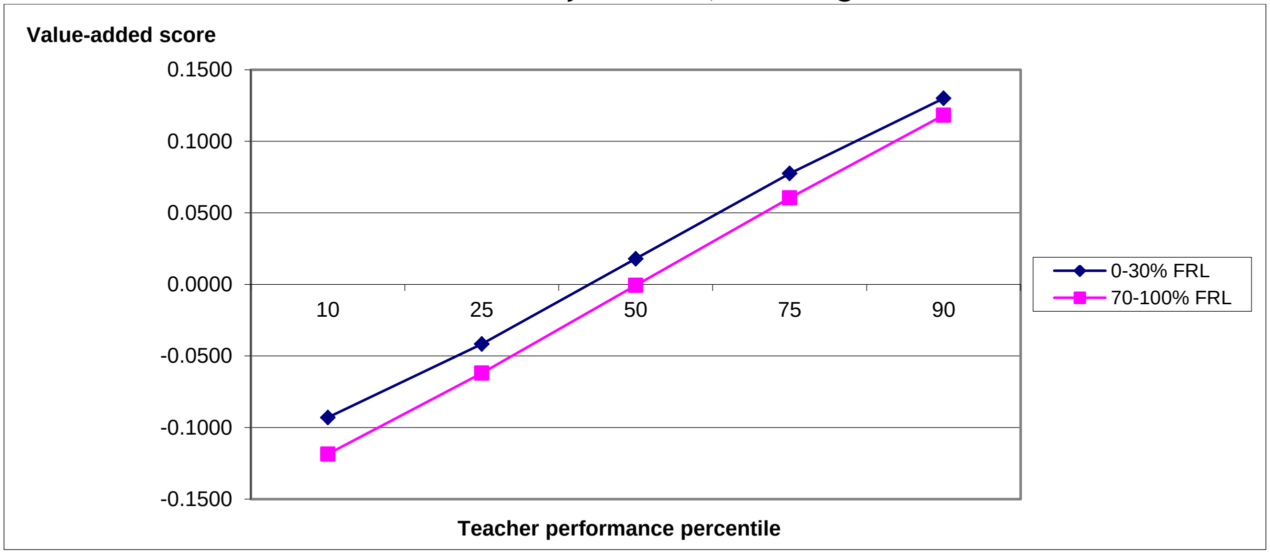
Appendix Figure A1B. Teacher Value Added at Various Percentiles for High- and Lower-Poverty Schools, Reading Teachers in North Carolina



Appendix Figure A2A. Teacher Value Added at Various Percentiles for High-and Lower-Poverty Schools, Math Teachers in Florida



Appendix Figure A2B. Teacher Value Added at Various Percentiles for High- and Lower-Poverty Schools, Reading Teachers in Florida



Appendix Table A4. Sources of Predicted Mean Difference in FGLS Estimates of Teacher Effects, by State, Subject and School Poverty Category

Subject, predicted value-added, overall difference and sources of difference	Florida		North Carolina	
	0-30% FRL	*70-100% FRL*	*0-30% FRL*	*70-100% FRL*
Math				
Mean predicted value-added	0.0161**	0.0204**	0.0337**	-0.0254**
Overall Difference	-0.0043		0.0591**	
Sources of difference				
Characteristics	0.0030		0.0028*	
Marginal effect of characteristics	-0.0088		0.0548**	
Interaction of characteristics and marginal effects	0.0015		0.0015	
Reading				
Mean predicted value-added	0.0267**	-0.0003	0.0315**	-0.0215**
Difference	0.0271**		0.0530**	
Sources of difference				
Characteristics	0.0028*		0.0022	
Marginal effect of characteristics	0.0212**		0.0506**	
Interaction of characteristics and marginal effects	0.0031		0.0002	

Note: Difference is * significant at 5%, ** significant at 1%

Appendix Table A5. FGLS Estimates of the Determinants of Teacher Value Added, by State, Subject and School Poverty Category

Subject and teacher characteristic	Florida		North Carolina	
	0-30% FRL	*70-100% FRL*	*0-30% FRL*	*70-100% FRL*
Math				
3-5 years	0.0074	0.0028	0.0219	0.0351*
6-12 years	0.0132	0.0181	0.0361**	0.0370*
13-20 years	0.0116	0.0074	0.0288	0.0202
21-27 years	-0.0009	0.0114	0.0422**	0.0218
28 or more years	-0.0295	0.0084	0.0486**	0.0284
Graduate degree	0.0061	-0.0089	-0.0056	0.0093
Regular license	0.0346	0.0291*	0.0875*	0.0254
Reading				
3-5 years	-0.0091	0.0105	0.0278**	0.0276*
6-12 years	0.0173	0.0071	0.0360**	0.0195
13-20 years	0.0182	0.0024	0.0444**	0.0296*
21-27 years	0.0229	0.0059	0.0476**	0.0418**
28 or more years	0.0364**	-0.0053	0.0471**	0.0265
Graduate degree	0.0003	0.0044	-0.0056	-0.0084
Regular license	0.0545**	0.0327**	0.0112	0.0246

Note: Coefficient is * significant at 5%, ** significant at 1%

CALDER