NBER WORKING PAPER SERIES

HOW AND WHY DO TEACHER CREDENTIALS MATTER FOR STUDENT
ACHIEVEMENT?

Charles T. Clotfelter
Helen F. Ladd
Jacob L. Vigdor

How and Why do Teacher Credentials Matter for Student Achievement?
Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor
NBER Working Paper No. 12828
January 2007
JEL No. I21

ABSTRACT

Education researchers and policy makers agree that teachers differ in terms of quality and that quality matters for student achievement. Despite prodigious amounts of research, however, debate still persists about the causal relationship between specific teacher credentials and student achievement. In this paper, we use a rich administrative data set from North Carolina to explore a range of questions related to the relationship between teacher characteristics and credentials on the one hand and student achievement on the other. Though the basic questions underlying this research are not new - and, indeed, have been explored in many papers over the years within the rubric of the "education production function" - the availability of data on all teachers and students in North Carolina over a ten-year period allows us to explore them in more detail and with far more confidence than has been possible in previous studies. We conclude that a teacher's experience, test scores and regular licensure all have positive effects on student achievement, with larger effects for math than for reading. Taken together the various teacher credentials exhibit quite large effects on math achievement, whether compared to the effects of changes in class size or to the socio-economics characteristics of students, as measured, for example, by the education level of their parents.

Charles T. Clotfelter
Sanford Institute of Public Policy
Box 90245
Duke University
Durham, NC 27708
and NBER
charles.clotfelter@duke.edu

Helen F. Ladd
214A Terry Sanford Institute
Duke University
Box 90243
Durham, NC 27708
hladd@pps.duke.edu

Jacob L. Vigdor
Terry Sanford Institute of Public Policy
Duke University
Durham, NC 27708
and NBER
jacob.vigdor@duke.edu

Education researchers and policy makers agree that teachers differ in terms of quality and that quality matters for student achievement. The problem is that it is difficult to measure teacher quality. One strategy used in a number of recent empirical studies is to use teacher fixed effects in an equation explaining variation in student achievement. The variation in these estimated fixed effects is then interpreted as the variation in teacher "quality." Emerging from such studies is the general consensus that a one standard deviation difference in the quality of teachers as measured in this way generates about a 0.10 standard deviation in achievement in math and a slightly smaller effect in reading. (Rivkin, Hanushek and Kain, 2005; Rockoff 2004, and Aronson, Barrow and Sanders, 2003). Of more direct interest for policy makers, however, would be knowledge about which specific characteristics or credentials of teachers are most predictive of student achievement. Such information would enable policy makers concerned about either the overall level or the distribution of student achievement to design better licensure, salary and other teacher policies.

Despite prodigious amounts of research, debate still rages about the effect of specific teacher credentials on student achievement. Recently some researchers have tried to explain the variation in teacher "quality" (measured as just described by the variation in estimated intercepts) by variation in teacher credentials, but data limitations have often limited the analysis to only a few teacher credentials, such as teacher experience. The more traditional approach is simply to estimate in one step the relationship between student achievement and teacher credentials. Many such studies focus primarily on credentials such as years of experience or graduate degrees that have budgetary costs (see meta analyses of such studies by Hanushek (1997) and Hedges, Laine, and Greenwald,

1994). Others focus on particular credentials such as licenses or National Board Certification (Goldhaber and Brewer, 2004, Goldhaber and Anthony, 2005). We pursue this more direct approach in this paper.

Specifically, we use a rich administrative data set from North Carolina to explore a range of questions related to the relationship between teacher characteristics and credentials on the one hand and student achievement on the other. Though the basic questions underlying this research are not new -- and, indeed, have been explored in many papers over the years within the rubric of the "education production function" -- the availability of data on all teachers and students in North Carolina over a ten-year period allows us to explore them in more detail and with far more confidence than has been possible in previous studies. Particularly relevant to this study is that North Carolina has been testing all students in reading and math from grades three to eight since the early 1990s and that the tests are closely linked to the state's Standard Course of Study. Thus, students in North Carolina are being tested on the knowledge and skills that the state wants them to know. Further, the existence of a relatively sophisticated test-based accountability system gives teachers strong incentives to teach them that material.

The teacher credentials in which we are most interested are those that can be affected in one way or another by policy, either through incentives to induce teachers to change their credentials (such as by offering higher pay to teachers with master's degrees), by setting the rules on who can become a teacher (such as by licensing requirements) or by formal or informal decisions that determine how teachers with stronger or weaker credentials are distributed among schools. Ultimately, we are interested in determining what teacher-related policies are most likely to promote student

achievement and to reduce achievement gaps between various groups of students (such as minority and white students or low-income and more affluent students). Given that spending on teachers constitutes such a large share of education budgets and that effective teachers are currently so unevenly distributed across classrooms, additional research on the link between teacher credentials is both valuable and important for education policy.

This paper builds on our previous cross-sectional research on teacher credentials and characteristics (Clotfelter, Ladd and Vigdor, 2006). In that study, we used a portion of the same North Carolina data to estimate a model of fifth grade student achievement in 2002, paying particular attention to the sorting of teachers and students among schools and classrooms. Emerging from that research was abundant evidence that teachers with stronger credentials tend to teach in schools with more advantaged and higher performing students and, to a far lesser extent, that similar matching occurs across classrooms within schools. As we document in that paper, without careful attention to this nonrandom matching of teachers to students, the estimated effects of teacher credentials on student achievement are likely to be biased upward. A major purpose of our earlier paper was to show how a large administrative data set can be used to overcome such statistical problems in cross sectional analysis. After making the appropriate statistical adjustments we concluded that the two credentials most consistently linked to student achievement were teacher experience -- with the largest effects emerging within the first few years of teaching -- and teacher test scores.

This new research differs in its use of longitudinal data covering all the North Carolina students in grades 3,4, and 5 in years 1995-2004 for whom we can identify their

teachers of math and reading. As a longitudinal study, this paper builds most closely on Jonah Rockoff's (2004) study of the impact of individual teachers for three districts in New Jersey and Hanushek, Kain , O'Brien and Rivkin's (HKOR) (2005) study of the market for teacher quality that is based on data for one district in Texas. We are working with richer and larger data sets than Rockoff and HKOR since our data cover a whole state. Moreover, we have better achievement measures than in HKOR, and richer sets of teacher credentials and student characteristics than either Rockoff or HKOR. We believe that the North Carolina data set is the only statewide data set that permits the matching of student test scores to specific teachers.

Our basic approach is to estimate a variety of models beginning with a standard value-added model in which student achievement in the current year is estimated as a function of the student's achievement in the prior year, and student, teacher, and classroom characteristics in the current year, and progressing to models that include student fixed effects. These fixed effects, which can be included only when longitudinal data are available, provide powerful protection against the left-out variable bias that typically plagues research of this type. By controlling for the time-invariant characteristics of students that may affect achievement – both those that are observed and those that are unobserved – the fixed effect approach eliminates many of the statistical problems that arise in a linear model because of the non-random matching of teachers to students.

The reliance on longitudinal data is also advantageous in that it permits us to explore in some detail the mechanisms through which teacher credentials exert their impacts. As one example, in our previous cross sectional work we were not able to

6

determine if our finding of a positive relationship between years of teaching experience and student achievement was the result of individual teachers becoming more effective as they gained experience or of higher rates of attrition for lower-quality than for higher quality teachers. With longitudinal data, we are able to shed additional light on that issue. Other examples include the potential for more detailed analysis of the effects of master's degrees, licensure policies, and National Board Certification.

The rest of the paper is structured as follows. In section II we present our conceptual framework. In section III we describe the North Carolina data with particular attention to the challenge of identifying the correct math and reading teachers for each student. Section IV presents our results for five basic models and section V puts the estimated magnitudes into perspective. In section VI we explore a number of elaborations of the basic model that permit us to generate greater insights into the mechanism through which the various credentials affect student achievement. The paper ends with a brief concluding discussion.

I.   Empirical framework

The starting point for our analysis is the observation that education is a cumulative process. In the context of a very simple model in which the only component of education that matters is teacher quality and in which all other determinants of achievement such as student background, ability, and motivation are set aside, we can write

$$A_{it} = f(TQ_{it}, TQ_{i\,t-1}, \ldots) + \text{error}$$

where A refers to student achievement and TQ refers to teacher quality. This equation expresses student i's achievement in year t as a function of the quality of her teacher in that year and in all previous school years plus a random error.

Two additional assumptions permit us to transform this relationship into one that can be used to estimate the effects of teacher quality in year t on the student's achievement in year t, controlling for the effects of teacher quality in all prior years. One assumption is that the effect of teacher quality on a student's achievement in the contemporaneous year is constant across years and that the relationship is linear. The second is that any decay in student achievement, or knowledge, from one year to the next occurs at a constant rate. As a result, the rate at which a student's knowledge persists from one year to the next is also constant. Letting β be the effect of TQ and α the rate at which knowledge persists, we can rewrite equation 1 as

$$A_{it} = \beta TQ_{it} + \alpha \beta TQ_{it-1} + \alpha^2 \beta TQ_{i\,t-1} + \alpha^3 \beta TQ_{it-2} + \ldots + error_{it}$$

and, after rearranging terms, as

$$A_{it} = \beta TQ_{it} + \alpha(\beta TQ_{it-1} + \alpha \beta TQ_{it-2} + \alpha \beta TQ_{it-3} + \ldots) + error_{it}$$

Given that the expression within the parenthesis is simply A_{it-1}, we end up with

$$A_{it} = \beta TQ_{it} + \alpha A_{it-1} + error_{it}.$$

Thus, the effects on current achievement of the student's prior teachers are captured by the lagged achievement term. If the prior year achievement has no effect on current achievement, the persistence coefficient, α, would be zero.

Models of this form, which are typically referred to as value-added models, are commonly used in the literature to estimate β, namely the effect of current teachers on current achievement. Their popularity comes largely from their simplicity and intuitive appeal. Logically, it makes sense that one would want to control statistically for the achievement, or knowledge, that the student brings to the classroom at the beginning of

the year when estimating the effect of her current teacher. In addition, the value-added

model is flexible in that it does not impose a specific assumption about the rate at which

knowledge persists over time; instead it allows that rate to be estimated. Nonetheless, as

we discuss further below, such models are still subject to various statistical concerns that

could lead to upward or downward estimates of β, the key coefficient of interest.

Our empirical strategy in this paper is to start with a more elaborate form of this

simple value added model (see model 1 below). We then modify the model in various

ways in an effort to address some of the statistical concerns. One modification is to add

fixed effects, either for schools or for students. Another is to change the dependent

variable from the level of achievement to the gain in achievement, defined as $A_{it} - A_{it-1}$.

We report results for the following five models.

Model 1. Value- added model with no fixed effects.

We start with a more elaborate form of the basic value-added model. Thus, in

model 1 the achievement of a student in year t is specified as a function of the student's

prior achievement; teacher characteristics (both those that are time invariant and those

that vary over time), class room characteristics; and student characteristics (both those

that are time invariant and those that vary over time).[1] The relevant variables or vectors

of variables are defined as follows.

> A_{it} is achievement of student i in year t as measured by a normalized test score in
> reading or math.

> A_{it-1} is achievement of the ith student in the prior year.[2]

---

[1] For thoughtful discussions of value added models see Todd and Wolpin, 2003) and Boardman and
Murnane, 1979.

[2] For simplicity, the lagged achie   vement term refers here, and also in the basic regressions, to the same
subject as the dependent variable. As we note below we also estimated some regressions in which we
included lagged achievement terms for both math and reading as well as squared terms for each of them, on
the ground that prior achievement in both math and reading could affect current year achievement in either
subject.

TCF is a vector of teacher characteristics, such as the teacher's race and gender, that are fixed over time for any specific teacher.

TCV_t is a vector of teacher characteristics that vary over time, including, for example, years of teaching experience, attainment of higher degrees, or attainment of a particular type of license.

C_t is a vector of classroom characteristics that vary depending on the student's classroom each year. These include class size and characteristics of peers.

SCF_i is a vector of measurable student characteristics that are fixed over time, such as a student's race, gender, and age in grade 3. For reasons explained below we also include in this category the education level of the student's parents and whether the child is eligible for a subsidized lunch even though such characteristics could potentially change over time. In some of our models these variables are replaced with student fixed effects that capture the effects not only of the time-invariant student characteristics that can be observed and measured but also those that are unobserved.

SCV_{it} = a vector of student characteristics that vary over time. These include indicator variables for thing such as grade repetition or movement to a new school.

u_{it} is an error term

Specifically, we estimate the following value added model:

$$A_{it} = \alpha A_{it-1} + \beta_1 SCF_i + \beta_2 SCV_{it} + \sum_j [\beta_3 TCF_t^{(j)} + \beta_4 TCV_t^{(j)} + \beta_5 C_t^{(j)}] D_{it}^{(j)}$$

(1)

In addition to the variables defined above, $D_{it}^{(j)}$ is an indicator variable for whether the

student had the jth teacher in year t. The coefficients β_1 to β_5 are vectors rather than

individual parameters, and α is the persistence parameter. Of most interest are the

vectors of coefficients that relate to the teachers, that is, the parameter vectors denoted by

β_3 and β_4.

.   The estimates of the teacher coefficients are subject to at least two sources of bias.

First, and most important, the coefficients will be biased if teachers with stronger

qualifications are matched in some systematic way with students exhibiting characteristics that are not fully controlled for in the model. To the extent that there is positive matching in the sense that teachers with stronger qualifications are matched to students who are educationally more advantaged along dimensions that are hard to measure, the coefficients would be biased upward. This bias arises because too much of the achievement of the high achieving students would be attributed to the teacher variables rather than to the unobserved characteristics of the students.

Second, the inclusion of the lagged achievement variable on the right hand side of the equation is a problem, both because the variable is likely to be noisy and because any correlation of achievement over time would render the variable endogenous. A noisy variable is one in which the random error is large relative to the true signal. The lagged test score is likely to be noisy both because the test itself is an imperfect means of gauging student knowledge and because a student's performance on any one test is subject to the vagaries of the testing environment, such as how well the student feels on the day of the test or how crowded the room is. In general, measurement error of this type generates a downward bias in the coefficient of the variable itself and also biases the coefficients of other variables in hard–to-predict ways.[3] The correlation between the lagged achievement variable and the error term in the current achievement equation introduces additional bias of undetermined sign. The standard solution for both problems is to use the statistical technique of instrumental variables, provided a good instrument is available. Though a twice lagged achievement variable could potentially serve that

---

[3] Though the dependent variable is subject to measurement error as well, that error simply shows up in the error term and ins not a source of bias.

purpose, we have chosen not to use that approach in this paper given the short length of our data panels..

Model 2. Value-added model with school fixed effects.

In prior research we have documented that teachers are indeed positively matched to students in North Carolina. (Clotfelter, Ladd and Vigdor, 2006). Moreover, we find that most of this positive matching occurs at the school, rather than the classroom, level. This pattern of positive matching across schools largely reflects the use of a single salary schedule across schools within each district, that teachers value working conditions -- which are determined in part by the characteristics of their students -- as well as their salaries, and a typical internal transfer policy that favors more highly qualified teachers. Given that more advantaged students are often deemed easier and more rewarding to teach than those from disadvantaged backgrounds, highly qualified teachers have an incentive to move away from schools with large proportions of disadvantaged students in favor of schools with more advantaged and easier-to-educate students.

Model 2 directly addresses the upward bias of the teacher coefficients that arises from this positive matching by the inclusion of school fixed effects. In essence the model includes 0-1 indicator variables for each school, and thereby controls for the unchanging characteristics of each school, both those characteristics that are observed and measurable and those that are not. The inclusion of the school fixed effects means that the effects on student achievement of the teacher qualification variables are identified only by the variation across classrooms within each school, and not at all by how teachers are distributed across schools. As a result, any upward bias in the estimates caused by the

positive matching of teachers and students across schools is eliminated provided the characteristics of schools enter the model in a linear manner.

Though model 2 goes a long way toward eliminating the bias associated with the nonrandom matching of teachers and students across schools, it does not address the bias that may arise from any nonrandom matching of students to teachers across classrooms within schools. Our prior research suggests that any nonrandom matching of this type is likely to be small. Nonetheless some bias may well remain. In addition, this model is still subject to any bias caused by the correlation between the lagged achievement term and the error term. [4]

Model 3. Gains model, with school fixed effects

One potential solution to the statistical problems associated with having the lagged dependent variable on the right side of the equation is to move that variable to the left side and to use gains in achievement, that is $A_{it} - A_{it-1}$, as the dependent variable. This gains specification would be equivalent to the previous model only if there were no decay in knowledge from year to year, that is, only if the persistence coefficient in the value-added model were equal to 1.

In the more likely case in which the true coefficient on the lagged term is less than one, so that the decay rate $(1-\alpha)$ is positive, the gains model is misspecified and, under reasonable assumptions, will lead to downward biased estimates of the effects of teacher

---

[4] As emphasized by Todd and Wolpin (2003), implicit in this general form of the value-added specification is the assumption that the effects of the input coefficients decay at the same rate for all years. In addition, unless the ability endowment also decays at the same rate at the input effects, the lagged term (which represents baseline achievement) will be correlated with the error time and the coefficients cannot be consistently estimated with OLS. As Todd and Wolpin state : "Any value-added model that admits to the presence of unobserved endowments must also recognize that baseline achievement will then logically be endogenous." Page F 21. Moreover, that endogeneity will affect not only the coefficient of the lagged term but also the estimates of all the input effects.

qualifications on current achievement. The logic here is as follows. Missing from the right side of the gains equation is the lagged achievement variable, which in this case would have entered with a coefficient of $-(1-\alpha)$ to capture the loss in achievement due to decay. Because this variable is missing, the other variables in the equation, such as the teacher qualification variables, will pick up some of the negative effect to the extent that qualifications of teachers facing particular students are positively correlated from one grade to the next.

Thus, the gains model solves one problem but introduces another. On balance, we prefer model 2 based on our intuition that the imposition of an incorrect assumption about the rate of decay is likely to create more bias than does the presence of the lagged achievement term on the right hand side of the equation. At the same time, given the statistical problems caused by the inclusion of the lagged achievement term in that model, we cannot conclude with certainty that model 2 is preferred.

Model 4. Levels model with student fixed effects (and no lagged achievement variable)

None of the first three models makes full use of the longitudinal data that are available for this study. The next two models do so by including student fixed effects, first in a model with student achievement as the dependent variable (model 4) and second in a model with gains in achievement as the dependent variable (model 5). Though neither model generates unbiased coefficients of the teacher variables, under reasonable assumptions the coefficients that emerge from the two models should bracket the true coefficients.

Like model 2, the dependent variable in model 4 is the level of student achievement. This model differs in that it includes student, rather school, fixed effects

and excludes the lagged achievement term. Compared to model 2, this model more effectively eliminates any bias associated with the nonrandom matching of teachers and students. This conclusion follows because the presence of student fixed effects means that the only variation used to estimate the coefficients of interest is variation within, not across, individual students. Thus, the question is not whether one student who is taught by a teacher with a given qualification achieves on average at a higher level than a different student facing a teacher with a lower qualification but rather whether any single student does better or worse when she is taught by a teacher with the higher qualification than when that same student is taught by a teacher with a lower qualification. As a result of these within-students comparisons, any nonrandom matching of teachers and students either across schools or across classrooms within schools becomes irrelevant, provided the effects are linear.

Ideally, it would be useful to include lagged achievement as an explanatory variable is such a model but that strategy would not generate sensible results in the context of our data. As described below, we have data on many cohorts of students but for each student we have only three years of test scores, typically those for grades 3, 4 and 5. The shortness of each panel renders it undesirable to include the lagged dependent variable along with student fixed effects, given that the latter picks up each student's average achievement over the three-year time span. The problem is that the only variation remaining in the lagged achievement term variable for each student is now the variation around the average, a large portion of which in a short panel is likely to be random error. Thus the problem of excessive noise relative to true signal becomes acute in this case. Moreover, whatever variation remains continues to be correlated with the error term of

the model. In the context of longer panels, it may be possible to address these problems

with the use of twice lagged achievement as an instrumental variable. We do not have the

luxury of estimating such a model in this case.

The fact that the lagged term is missing from the equation leads to downward

biased estimates of the relevant coefficients, regardless of how the teacher characteristics

for specific students are correlated across grades.[5] Only if knowledge did not persist at all

from one year to the next, would this model generate unbiased estimates. Note that the

more that knowledge persists from one year to the next (that is, the larger is α), the larger

will be the downward bias in the coefficients of interest.

Model 5. Gains model with student fixed effects.

As discussed above, we can incorporate the lagged achievement term directly by

respecifying the dependent variable as the gains in achievement rather than the levels of

achievement. For model 5, we follow that strategy, but in contrast to model 3, include

student, rather than school, fixed effects.

In stark contrast to the gains model without student fixed effects (model 3) ,

model 5 generates upward biased estimates of the effects of teacher characteristics. In a

simple model, the size of the bias would be proportional to the decay rate $(1-\alpha)/2$ and

would not depend on how teacher qualifications are correlated over time.[6] When there is

no decay (so that $\alpha = 1$), this expression equals 0 and there is no bias. The greater is the

rate of decay, the more the coefficients of interest will be biased upward. In the extreme

case in which there is no persistence of knowledge from one year to the next, (so that α

---

[5] The discussion of bias in this model and in the next draws heavily on the insights from a note by Steven
Rivkin (2006) and an application to the effects on achievement of school racial composition (Hanushek,
Kain and Rivkin (2006).
[6] See Rivkin (2006), equation 10, p.7.

= 0), the estimated coefficients from the gains model with student fixed effects could be too large by half.

The following table summarizes the five models. Our preferred models are 4 and 5 because they include student fixed effects and hence protect again any bias associated with nonrandom matching of teachers and students across classrooms in the context of a linear model. Though neither of the preferred models generates unbiased estimates of the effects of teacher qualifications, we are quite confident that together they will bracket the true coefficients.

Summary table

	Dependent variable (Achievement)	Lagged achievement included?	Type of fixed effect	Likely direction of bias of effects of teacher credentials
Model 1	Levels	yes	none	upward
Model 2	Levels	yes	school	unclear, but small
Model 3	Gains	no	school	downward
Model 4	Levels	no	student	downward
Model 5	Gains	no	student	upward

II. The North Carolina Data.

The data we use for this study are derived from administrative records maintained by the North Carolina Education Research Data Center, housed at Duke University. We link several different sets of records to form the database used for this analysis. Student information, including race, gender, participation in the federal free and reduced price lunch subsidy program, education level of the parents, and standardized test scores are derived from student test records. The teacher data come from a state-maintained archive of personnel records. For each teacher, information is available on type of license, licensure test scores, including the test taken and the year it was administered;

undergraduate institution attended, whether the teacher has any advanced degrees or is National Board Certified, and the number of years of teaching experience.

Crucial for this analysis is the identification of each student's actual math and reading teacher. For self-contained classrooms, which are common in elementary schools, the student might well have the same teacher for math and reading. In other school structures, the teachers could well differ. Although each record of a student's test score in either math or reading identifies the teacher who administered the test, we were not willing simply to assume that this proctor was in fact the student's teacher of that subject. We first checked to make sure that the student's proctor was a valid math (or reading) teacher. We deemed a proctor to be the actual teacher if she taught a course with a state-specified code for the relevant subject during the semester in which the test was administered. [7] We supplemented this approach with two other identification strategies. When the proctor was not a valid math or reading teacher, but the student was in a school and grade in which a single math (reading) teacher had at least a 95 percent share of the math (reading) students, we assigned that teacher to the student. Finally, in a few cases we were able to make either an exact or an approximate match, based on the characteristics of the group of students taking the test together and our information from school activity reports about the characteristics of the classes taught by the proctor of those students during that semester. [8]

---

[7] In particular, a teacher was deemed to be a valid math teacher if that teacher taught a course with one of the following subject codes: 0000, 0120,0123,01234,0129,0230,0239,0290,0420,0423 or any code between and including 2001 and 2078. A teacher was deemed a valid reading teacher is that teacher taught a course with one of the following subject codes: 0000,0109,0120,0123,0124,0129,0130,0140, or any coed between and including 1001 and 1038.

[8] The composition of students taking an end-of-grade exam together with a particular teacher proctor was compared to the compositions of students in the classes taught by that teacher during the semester the test was given. In the event of an exact match in the total number of students and in the numbers of male,

These procedures allowed us to identify math and reading teachers for at least 75 percent of all students in grades 3, 4 and 5 during the period 1994/05 to 2003/04. For grades 6 to 8 we were far less successful. For 2002/03, for example, we were able to match only 34 percent of the sixth grade students, only 29 percent of the seventh graders and only 26 percent of the eighth graders to their math teachers. These low match rates in the higher grades forced us to limit our analysis to students in grades 3, 4 and 5. The match rates for each cell for both math and reading are shown in table 1. Appendix table A1 reports the total number of students in each grade and year for which test scores are available in each subject.

In all our regressions, the dependent variable is a standardized end-of grade test score in either reading or math for each student or the year-to-year change in that variable. As noted earlier, these tests are directly related to the state's standard course of study for each grade. Each test score is reported on a development scale score, but we converted all the scale scores to standardized scores with means of zero and standard deviations equal to one based on the test scores in each subject in each grade in each year. This standardization makes it possible to compare test scores across grades and years.

Teacher credentials and characteristics

The basic measures of teacher credentials include licensure test scores, years of teaching experience, type of license, quality of undergraduate college, whether the teacher has an advanced degree, and licensure test scores. We describe here the measures

---

female, white , and nonwhite students, we treated the teacher of that class as the relevant teacher for that group of students. For an approximate match, we determined the percentage difference between the class taught by that teacher and the group of students the teacher proctored for each characteristic. A match was considered approximately accurate if the square root of the sum of the squared percentage differences did not exceed 0.125.

that we use in the basic equations. In section VI, we describe additional forms of some of the credentials.

Teacher test scores. From the early 1960s through the mid-1990s, all elementary school teachers in North Carolina were required to take either the Elementary Education or the Early Childhood Education test. Included in the former was material on curriculum, instruction and assessment. Starting in the mid-1990s, teachers were required to take both that basic elementary test and a test that focused on content. We normalized test scores on each of these tests separately for each year the test was administered based on means and standard deviations from test scores for all teachers in our data set, not just those in our subset of teachers matched to students. For teachers with multiple test scores in their personnel file, our teacher test score variable is set equal to the average of all scores for which we can perform the normalization.

Years of teaching experience. We measure years of teaching experience as the number of years used by the state to determine a teacher's salary. Thus, this measure counts all the years of teaching whether in the State of North Carolina, or elsewhere, for which the state has given the teacher credit.[9] Because of our own prior research and that of others (e.g. Hanushek, Kain, O'Brien and Rivkin, 2005), we expect the returns to additional years of experience to be highest in the early years. We allow for this nonlinearity by specifying years of experience as a series of indicator variables, with the base or left-out category being no experience.

---

[9] The teacher experience variable was missing for some teachers. In cases where it was possible to observe experience levels in payroll records from other years, we imputed values. In cases where observations from other years' payroll data were inconsistent with the recent record, we put more weight on the more recent record.

Licensure information. The state of North Carolina has many types of licenses which we have divided into three categories: regular, lateral entry and "other." Lateral entry licenses are issued to individuals who hold at least a bachelors degree with a minimum 2.5 GPA and the equivalent of a college major in the area in which they are assigned to teach. Such teachers must affiliate with colleges and university to complete prescribed coursework. Currently the licenses are issued for two years and can be renewed for a third year. Because lateral entrants who remain in teaching eventually convert to a regular license, we include an additional variable to identify those teachers in a later year who initially entered as a lateral entrant. The "other" category includes a variety of provisional, temporary, and emergency licenses.

Graduate degree. For the basic regression we include a single variable to indicate whether the teacher has a graduate degree of any type such as a masters that leads to a higher salary, Ph.D., or other advanced degree including those that do not affect the teacher's salary. In subsequent analysis we break these degrees down by type and by when the teacher earned the advanced degree.

National Board Certification. North Carolina has been a leader in the national movement to have teachers board certified by the National Board for Professional Teaching Standards (NBPTS), and provides incentives in the form of a 12 percent boost in pay for teachers to do so. Such certification, which requires teachers put together a portfolio and to complete a series of exercises and activities designed to test their knowledge of material for their particular field, takes well over a year and is far more difficult to obtain than state licensure. For this study we obtained the names of all North Carolina certified teachers by year and grade from the NBPTS and then the NC Education Research Data

Center matched those names to our information on all North Carolina teachers. Despite significant effort, we were able to match only about 90 percent of the teachers.[10] As of 2004, our matched sample of math and reading teachers included about 300 board certified teachers in each of the grades 3-5.

Quality of undergraduate institution. Available for each teacher is the name of the undergraduate institution from which she graduated. Following standard practice we assign to each institution a competitive ranking based on information for the 1997-98 freshman class from the Barron's College Admissions Selector. Barron's reports seven categories which we aggregated to four categories: uncompetitive, competitive, very competitive, and unranked. Many of the state's teacher preparation programs are offered by state institutions in the competitive category.

Teacher characteristics. In addition to these measures of teacher credentials, the basic equations also include a number of standard teacher characteristics. These include the indicator variables for whether the teacher is male, black, Hispanic or "other race." In addition, in light of the work by Dee (2005) we include a variable indicating whether or not the teacher is the same race as the student. Similarly, we include a variable indicating whether the teacher is the same gender as the student.

Student characteristics

As discussed earlier, student-specific variables come in two forms: those that measure relatively permanent characteristics and those that vary over time. Only the ones

---

[10] For each Board Certified teacher we had the teacher's name, and after 2000, the teacher's school in most cases. In addition to the standard problems of matching names with potential variation in spelling, two other explanations help to account for the incomplete match. First is that some Board Certified teachers are teaching in private schools and, hence, are not in our data set. Second, we are likely to lose some teachers whose names changed between the fall when they were identified as Board Certified and the spring term, or who stopped being a NC teacher during that period.

that vary over time are used in models 4 and 5 because of the presence of the student fixed effects in those models. Included in the permanent category are the student's race, gender, and age in grade three.[11] In addition, we include indicators for whether the student is classified as limited English, gifted, or special needs. Though each of these latter characteristics may vary over time, we exclude them from the models with student fixed effects on the ground that such variation is likely to be small.

In addition, data considerations force us to treat as nonvarying two measures of the student's family background: family income as proxied by whether or not the student is eligible for a subsidized lunch and the education level of the student's parents. Because student-specific information on subsidized lunch is missing from the early years of our data set, we have chosen to assign a single status to each student based on her subsidized lunch status in 7th or 8th grade, years for which we have information for all cohorts.[12] Given that middle school students are often more reluctant than students in earlier grades to apply for subsidized lunches, this variable is a relatively conservative measure of family poverty. The problem with our information on parental education is that because it is based on teacher reports it exhibits a lot of noise over time. Though the actual education level of the parents could change as parents earn more degrees, marry or divorce, we suspect that most of the variation reported on the student's test score record over time is noise. Hence, we identify a parent as having a college degree or more, a high

---

[11] We include the age variable to account for the fact that student start school at different ages and hence at different stages of their development. The age variable also picks up the effect of repeating or skipping earlier grades, but not of repeating the current grade which is represented by a separate variable.

[12] Specifically, for a student for whom we have data from either 7th or 8th grade, we assign the value 1 to the students if she was on free or reduced price lunch in either of those years. For a student for whom we do not have that data for either of those grades but for whom we have such information for an earlier year we use the earlier information. The goal is to use the most consistent data available while retaining as large a sample as possible.

school degree or more but no college degree, or no high school degree based on the most common identification for a students during the grades 3-8.[13]

The student-level variables that change over time include indicator variables for the following: the student is repeating the grade, the student has transferred into the school that year, and the student has transferred into the school as part of a structural change, such as from a school than ends in fourth grade to a school that begins in fifth grade.[14] These time-varying variables are included in all five models. In addition, we include in models 1 and 2, each student's lagged test score in the specified subject in the prior year.

Classroom characteristics

The classroom variables include the class size, and several measures of the characteristics of the student's peers. These measures include the percentages of students who are nonwhite and who are on subsidized lunch, and the percentages of students whose parents have college degrees, have high school degrees, or are high school drop outs. Finally, in the model with lagged terms, we also included the lagged class average math or reading score as a proxy for the average ability level in the class. Each of these peer variables is based on the other students in the class, that is, excluding the student for whom the peer groups are being defined.

III. Basic results

Tables 2 and 3 summarize the results for the teacher variables for the five models for math and reading, respectively. In all cases, the reported regressions are based on as

---

[13] For example, we identify a parent as having a college degree if for a majority of years for which we have information for that student, the parent is identified as having a college degree.

[14] Though Rockoff (2004) also includes a variable indicating that a student will repeat the grade, that variable is potentially endogenous and, hence, does not belong in the equation.

much data as possible. Hence, because model 4 includes no lagged term either as an independent variable or as part of the depending variable, the reported results are based on the achievement of 3rd, 4th and 5th graders, rather than just of 4th and 5th graders as in the other models.[15] To highlight the findings for teacher characteristics and credentials, the tables exclude the coefficients of the time-invariant student variables included in models 1 and 2 and the time-varying student variables included in all five models. The coefficients for those variables are reported in appendix tables A2 and A3. Of interest is that many of the student-specific variables, such as parental education, exert larger effects on reading than on math.

The findings for the teacher variables are remarkably consistent across the five models, and, perhaps most importantly, are consistent with the predictions about the direction of the biases for the estimated effects of the teacher credentials. Comparing the results for the teacher experience variables for math, we find, as expected, that the coefficients from model 1 are higher than those for models 2 and 3. In addition, the results tend to confirm the prediction that models 4 and 5 provide upper and lower bounds on the true effects. We refer to these lower and upper bounds as we quantify effects in the following discussion.

The lagged achievement terms, which appear only in models 1 and 2, enter with coefficients of about 0.70 for math and about 0.68 for reading. Given that any bias in these coefficients is likely to be toward zero because of measurement error, these estimates are consistent with the conclusion that the rate at which knowledge persists

---

[15] We have also estimated model 4 equations for the sample restricted to 4th and 5th graders. The estimated patterns are similar to those presented, with the coefficients of the teacher credentials variables typically somewhat smaller than the ones reported in table 4. We prefer the reported results on the ground that longer panels are preferred to shorter ones.

from one year to the next is well over 50 percent. That would render the results for

teacher credentials from model 5 (which would generate unbiased estimates of teacher

credentials if the persistence rate were 100 percent) somewhat closer to the truth than

those in model 4 (which would generate unbiased results if the persistence rate were

zero.) [16]

Teacher characteristics

We turn first to the effects of teacher characteristics. For math achievement, we

find quite consistent results across models, with male teachers generating less positive

results than female teachers, and black teachers less positive results than white teachers.

For reading, in contrast, no differences emerge in the effectiveness of black and white

teachers in any of the models. Of most interest is the finding that when a student and a

teacher are the same race, the effects on student achievement are positive, with an effect

size of about 0.020 to 0.029 standard deviations for math and 0.013 to 0.020 for reading.

Thus our results confirm those of Dee (2005) based on his study of data from the

Tennessee Star Experiment in which students were randomly assigned to classrooms and,

hence, to teachers.

Teacher credentials

Though the estimated patterns of results for teacher credentials are similar for

math and reading, in almost all cases the estimated achievement effects of the various

teacher credentials are larger for math. That pattern is not surprising given the common

view that schools have a larger role to play relative to families in the teaching of math

---

[16] To account for the possibility that achievement in math or reading may be affected by prior achievement in both subjects and in a nonlinear way, we estimated other versions of model 2 that included lagged test scores in both subjects in both linear and squared form (results not shown). The addition of those variables has no noticeable effect on the estimates of the teacher credentials in model 2.

than in reading and also in light of the finding that the student characteristics exhibit larger impacts on reading than on math achievement in these models. Thus, we focus the text discussion on the math results as shown in Table 2.

Consistent with other studies (see, in particular, Hanushek, Kain, O'Brien and Rivkin 2005 and Clotfelter, Ladd and Vigdor, 2006), we find clear evidence that teachers with more experience are more effective than those with less experience. Compared to a teacher with no experience, the benefits of experience rise monotonically to a peak in the range of 0.092 (from model 4) to 0.119 (from model 5) standard deviations after 21-27 years of experience, with more than half of the gain occurring during the first couple of years of teaching. In section VI, we explore these findings in more detail.

Teacher licensure also seems to matter. Here the base case is a teacher with a regular license. Most clear are the negative effects on achievement for those with "other" types of provisional or emergency licenses, with the estimates ranging from -0.033 to – 0.059 across models 4 and 5. Teachers operating under a lateral entry license exhibit a statistically significant negative average effect on student achievement, but only in model 4 and it is not clear whether that negative effects persists after the lateral entrant receives a regular license. These results for lateral entrants appear to be quite consistent with the more detailed investigation of pathways into teaching in New York State by Boyd et al. (2006). That study found that teachers with reduced coursework prior to entry often exhibited smaller initial gains that other teachers, but that the differentials were small and disappeared as the cohort matured.

Despite that fact that teachers are rewarded in the form of higher salaries for having a master's degree, the variable denoting having a graduate degree exerts no

statistically significant effect on student achievement and in some cases the coefficient is negative. We explore the reasons for that finding in section VI. In contrast, teachers who are National Board certified appear to be more effective (with coefficients of 0.020 to 0.027) than those who are not. However, from these basic regressions we cannot tell whether this greater effectiveness is because the teachers who become Board certified are the more effective teachers to begin with or whether the rigorous process of Board certification makes them better teachers. We return to that issue below.

We have included a measure of the competitiveness of the teacher's undergraduate institution since that is a common measure of teacher credentials used in other studies. Perhaps because of the rich set of other measures included in this study, and in particular the inclusion of the teacher's test score, these variables exhibit only very small effects, at most, on student achievement. The clearest finding is that a teacher who comes from an undergraduate institution ranked as competitive appears to be somewhat more effective on average than one from an uncompetitive institution. Coming from an elite and very competitive institution, however, apparently does not make a teacher any more effective on average relative to teachers from other institutions.

The final measure of teacher credentials, the teacher's test score, enters as expected with a positive and statistically significant coefficient. The coefficient implies that a teacher whose test scores were one standard deviation above the average would increase student achievement by 0.011 to 0.015 standard deviations. We provide more disaggregated results in section VI. .

Classroom Characteristics

Among the classroom characteristics the most consistent results emerge for the class size variable and the percent of students on free and reduced price lunch. Class size consistently enters with a negative coefficient of about 0.002 to 0.005. This finding implies that reducing class sizes in elementary schools by five students would increase student achievement by about 0.010 to 0.015 standard deviations on average.

Also consistently negative for math achievement are the effects of concentrations of poor children in a student's classroom. A 10 percentage point increase in the fraction of students in the class receiving subsidized lunches decreases math achievement by about 0.005 standard deviation. In addition, students in classrooms with greater proportions of parents who are high school graduates appear to do less well than those in which there are more college graduates.

V. Interpreting the magnitudes

Each teacher brings to the classroom a bundle of personal characteristics and credentials. Hence, we illustrate the magnitudes of the estimated teacher effects by comparing teachers with different bundles of attributes. Consider for example a baseline teacher with the following relatively typical attributes listed in the first column of Table 4. The teacher has 10 years of experience, attended a competitive undergraduate college, and has a regular license, an average test score, and a graduate degree. In addition, we assume somewhat less typically that she is National Board certified. We then compare her to a teacher with far weaker credentials as described in the second column. That teacher has no teaching experience, attended a non-competitive undergraduate college,

does not have a regular license, has a test score one standard deviation below average, does not have an advanced degree and is not Board certified.

Based on the lower bound estimates from model 4 and the upper bound estimates from model 5, the following two columns depict the reasonable range of differential effects on student achievement in both math and reading, all other factors held constant. We remind the reader that the true effects are likely to be somewhat closer to the model 5 than the model 4 estimates. The first observation to emerge from the calculations is the far larger adverse effects from having a teacher with weak credentials on math than on reading achievement. For math the total effects of having the weak teacher range from –0.150 to –0.206 standard deviations and for reading from -0.081 to –0.120. Second, the biggest differentials are associated with experience and licensure status. Of course, by assuming the subject teacher has no teaching experience we have magnified the effects of experience. If, instead, the subject teacher had one or two years of experience, the total effect in math would have ranged from –0.093 to –0.134 and for reading would have been -0.049 to –0.077. Though the comparison in Table 4 is merely illustrative, it does provide some information with which to evaluate the magnitude of the estimated effects. The question is whether these teacher effects are large or small.[17]

Relative to the estimated effects of class size, the effects of teacher credentials.appear to be quite large. Based on the estimated coefficients for the class size variable, an increase of 5 students in an elementary school class would reduce student

---

[17] Another way to address the question of magnitude is to take account of the fact that some of the variation in the dependent variable is measurement error. To account for that variation, one might want to gross up the estimates of the teacher credentials so that they reflect not the effects on measured test scores but rather the effects on a more accurate measure of learning. Doing so would make the effects of teachers seem larger than in this study, but would have no effect on the comparisons of relative magnitudes that we discuss here.

achievement in math by about 0.015 to 0.025 standard deviations and in reading by about

0.010 to 0.020 standard deviations, far smaller effects that those associated with having a

teacher with weak credentials.

An alternative comparison is to the effects of demographic characteristic such as

parental education. From the results for student characteristics reported in the appendix

for model 2 (the model with school, but not student, fixed effects), we see that, relative to

having a parent who is a college graduate, having a parent without a college degree

reduces predicted math achievement by about 0.11 standard deviations if the parent has a

high school degree and by another 0.11 standard deviations if the parent is a high school

drop out. The effects are slightly larger for reading: 0.11 for those with high school

degrees and another 0.14 for high school drop outs. Thus, for math, having a teacher

with weak credentials has negative effects generally comparable in size to those

associated with having poorly educated parents. For reading, the negative effects

associated with having a teacher with poor credentials, though still harmful for

achievement, are not as harmful as having poorly educated parents.

Thus, we conclude that a variety of teacher credentials matter for student

achievement and that the effects are particularly large for achievement in math. As a

result how teachers with differing qualifications are distributed among classrooms and

schools matters. To the extent that the teachers with weaker credentials end up in

classrooms with the more educationally disadvantaged children, schools would tend to

widen, rather than reduce, the already large achievement gaps associated with the

socioeconomic differences that students bring to the classroom. In related research, we

have documented that such disparities exist both by race of the student and poverty level

31

of the school (Clotfelter, Ladd and Vigdor, 2002, and Clotfelter, Ladd and Vigdor, 2006, and Clotfelter, Ladd, Vigdor and Wheeler, 2006)

VI. More detailed analysis of specific teacher credentials.

We now turn to more detailed analysis of four of the teacher credentials: graduate degrees, National Board Certification, teacher test scores, and teacher experience. For this analysis we report results only for models 4 and 5, on the ground that they are likely to bracket the true results, and only for the specific credentials of interest. In all cases, the new variables are embedded in the full models.

Graduate degrees

One of the most counterintuitive findings to emerge from the basic models is the small or negative effects of having a graduate degree. Most of those degrees are master's degrees that generate higher salaries for teachers. A negative coefficient would suggest that having such a degree is not associated with higher achievement. Thus, if the goal of the salary structure were to provide incentives for teachers to improve their teaching, the higher pay for master's degrees would appear to be money that is not well spent, except to the extent that the option of getting a master's degree keeps effective experienced teachers in the profession.

The first step in our supplemental analysis is to disaggregate the degrees by type: master's, advanced, and Ph.D. As we noted earlier the category of advanced degree generally applies to graduate degrees that do not increase teacher salaries and teachers are not required to report them. The first row of Table 5 replicates the coefficients that emerged for both math and reading from the basic model. The second panel shows the

disaggregated results. Emerging most clearly is the relatively large negative effects for advanced degrees and the very small -- and in half the cases not statistically significant -- negative coefficients for master's degrees. Thus, it appears that the negative effect in the basic model is more attributable to the advanced degrees than to the master's degrees. The large negative coefficient on the Ph.D variable for math is probably an anomaly given the small number of elementary school teachers with Ph.Ds.

In the third panel we disaggregate the master's degrees by the period during which the teacher earned the degree. The estimates indicate that the teachers who received their degree prior to entering teaching or any time during the first five years of teachers were no less or no more effective than other teachers in raising student achievement. In contrast, those who earned their master's degree more than five years after they started teaching appear to be somewhat less effective on average than those who do not have master's degrees. Whether this negative effect means that those who seek master's degrees at that stage in their career are less effective teachers in general or whether having a master's degree makes them less effective cannot be discerned with complete confidence from this analysis. The observation that the earlier master's degree has no effect, however, suggests that the negative sign is more attributable to who selects into that category than to any negative effect of the degree itself.

National Board Certification

The positive association that emerges between National Board Certification and student achievement in the basic models confirms that teachers with that credential are more effective, on average, than those who are not certified. However, those results alone cannot distinguish whether the National Board is simply identifying the most effective

teachers from whether the process itself (or possibly the recognition associated with certification) makes the teachers more effective than they otherwise would have been. Table 6 sheds light on this issue by reporting results for four NBCT indicator variables embedded in models 4 and 5. NBCT-2 and NBCT-1 take on the value 1 two years and one year, respectively, prior to the year in which the teacher becomes certified. NBCTcurrent takes on the value 1 in the academic year in which the teacher is certified and NBCTpost represents each subsequent year that she is certified.

The positive and statistically significant coefficients for NBCT-2 indicate that the Board does indeed confer certification on the more effective teachers, as would be appropriate to the extent that the policy goal is to reward effective teachers. These coefficients range from 0.024 to 0.055 standard deviations for math and from 0.026 to 0.038 standard deviations for reading. The fact that the NBCTcurrent and post coefficients are all lower (although not in a statistically significant sense) than the NBCT-2 coefficients provides no support for the hypothesis that the certification process makes teachers more effective than they otherwise would be. If anything, the lower coefficients on the NBCTpost variables suggest that teachers may be less effective – where effectiveness is measured by success in raising test scores—after receiving certification than before. Again, though, the differences between the NBCT-2 and NBCTpost coefficients are not statistically significant at standard levels.

These findings are fully consistent with those of Goldhaber and Anthony's more detailed study (2005) of National Board Certification in North Carolina for a somewhat earlier, but overlapping, time period. With these findings in hand, a logical next question for policy makers is the extent to which National Board Certification succeeds in keeping

the more effective teachers in the profession longer than would otherwise be the case. If certification has no effect on retention, then any additional salary paid to certified teachers is simply a reward for good work but not a way to improve student achievement. A second policy question relates to how Board certified teachers are distributed among schools and classrooms. As shown for fifth grade classrooms in Clotfelter, Ladd and Vigdor (2006), Board certified teachers tend to be in schools with whiter student bodies and with somewhat smaller proportions of low-income students compared to non-certified teachers. Even more dramatic are the patterns across classrooms within schools. Compared to the classrooms of non-certified teachers within each school, the classrooms served by certified teachers were more advantaged along every dimension: race, income, education level of the parents, and mean average prior year test scores. These and other policy-relevant questions are the topic of ongoing research both by ourselves and of others.

Teacher test scores

In the basic model, teacher test scores are simply normalized test scores averaged over all the tests taken by each elementary school teacher. As shown in the top row of Table 7, higher average test scores are associated with higher math and reading achievement, with far larger effects for math than for reading. To test for nonlinear patterns, in the second panel of the table we disaggregated the test scores into a series of indicator variables.

The results for math achievement are quite striking and exhibit some clear nonlinearity. Specifically, having a teacher at one of the extremes of the distribution has a big effect on achievement relative to having an average teacher. Referring to the results

for model 5 (based on achievement gains), we see that teachers who scored 2 or more standard deviations above the average boosted student gains by 0.068 standard deviations relative to the average teacher, and teachers who scored 2 or more standard deviations below the average reduced achievement gains by 0.062 standard deviations. The overall difference between teachers at the two extremes is a whopping 0.130 standard deviations, which is far larger than the 0.060 standard deviations that would be predicted from the linear specification. A similar nonlinear pattern emerges from the model 4 results for math, but the difference between the extremes is far smaller at 0.074 standard deviations. For reading, all the effects are much smaller and any nonlinearities are hard to detect.

The third panel disaggregates the tests into two types, both of which changed form over time. One type is the basic elementary education test and the other is more content based. [18] In addition to the variables reported in the table, we also included two indicator variables for whether or not a teacher took each of the tests. The results are inconclusive in that they vary depending on the model. Model 4 suggests that the elementary education test is a better predictor of student achievement in both math and in reading. According to the gains model, however, in math, the coefficient for the content test is slightly larger, although still indistinguishable from that for elementary education, and neither coefficient is distinguishable from zero for reading.

Teacher experience

Though the positive results by years of teacher experience are clear and robust to various model specifications, the thorny issue remains of whether the rising returns to

---

[18] Most teachers took only one test of each type but if they took more than one we averaged over the two tests. The first test is either the 0010 test or the 0011 test. The second is either the 0020 test or the 0012 test.

experience reflect improvement with experience or differentially higher attrition of the less effective teachers. In his study of 3 New Jersey school districts, Rockoff (2004) found that, even after including teacher fixed effects to control for the permanent characteristics of teachers, years of experience still emerged as a determinant of student achievement.

We shed light on this issue by making use of information on teacher longevity as shown in Table 8. Specifically, we have added to models 4 and 5, an indicator variable for whether or not the teacher remains a North Carolina teacher for at least three years and an interaction term between that variable and the indicator variable for 1-2 years of experience.[19] The negative coefficients of –0.019 and –0.033 on the indicator variable in the math equations suggests that the teachers who stay may be less effective on average than the ones who leave, a finding that is inconsistent with the differential attrition explanation for the rising returns to experience. Moreover, the fact that the interaction terms, in both the math and the reading equations, are not statistically significant suggests that differential attrition does not generate upward biased estimates of the returns to experience in the early years of teaching. Hence, we conclude that the returns to education that emerge from our basic model are primarily attributable to learning from experience.[20]

---

[19] For this analysis we have deleted from the full sample any students taught in a particular year by any recent young teacher for whom we do not yet have enough information to determine whether they remain a North Carolina teacher for at least three years.

[20]. We have not tried to add teacher fixed effects to any of our models that include either school or student fixed effects, largely because of the technical difficulties of doing so. We have, however, included teacher fixed effects in model 1. Consistent with the findings from the other test described in the text, we find that the new coefficients for the teacher experience variables increase slightly (and range from 0.069 for 1-2 years of experience to 0.135 for 21-27 years of experience) for math once the teacher fixed effects are included and a similar pattern emerges for reading (with the new coefficients ranging from 0.043 to a maximum of 0.101).

VI. Conclusion

As we noted earlier, the basic approach of the paper is not new. In fact, it is the approach used in a large number of papers in the tradition of education production functions. What differentiates this paper from other work is the richness and coverage of the administrative data on which the analysis is based. To our knowledge, no other statewide data base allows the matching of students with their specific teachers The availability of longitudinal data allows us to use student fixed effects which helps to minimize the statistical problems that have plagued much of the previous work on teacher credentials.

Our findings for the experience variables are fully consistent with those of other studies, including our own previous cross sectional analysis. In particular, close to half the achievement returns to experience arise during the first few years of teaching but returns continue to rise throughout most of the experience range. Furthermore it appears that all of the returns are attributable to experience *per se* rather than to differential rates of attrition between more or less effective teachers. In addition to the findings for experience, we also document that the state's licensure tests provide policy relevant information, especially with respect to the teaching of math; that the form of licensure matters; that the National Board Certification process appears to identify effective teachers but does not make them more effective; and that master's degrees obtained after 5 years of teaching are associated with negative effects on student achievement.

Taken together the various teacher credentials appear to have quite large effects on math achievement, whether compared to the effects of changes in class size or to the socio-economics characteristics of students, as measured, for example, by the education

38

level of their parents. The effects of teacher credentials on reading achievement are noticeably smaller, especially relative to the effects of family background. Thus, even highly credentialed teachers are not likely to offset the effects of educationally impoverished family backgrounds on student achievement in reading. For math, however, the results are potentially more optimistic. The real challenge for policy makers is to find ways to direct the teachers with strong credentials to the students who most need them.

References

Aaronson, D., L. Barrow and W. Sander. (2003) "Teachers and Student Achievement in the Chicago Public High Schools." Federal Reserve Bank of Chicago manuscript.

Ballou, D., W. Sanders and P. Wright (2004) "Controlling for Student Background in Value-Added Assessment of Teachers," *Journal of Educational and Behavioral Statistics* 29 (Spring), 37-66.

Boardman, E.E. and R.J. Murnane. (1979) "Using Panel Data to Improve Estimates of the Determinants of Educational Achievement." *Sociology of Education* v.52 pp.113-121.

Boyd, D., P. Grossman, H. Lankford, S. Loeb, and James Wyckoff. 2006. "How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement." *Education Finance and Policy*, vol. 1, Issue 2 (Spring), pp.176-216.

Clotflelter, C.T., H.F. Ladd, and J.L. Vigdor, 2002 "Who Teaches Whom? Race and the Distribution of Novice Teachers." *Economics of Education Review*, 24 (2005), pp. 377-392.

Clotfelter, C.T., H.F. Ladd, and J.L.Vigdor. 2006 . "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources,* vol. XLI, N. 4 (Fall), pp. 778-820.

Clotfelter, C.T., H.F. Ladd, J.L.Vigdor, and J. Wheeler. 2006. "High Poverty Schools and the Distribution of Principals and Teachers." Sanford Institute Working Paper. www.pubpol.duke.edu/research/papers/SAN06-08.pdf>

Dee, T.S., "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" 2005. *American Economic Review* 95 (May) , 158-165.

Goldhaber, D. and E. Anthony (2005) "Can Teacher Quality be Effectively Assessed?" updated version of an Urban Institute Working Paper.

Goldhaber, D.D. and D.Brewer (2000) "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." *Educational Evaluation and Policy Analysis* v.22 pp.129-45.

Hanushek, E.A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis* 19(2): 141-164.

Hanushek, E. A., J. F. Kain, D. M. O'Brien and S.G. Rivkin (2005) "The Market for Teacher Quality," unpublished paper, Stanford University, January.

Hanushek, E.A. J.F. Kain, and S.G. Rivkin (2006). "New Evidence about Brown v. Board of Education: The Complex Effects of school Racial Composition on Achievement. (Unpublished paper). March.

Hedges, L.V., R. Laine, and R. Greenwald (1994) "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes," *Educational Researcher*23 (April): 5-14.

Lankford, H., S. Loeb, and J. Wyckoff. (2002) "Teacher Sorting and the Plight of Urban Schools," *Educational Evaluation and Policy Analysis* v.24 pp.37-62.

Lalonde, R., "Evaluating the Econometric Evaluations of Training Programs with experimental Data," *American Economic Review* 76, 604-620

Nye, B., S. Konstantopoulos, and L.V. Hedges (2004) "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis* v.26 n.3

Oakes, J. (1995) "Ability Grouping, Tracking and Within-School Segregation in New Castle County Schools," Report to the U.S. district Court for the District of Delaware in the case of *Coalition to Save our Children v. State Board of Education, et al.* December 9, 1994 (corrected January 1, 1995).

Public Education Association (New York City). (1955) *The Status of the Public School Education of Negro and Puerto Rican Children in New York City*, Board of Education Commission on Integration, October.

Rivkin, Steven. 2006. "Cumulative Nature of Learning and Specification Bias in Education Research. (unpublished manuscript) January 2006.

Rivkin, S.G., E.A. Hanushek, and J.F. Kain (2005) "Teachers, Schools and Academic Achievement." *Econometrica* v.79 pp.418-458.

Rockoff, J.E. (2004) "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review Papers and Proceedings* (May 2004), 247-252.

Summers, A.A. and B.L. Wolfe (1977) "Do Schools Make A Difference?" *American Economic Review* v.67 pp.639-652.

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

Table 1: Proportions of all students with test scores for whom we can match math and reading teachers, by grade and by year.

Math	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
3rd	.828	.747	.821	.833	.838	.857	.862	.833	.814	.799
4th	.802	.742	.811	.805	.827	.844	.848	.817	.800	.781
5th	.772	*	.792	.790	.807	.814	.827	.807	.784	.773
Reading										
3rd	.827	.750	.821	.831	.838	.857	.860	.831	.815	.799
4th	.811	.735	.814	.808	.833	.851	.855	.823	.808	.787
5th	.789	*	.799	.794	.811	.821	.834	.810	.801	.785

See text for the matching criteria. Test score records are missing for 5th grade students in 1996. See appendix table A1 for the total number of student records available for the same grades and years.

Table 2. Achievement models for math [a]					
	(1) Levels No FE 4th and 5th graders (with lags)	(2) Levels School FE 4th and 5th graders (with lags)	(3) Gains School FE 4th and 5th graders	(4) Levels Student FE 3rd, 4th and 5th graders (without lags)	(5) Gains Student FE 4th and 5th graders
Student characteristics that are constant over time [b]	Yes	Yes	Yes	No	No
Student characteristics that vary over time [b]	Yes	Yes	Yes	Yes	Yes
Lagged math score	0.071** (0.004)	0.704** (0.002)	---	---	---
Teacher characteristics					
Male	-0.010* (0.004)	-0.010** (0.004)	-0.009* (0.004)	-0.013** (0.004)	-0.019** (0.007)
White (base)	--	--	--	--	--
Black	-0.026** (0.004)	-0.031** (0.004)	-0.030** (0.004)	-0.026** (0.003)	-0.037** (0.007)
Hispanic	0.033 (0.021)	0.011 (0.021)	0.013 (0.022)	-0.006 (0.017)	0.021 (0.041)
Other race	-0.042** (0.013)	0.016 (0.015)	0.013 (0.016)	0.042** (0.011)	0.055* (0.027)
Same race as student	0.026** (0.002)	0.017** (0.002)	0.018** (0.002)	0.020** (0.002)	0.029** (0.005)
Same gender as student	0.004* (0.002)	0.004* (0.002)	0.006** (0.002)	0.003 (0.002)	0.009* (0.004)
Teacher credentials					
No experience (base)	--	--	--	--	--
1-2 years	0.066** (0.005)	0.063** (0.005)	0.062** (0.005)	0.057** (0.004)	0.072** (0.009)
3-5 years	0.082** (0.005)	0.078** (0.005)	0.072** (0.005)	0.072** (0.004)	0.091** (0.009)
6-12 years	0.087** (0.005)	0.083** (0.005)	0.078** (0.005)	0.079** (0.004)	0.094** (0.009)
13-20 years	0.088** (0.005)	0.091** (0.005)	0.086** (0.005)	0.082** (0.004)	0.102** (0.009)
21-27 years	0.103** (0.005)	0.104** (0.005)	0.098** (0.005)	0.092** (0.004)	0.119** (0.009)
28+ years	0.097** (0.006)	0.094** (0.006)	0.086** (0.006)	0.085** (0.005)	0.109** (0.010)
Regular license (base)	--	--	--	--	--
Lateral entry	-0.031 (0.019)	-0.024 (0.018)	-0.010 (0.020)	-0.033* (0.017)	-0.022 (0.035)

Interact continuing/lateral entry	-0.017 (0.015)	-0.007 (0.015)	-0.025 (0.016)	-0.021 (0.013)	-0.037 (0.026)
Other license	-0.034** (0.004)	-0.028** (0.004)	-0.031** (0.005)	-0.033** (0.004)	-0.059** (0.008)
Graduate degree	-0.004 (0.002)	-0.004 (0.002)	-0.004 (0.002)	-0.003 (0.002)	0.002 (0.004)
National Board Certified	0.021** (0.006)	0.020** (0.006)	0.018** (0.006)	0.020** (0.005)	0.028* (0.011)
Undergraduate institution non-competitive (base)	--	--	--	--	--
Competitive	0.008** (0.003)	0.007* (0.003)	0.007* (0.003)	0.007** (0.002)	0.010 (0.005)
Very Competitive	0.011** (0.004)	0.004 (0.004)	0.002 (0.004)	0.001 (0.003)	0.006 (0.007)
Unranked	-0.010 (0.007)	-0.011 (0.007)	-0.012 (0.007)	-0.007 (0.006)	-0.013 (0.013)
Mean teacher test score	0.014** (0.001)	0.013** (0.001)	0.012** (0.001)	0.011** (0.001)	0.015* (0.003)
Classroom characteristics					
Class size	-0.003** (0.000)	-0.002** (0.000)	-0.003** (0.000)	-0.002** (0.000)	-0.005** (0.001)
Percent nonwhite	0.007 (0.006)	-0.002 (0.011)	0.046** (0.012)	0.017* (0.008)	0.022 (0.019)
Percent subsidized lunch	-0.045** (0.008)	-0.093** (0.009)	-0.060** (0.009)	-0.039** (0.008)	-0.084** (0.019)
Percent college grad (base)	--	--	--	--	--
Percent hs grad	-0.044** (0.007)	-0.100** (0.009)	-0.057** (0.009)	-0.026** (0.007)	-0.035* (0.018)
Percent hs dropout	0.003 (0.014)	-0.091** (0.018)	-0.005 (0.019)	0.006 (0.015)	0.059 (0.036)
Lagged class average reading score	0.031** (0.004)	-0.023** (0.005)	--	--	--
Constant	0.537** (0.019)	0.614** (0.020)	0.313** (0.018)	0.004 (0.009)	0.051* (0.021)
R-squared	0.7150	0.7232	0.0683	0.9062	0.5021
Observations	1,089,132	1,089,132	1,089,132	1,805,638	1,089,503

Notes.
a. Based on a panel data set of 3rd, 4th and 5th graders from 1995 to 2004. Dependent variable is student achievement in each year. FE stands for fixed effect; Robust standard errors are in parentheses; ** signifies statistical significance at the 0.01 level and * at the 0.05 level.
b. The coefficients and standard errors for these variables are in appendix table 2.

Table 3. Achievement models for reading [a]					
	(1) Levels No FE 4th and 5th graders (with lags)	(2) Levels School FE 4th and 5th graders (with lags)	(3) Gains School FE 4th and 5th graders	(4) Levels Student FE 3rd, 4th and 5th graders (without lags)	(5) Gains Student FE 4th and 5th graders
Student characteristics that are constant over time [b]	Yes	Yes	Yes	No	No
Student characteristics that vary over time [b]	Yes	Yes	Yes	Yes	Yes
Lagged reading score	0.683** (0.002)	0.679** (0.002)	--	---	---
Teacher characteristics					
Male	-0.011** (0.003)	-0.011** (0.003)	-0.009** (0.003)	-0.015** (0.003)	-0.010 (0.006)
White (base)	--	--	--	--	--
Black	0.001 (0.003)	-0.000 (0.003)	0.001 (0.003)	0.003 (0.002)	0.006 (0.006)
Hispanic	0.014 (0.015)	-0.009 (0.016)	-0.014 (0.017)	-0.002 (0.014)	-0.018 (0.035)
Other race	-0.045** (0.010)	0.006 (0.011)	0.006 (0.012)	0.035** (0.009)	0.024 (0.021)
Same race as student	0.008** (0.002)	0.007** (0.002)	0.010** (0.002)	0.013** (0.002)	0.020** (0.005)
Same gender as student	-0.006** (0.002)	-0.006** (0.002)	-0.005* (0.002)	-0.001 (0.002)	-0.004 (0.005)
Teacher credentials					
No experience (base)	--	--	--	--	--
1-2 years	0.042** (0.004)	0.041** (0.004)	0.038** (0.004)	0.032** (0.003)	0.043** (0.007)
3-5 years	0.057** (0.004)	0.056** (0.004)	0.049** (0.004)	0.046** (0.003)	0.064** (0.008)
6-12 years	0.067** (0.004)	0.063** (0.004)	0.057** (0.004)	0.053** (0.003)	0.072** (0.007)
13-20 years	0.074** (0.004)	0.070** (0.004)	0.065** (0.004)	0.062** (0.003)	0.082** (0.008)
21-27 years	0.084** (0.004)	0.080** (0.004)	0.075** (0.004)	0.067** (0.003)	0.096** (0.008)
28+ years	0.083** (0.004)	0.078** (0.004)	0.071** (0.004)	0.063** (0.004)	0.090** (0.009)
Regular license (base)	--	--	--	--	--
Lateral entry	0.027 (0.014)	0.039** (0.014)	0.044** (0.015)	0.010 (0.014)	0.048 (0.028)
Interact continuing/lateral	-0.022	-0.008	-0.022	-0.019	-0.021

entry	(0.012)	(0.012)	(0.012)	(0.012)	(0.023)
Other license	-0.012**	-0.008*	-0.012**	-0.017**	-0.024**
	(0.003)	(0.003)	(0.004)	(0.003)	(0.007)
Graduate degree	-0.009**	-0.007**	-0.007**	-0.004**	-0.008*
	(0.002)	(0.002)	(0.002)	(0.001)	(0.004)
National Board Certified	0.014**	0.018**	0.016**	0.012**	0.012
	(0.005)	(0.005)	(0.005)	(0.004)	(0.010)
Undergrade institution non-competitive (base)	--	--	--	--	--
Competitive	-0.000	0.000	-0.001	0.001	0.004
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Very Competitive	0.004	-0.003	-0.005	-0.007**	-0.001
	(0.003)	(0.003)	(0.003)	(0.002)	(0.005)
Unranked	-0.015**	-0.012*	-0.012*	-0.007	-0.014
	(0.005)	(0.005)	(0.005)	(0.004)	(0.011)
Mean teacher test score	0.007**	0.005**	0.004**	0.003**	0.004*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Classroom characteristics					
Class size	-0.002**	-0.002**	-0.002**	-0.002**	-0.004**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Percent nonwhite	-0.010*	-0.003	0.026**	0.004	0.021
	(0.005)	(0.009)	(0.009)	(0.007)	(0.018)
Percent subsidized lunch	-0.009	-0.031**	-0.026**	-0.001	-0.009
	(0.006)	(0.007)	(0.007)	(0.007)	(0.017)
Percent college grad (base)	--	--	--	--	--
Percent hs grad	-0.032**	-0.050**	-0.033**	-0.016	-0.037*
	(0.005)	(0.008)	(0.007)	(0.006)	(0.016)
Percent hs dropout	-0.031**	-0.043**	-0.002	-0.001**	0.051
	(0.011)	(0.014)	(0.015)	(0.012)	(0.032)
Lagged class average reading score	0.060**	0.021**	--	--	--
	(0.004)	(0.005)			
Constant	0.478**	0.505**	0.166**	0.004	0.007
	(0.016)	(0.017)	(0.015)	(0.007)	(0.018)
R squared	0.6851	0.6885	0.0432	0.8969	0.4862
Observations	1,096,478	1,096,478	1,096,516	1,814,704	1,096,724

Notes.

a. a. Based on a panel data set of 3rd, 4th and 5th graders from 1995 to 2004. Dependent variable is student achievement in each year. FE stands for fixed effect; Robust standard errors are in parentheses; ** signifies statistical significance at the 0.01 level and * at the 0.05 level.

b. The coefficients and standard errors for these variables are in appendix table 2.

Table 4. Effects on achievement: subject teacher vs. baseline teacher

Baseline teacher	Subject teacher *(weak credentials)*	Difference in achievement (lower and upper bound estimates)[a]			
		Math		Reading	
		low	high	low	high
10 years of experience	No experience	-0.079	-0.094	-0.053	-0.072
Competitive undergraduate college	Non competitive undergraduate college	-0.007	-0.010	*	*
Regular license	Other license	-0.033	-0.059	-0.017	-0.024
Licensure test score is average	Licensure test is 1 SD below the average	-0.011	-0.015	-0.003	-0.004
Graduate degree	No graduate degree	*	*	+0.004	+0.008
National Board Certified	Not National Board Certified	-0.020	-0.028	-0.012	-0.012
Total difference		-0.150	-0.206	-0.081	-0.120
a. Lower bound estimates are from model 4 and upper bound estimates are from model 5 in Tables 3 and 4. *signifies coefficient is not statistically significant					

Table 5. Achievement effects of graduate degrees[a]				
	Model 4		Model 5	
	Math	Reading	Math	Reading
Basic model				
Graduate degree	-0.003 (0.002)	-0.004** (0.001)	+0.002 (0.004)	-0.008** (0.004)
Disaggregated by degree				
Master's degree	-0,002 (0.002)	-0.003* (0.001)	0.003 (0.002)	-0.007* (0.004)
Advanced degree	-0.045** (0.012)	-0.025** (0.010)	-0.052* (0.025),	-0.048* (0.020)
Ph.D.	-0.093** (0.023)	-0.031 (0.019)	-0.078) (0.056)	-0.021 (0.064)
Master's by time				
MA before teaching	-0.001 (0.003)	-0.005 (0.003)	0.009 (0.007)	-0.009 (0.006)
MA 1-5 years into teaching	0.004 (0.003)	0.001 (0.002)	0.007 (0.006)	-0.005 (0.005)
MA degree 5+ years into teaching	-0.010** (0.003)	-0.007** (0.002)	-0.008 (0.006)	-0.010* (0.005)
Notes. a. The entries in the first row are identical those reported for graduate degree in Tables 2 and 3 for models 4 and 5. Subsequent entries are based on those same full models but with the graduate degree variable replaced first by the three disaggregated degree variables and then by the three master's-by-time variables. Robust standard errors are in parentheses; ** signifies the coefficient is statistically significant at the 0.01 level and * at the 0.05 level.				

Table 6. National Board Certification				
	Model 4		Model 5	
	Math	Reading	Math	Reading
NBCT –2	0.024** (0.008)	0.026** (0.006)	0.055** (0.019)	0.038** (0.014)
NBCT-1	0.018** 0.008)	0.016** (0.006)	0.061** (0.017)	0.026* (0.013)
NBCTcurrent	0.018** (0.007)	0.016** (0.006)	0.046** (0.016)	0.035** (0.013)
NBCT post	0.022** (0.005)	0.0116** (0.004)	0.041** (0.010)	0.023** (0..008)

Notes.
These results are based on the full models reported in Tables 2 and 3 with the NBCT variable in those equations replaced by the four NBCT variables listed here. NBCT-2 (or –1) takes on the value 1 for two years (or 1 year) before the teacher is certified. NBCTcurrent takes on the value 1 for a teacher the year she is certified. NBCT post takes on the value 1 for any year after a teacher is certified.. Robust standard errors are in parentheses; ** signifies the coefficient is statistically significant at the 0.01 level and * at the 0.05 level.

Table 7. Teacher test scores[a]				
	Model 4 (levels)		Model 5 (gains)	
Basic model	Math	Reading	Math	Reading
Test score	0.011**	0.003**	0.015**	0.004*
	(0.001)	(001)	(0.003)	(0.002)
Non-linear average test score (SDs)				
>= 2	0.032**	0.008*	0.068**	0.002*
	(0.010)	(0.007)	(0.025)	(0.009)
1.5 to 2	0.012**	0.004	0.016	0.002
	(0.005)	(0.004)	(0.011)	(0.008)
1- to 1.5	0.022**	0.011**	0.026**	0.009
	(0.003)	(0.002)	(0.007)	(0.006)
0.5 to 1	0.008**	0.001	-0.007	-0.003
	(0.002)	(0.002)	(0.005))	(0.004)
-0.5 to 0.5 (base)	--	--	--	--
-0.5 to - 1	-0.003	0.003	-0.008	-0.004
	(0.003)	(0.002)	(0.006)	(0.005)
-1 to -1.5	-0.012**	-0.005	-0.017*	-0.010
	(0.003)	(0.003)	(0.008)	(0.007)
-1.5 to -2	-0.022**	-0.011**	-0.024*	-0.016
	(0.005)	(0.004)	(0.006)	(0.010)
<= -2	-0.042*	-0.017*	-0.062**	-0.010
	(0.008)	(0,007)	(0.019)	(0.016)
Test scores by type				
Elementary	0.012**	0.004**	0.010**	0.003
	(0.001)	(0.001)	(0.002)	(0.002)
Content	0.006**	0.002	0.014**	0.001
	(0.001)	(0.001)	(0.002)	(0.003)

Notes.

a. The entries in the first row are identical those reported for test scores in Tables 2 and 3 for models 4 and 5. Subsequent entries are based on those same full models but with the single variable replaced first by the series of indicator variables and then by the two types of tests. Robust standard errors are in parentheses;** signifies the coefficient is statistically significant at the 0.01 level and * at the 0.05 level.

Table 8. Teacher experience [a]				
	Model 4		Model 5	
	Math	Reading	Math	Reading
With interactions				
Teacher will stay 3+ years	-0.019** (0.007)	-0.009 (0.006)	-0.033* (0.016)	-0.022 (0.014)
1-2 years	0.059** (0.006)	0.029** (0.005)	0.061** (0.014)	0.032** (0.012)
Interact 1-2 years with stay 3+ years	0.002 (0.008)	0.007 (0.007)	0.023 (0.018)	0.020 (0.015)
a. Based on the full models reported for model 4 and 5 in Tables 2 (math) and 3 (reading), with the sample modified to test scores for students taught by recent young teachers for whom we do not have information on whether they remained in teaching for 3 or more years. . Robust standard errors are in parentheses; ** signifies coefficient is significant at the 0.01 level and * at the 0.05 level.				

Table A1: Number of students with math and reading scores, grades 3-5, 1995-2004.

Math	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
3rd	88,901	90,505	94, 872	98,838	101,006	101,585	102,152	100,732	102,543	101,747
4th	88,227	89,174	91,416	94,492	98,398	100,003	100,394	101,095	100,683	103,059
5th	86,146	*	90,250	91,917	95,240	98,562	100,219	101,121	103,864	102,656
Reading										
3rd	88,820	90,444	94,614	98,383	100,509	101,075	101,635	100,090	102,119	101,303
4th	88,161	89,168	91,223	94,102	97,918	99,460	99,715	100,399	100,284	102,580
5th	88,136	*	90,092	91,557	94,776	98,101	99,630	100,291	103,431	102,174

*test score records are missing for 5th grade students in 1996.

Table A2. Student characteristics for math achievement models [a]					
	(1) Levels No FE 4th and 5th graders (with lags)	(2) Levels School FE 4th and 5th graders (with lags)	(3) Gains School FE 4th and 5th graders	(4) Levels Student FE 3rd, 4th and 5th graders (without lags)	(5) Gains Student FE 4th and 5th graders
Student characteristics					
Male	0.010** (0.002)	0.010** (0.002)	0.004* (0.002)	--	--
Black	-0.086** (0.002)	-0.098** (0.002)	0.029** (0.002)	--	--
Hispanic	0.075** (0.004)	0.065** (0.004)	0.074** (0.004)	--	--
Other race	0.038** (0.004)	0.045** (0.003)	0.063** (0.004)	--	--
Age in grade 3	-0.004** (0.000)	-0.004** (0.000)	-0.003** (0.000)	--	--
Parents are college graduates	base				
Parents are high school graduates	-0.109** (0.001)	-0.112** (0.001)	-0.035** (0.001)	--	--
Parents are high school dropouts	-0.217** (0.003)	-0.223** (0.003)	-0.053** (0.003)	--	--
Limited English	0.010 (0.006)	0.004 (0.005)	0.085** (0.006)	--	--
Gifted	0.282** (0.003)	0.295** (0.003)	0.001 (0.002)	--	--
Special needs	-0.145** (0.002)	-0.151** (0.002)	0.013** (0.002)	--	--
Subsidized lunch	-0.050** (0.001)	-0.052** (0.001)	-0.010** (0.001)	--	--
Repeating grade	0.529** (0.005)	0.490** (0.005)	0.773** (0.004)	0.445** (0.004)	0.877** (0.008)
School change	-0.025** (0.002)	-0.024** (0.002)	-0.002 (0.002)	-0.008** (0.002)	0.000 (0.004)
Structural school change	-0.048** (0.005)	-0.001 (0.006)	-0.035** (0.006)	-0.041** (0.005)	-0.065** (0.010)
Lagged math score	0.710** (0.002)	0.704** (0.002)	--	--	--
aSee Table 2 for the full set of variables in this model.					

Table A3. Student characteristics for reading achievement models [a]

	(1) Levels No FE 4th and 5th graders (with lags)	(2) Levels School FE 4th and 5th graders (with lags)	(3) Gains School FE 4th and 5th graders	(4) Levels Student FE 3rd, 4th and 5th graders (without lags)	(5) Gains Student FE 4th and 5th graders
Student characteristics					
Male	0.030** (0.002)	-0.030** (0.002)	-0.001 (0.002)	--	--
Black	-0.127** (0.002)	-0.131** (0.002)	-0.014** (0.002)	--	--
Hispanic	0.049** (0.004)	0.045** (0.004)	0.061** (0.004)	--	--
Other race	-0.018** (0.004)	-0.004 (0.003)	0.032** (0.004)	--	--
Age in grade 3	-0.003** (0.000)	-0.003** (0.000)	-0.001** (0.000)	--	--
Parents are college graduates	base				
Parents are high school graduates	-0.113** (0.002)	-0.114** (0.002)	-0.027** (0.002)	--	--
Parents are high school dropouts	-0.254** (0.003)	-0.254** (0.003)	-0.059** (0.003)	--	--
Limited English	-0.055** (0.006)	-0.062** (0.006)	-0.086** (0.006)	--	--
Gifted	0.229** (0.003)	0.240** (0.003)	-0.045** (0.002)	--	--
Special needs	-0.184** (0.003)	-0.192** (0.003)	0.036** (0.002)	--	--
Subsidized lunch	-0.062** (0.001)	-0.062** (0.001)	-0.012** (0.002)	--	--
Repeating grade	0.448** (0.005)	0.419** (0.006)	0.710** (0.005)	0.376** (0.005)	0.822** (0.009)
School change	-0.017** (0.002)	-0.016** (0.002)	0.002 (0.002)	-0.007** (0.002)	0.009 (0.005)
Structural school change	-0.045** (0.004)	-0.019** (0.004)	-0.049** (0.005)	-0.042** (0.004)	-0.081** (0.009)
Lagged reading score	0.683** (0.002)	0.679** (0.002)	--	--	--

a. See Table 3 for the full set of variables in this equation.