

NBER WORKING PAPER SERIES

TEACHER CREDENTIALS AND STUDENT ACHIEVEMENT IN HIGH SCHOOL:
A CROSS-SUBJECT ANALYSIS WITH STUDENT FIXED EFFECTS

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

Working Paper 13617
<http://www.nber.org/papers/w13617>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2007

We thank Aaron Hedlund and L. Patten Priestley for outstanding research assistance and the Spencer Foundation and the Center for the Analysis of Longitudinal Data in Education Research for financial support. This paper is a slightly revised the version that was dated September, 2007. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2007 by Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis
with Student Fixed Effects

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

NBER Working Paper No. 13617

November 2007, March 2008

JEL No. I21,J45

ABSTRACT

We use data on statewide end-of-course tests in North Carolina to examine the relationship between teacher credentials and student achievement at the high school level. The availability of test scores in multiple subjects for each student permits us to estimate a model with student fixed effects, which helps minimize any bias associated with the non-random distribution of teachers and students among classrooms within schools. We find compelling evidence that teacher credentials affect student achievement in systematic ways and that the magnitudes are large enough to be policy relevant. As a result, the uneven distribution of teacher credentials by race and socio-economic status of high school students -- a pattern we also document -- contributes to achievement gaps in high school.

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

Jacob L. Vigdor
Terry Sanford Institute of Public Policy
Box 90312 Duke University
Durham, NC 27708
and NBER
jacob.vigdor@duke.edu

Helen F. Ladd
Sanford Institute of Public Policy
Box 90245 Duke University
Durham, NC 27708
hladd@duke.edu

Introduction

Nearly all observers of the education process, including scholars, school administrators, policy-makers, and parents, point to teacher quality as the most significant institutional determinant of student achievement. At the same time, remarkably little is known about the relationship between teacher credentials and teacher quality, or about the policy levers that might be used to raise the quality of teachers and to ensure an equitable distribution of high quality teachers across schools and classrooms. This lack of knowledge is particularly troubling in light of the achievement-related accountability pressures on individual schools associated with both state-level accountability programs and the federal No Child Left Behind Act (NCLB) of 2001 that applies to schools across the country.

Though NCLB focuses primarily on the basic skills of reading and math in grades 3-8, policy makers are increasingly turning attention to the higher order skills taught in high school. This new attention to student achievement and other student outcomes at the high school level reflects the economic and political reality that even minimal participation in the economic and political life of an increasingly global and knowledge-based world requires high school skills.

In light of the availability over time of administrative test data in some states or districts for students in grades 3-8, it is not surprising that much of the recent research on the achievement effects of teacher credentials is based on students in those grades (Clotfelter, Ladd & Vigdor, 2006a, 2007a; Goldhaber & Anthony, 2007; Rockoff, 2004). In this paper, we shift the focus to high schools. At the high school level, most of the existing knowledge about the achievement effects of teacher credentials emerges

from studies based on national surveys such as the National Educational Longitudinal Survey (NELS) of 1988, the Baccalaureate and Beyond Longitudinal Study, and the Longitudinal Study of American Youth (Ehrenberg and Brewer, 1994; Monk, 1994; Monk and King, 1994, Goldhaber and Brewer, 1997b and 2000) and are somewhat dated. Though such panel data sets are useful in that they allow for value-added modeling and they include a rich set of student and teacher characteristics, the teacher credentials are self identified and are not always comparable across states; the test results included in such surveys are not linked to the specific curricula that the teachers are hired to teach; and it is difficult to control fully for the non-random sorting of teachers and students that can bias the results (Goldhaber, 2004; Goldhaber and Brewer, 1997a). An alternative is to turn to a state administrative data set, such as the rich data set on teachers and students available for North Carolina.

In contrast to most other states, North Carolina has long had a standard course of study at the high school level that culminates in end-of-course (EOC) tests in each of a number of subjects, such as English, algebra and biology. This fact makes it well suited for studying achievement at the high school level. For this research, we measure student achievement by test scores on the five EOC tests typically taken by North Carolina students in either the ninth or the tenth grades. Those test scores are matched with detailed administrative data on teacher characteristics and credentials. As we document below, we find compelling evidence that teacher credentials affect student achievement at the high school level in systematic ways that are large enough to be relevant for policy. As a result, the uneven distribution of teacher credentials by race and socio-economic status of high school students – a pattern we also document below – means that minority

students and those with less well educated parents do not have equal access to a high quality education at the high school level.

In addition to its substantive contributions to the literature on the causal linkages between the credentials of high school teachers and student achievement, this paper makes a methodological contribution by its use of student fixed effects in the context of a model estimated across subjects rather than, as is more typical in this literature, over time. The use of student fixed effects, whether in longitudinal studies or in a cross-subject study of this type, is advantageous because it mitigates one of the most serious statistical problems associated with the measurement of teacher effectiveness, namely the fact that teachers are not randomly distributed across classrooms, and hence across students.

In the following section, we set the stage by describing the policy context. Subsequent sections explain and justify the empirical framework, describe the data, and present the results. The paper concludes with a discussion of policy implications.

Background and Policy Context

We focus on teacher credentials because they are potentially important policy levers. All states currently impose various types of licensure requirements that affect who is allowed to teach. Further, the uniform salary schedules used by most states and districts attach financial rewards to certain credentials, namely years of experience and graduate education. Many states, including North Carolina, encourage their teachers to apply for National Board Certification. Underlying the analysis in this paper is the assumption that policy makers can make use of information on how teacher credentials of various types are linked with student achievement to promote policies designed to attract teachers with

the relevant credentials, to induce teachers to obtain those credentials, and to design mechanisms to assure that teachers, as defined by their credentials, are equitably distributed across schools.

To the extent, however, that teacher credentials are only weakly linked to student achievement, as some researchers believe to be the case, credentials would not be very powerful or useful policy levers for affecting the level and distribution of student achievement. Indeed some researchers and observers believe that teacher credentials are such poor predictors of student achievement that much of the current apparatus for preparing and credentialing teachers should be eschewed in favor of a new system in which teachers are hired (and fired) based not on their credentials but rather on their cognitive ability and their effectiveness in the classroom (Walsh, 2001).

The policy debate is lively and intense. On one side is the report of the National Commission on Teaching and America's Future (National Commission, 1996) that documents the high incidence of "unqualified" teachers and indicts the country's system of teacher training and licensure for not setting high enough standards and for failing to enforce the existing standards. On the other is the 2001 Report of the Abell Foundation (Walsh, 2001), which, in a harsh review of the literature on teacher credentials, argues for Maryland to deregulate its teacher licensing system. But opposing that position is a well-documented rebuttal by Linda Darling-Hammond (2002). Adding fuel to the fire is a recent empirical paper by Kane, Rockoff & Staiger (2006), who document that the classroom performance of teachers in New York City during their first two years of teaching is a more reliable indicator of a teacher's future effectiveness than the teacher's certification status.

In addition to the general debate about the desirability of teacher licensure and credentialing requirements, the research literature has focused on specific credentials that are currently growing in policy importance. One such credential is National Board Certification, which is available to teachers who successfully complete a rigorous application process (Goldhaber & Anthony, 2007; Ladd, Sass, & Harris, 2007; Cavaluzzo, 2004). Another is various forms of alternative entry into the teaching profession, including, for example, Teach for America, New York's Teaching Fellows, and a variety of state-sanctioned "lateral entry" programs (Boyd et al, 2006, Glzerman, Mayer & Decker, 2005).

The research in this paper builds on our prior work on teacher credentials at the elementary school level in North Carolina (Clotfelter, Ladd & Vigdor, 2006 and 2007). North Carolina is well suited to research at the elementary level because it has been testing students in grades 3-8 in math and reading since the early 1980s and these tests are matched to the state's standard course of study. Further, the state data on students and teachers are available to researchers in forms that permit the matching of students over time, and in many cases, the matching of students to their specific teachers. Our research on teachers in grades four and five documents not only that teacher credentials matter for student achievement at the elementary level, but also that are distributed in highly inequitable ways across schools (Clotfelter, Ladd & Vigdor, 2006a, 2007a and ; Clotfelter, Ladd, Vigdor & Wheeler, 2007).

North Carolina also serves as an excellent site for the study of teacher credentials at the high school level. Although many states now administer tests at the high school level, most of those tests are in the form of comprehensive high school exit exams or

minimum competency exams. Whatever the merits of such tests in assuring that students meet some specified level of achievement before they graduate, the tests are not very useful for examining the effectiveness of teachers. The main problem is that student outcomes cannot be attributed to the performance of particular teachers because the material covered on such tests goes well beyond that covered in a specific course. In addition, such tests can shed little light on how effectively teachers succeed in conveying high school level material because the material covered is often at a relatively low level -- one more appropriate to the middle school than to the high school. What is needed, instead, are tests that are external to the school, that relate to the material that teachers are hired to teach, and that the students are likely to take seriously. North Carolina is one of the few states that have had such tests at the high school level for many years.¹

Empirical Framework

The biggest challenge facing any study of the causal effect of teacher credentials on student achievement is the potential for bias that arises because students and teachers are not randomly assigned to classrooms. To the extent that teachers with stronger credentials are assigned to the classes with unobservably more able students, for example, a cross-section analysis that failed to address that assignment pattern would produce upward biased estimates of the achievement effects of teacher credentials. Alternatively, if policy makers try to compensate for the weakness of low-performing students by assigning them more qualified teachers, any estimates of teacher credential effects that did not take account of that assignment strategy would be subject to a negative bias. The statistical problems associated with this non-random sorting of teachers and students is

¹ For an overview of the use of comprehensive tests and end-of-course tests at the high school level in the South, see SREB, 2007.

exacerbated at the high school level because students have more opportunities to select their courses, and ability-tracking is more prevalent than at the elementary level.

The standard way to address this problem with the use of state administrative data at the elementary level has been for researchers to use longitudinal data that includes outcome measures, such as test scores in math, for each student for multiple years (Clotfelter, Ladd & Vigdor, 2007a and forthcoming; Kane, Rockoff & Staiger, 2006). The availability of multiple measures for each student makes it possible to include in the model student fixed effects and thereby to control statistically for unobservable time-invariant characteristics of students, such as their ability or motivation, that could be correlated with teacher credentials. This within-student estimation addresses the problem associated with the non-random assignment of teachers to students by identifying the effects of teacher credentials only by the within-student variation in teacher credentials during the time period of the data.² That approach is less suited to the high school level where multiple outcome measures for the same subject are not available over time. Nonetheless, similar benefits can be achieved when test scores are available for multiple subjects for the same student.

For the high school level, our starting point is a relatively standard education production function modified to refer to achievement test scores in several subjects taken by each student. Although these subjects could be taken in different grades or years (as is the case in our North Carolina data), we simplify the exposition at this point by ignoring the time dimension and assuming that all the subjects are taken in the same year. Each student i has test scores in multiple subjects, denoted by the subscript s . Since multiple

² The student fixed-effects method does not resolve bias associated with time-varying unobserved determinants of student achievement. As we discuss below, the analogous concern in this study is that some unobserved determinants of student achievement may vary across subjects.

teachers teach each subject, either within or across schools, we include a subscript j to denote the relevant teacher and a subscript k to denote the school.

Letting A_{ijsk} refer to the achievement of student i in subject s taught by teacher j , our preferred model takes the following form:

$$A_{ijsk} = \alpha + T_{ijsk}\beta + \lambda_i + e_{ijsk} \quad (1)$$

where T is a vector of variables that describe teacher j 's credentials and the characteristics of her classroom. Of particular interest for this paper are the teacher's characteristics (such as race or gender) and credentials (such as years of experience, type of license, and licensure test score), but T can also include variables such as the size of the class and the characteristics of the students in the class;

λ_i refers to a set of student specific fixed effects;

e_{ijsk} is a student-specific error term; and

α is constant term and β is a vector of parameters.

The inclusion of the student fixed effects means, as would be the case in longitudinal studies, that the effects of the T variables are estimated within students. In this case, that means they are based only on the variation in teacher credentials across the subjects for each specific student.

One difference from the longitudinal counterpart of this model is worth highlighting. In panel models, at least as they have been estimated with administrative data at the elementary level, education is explicitly modeled as a cumulative process. Because of that cumulative process, one or more lagged achievement variables must be included in the model to account for the achievement that the student brings to the classroom, and the failure to do so appropriately can lead to biased coefficients of the

teacher credentials (see discussion of bias in Clotfelter, Ladd & Vigdor, 2007a). In the context of our cross-subject model, the analogy would be to represent a student's knowledge at the beginning of the term by subject-specific test scores taken prior to the beginning of the instruction period. By not including these initial test scores (which, in any case, are not available), we are, in effect, assuming that a student's initial knowledge in a subject such as geometry is negligible. Any overall ability or achievement level, however, is captured by the student fixed effect.

Equation 1 is equivalent to the following equation:

$$(A_{ijsk} - \bar{A}_i^*) = (T_{ijsk} - T_i^*)\beta + (e_{is} - e_i^*) + (e_{ijsk} - e_{is}) \quad (1a)$$

where the variables with asterisks are the student-specific means of each variable. Thus, a student's achievement in subject s (with teacher j in school k) is measured not in absolute terms but relative to the average of her achievement based on all her tests. Similarly, a teacher's credentials are measured relative to the average credentials of all of the teachers of that student. The term $(e_{is} - e_i^*)$ refers to a student-specific error term that varies across subjects and the term $(e_{ijsk} - e_{is})$ refers to a subject-specific error term that varies with the unmeasured characteristics of the student's teacher in that specific subject.

This model will generate unbiased estimates of β provided that neither of the error terms is correlated with the relative – that is, demeaned -- teacher credentials or with each other. Potentially problematic is the student-specific error term that varies by subject. In the following discussion, we explain why we believe it is reasonable to assume there is little or no correlation between that term and the demeaned teacher credentials.

For the purposes of this discussion it is useful to provide some illustrative substance by referring to e_{is} as the student's ability in subject s . If student ability does not

differ by subject, then the term $(e_{is} - e_i^*)$ would be zero and would generate no statistical problem. If, however, student ability differs by subject, that term could potentially be correlated with the demeaned teacher credentials variable.

Table 1 provides basic evidence on the across-subject correlation in ability levels and course track assignments among high school students. The sample for the table is all North Carolina students who were in the 10th grade in 2002/03 for whom we can match test scores in their English I and Algebra I courses (regardless of the grade in which the student took the particular course), and for whom we have test scores on their eighth grade math and reading tests.³ We interpret the students' eighth grade test scores in math and reading as measures of their ability (or prior achievement) in those two fields, and have divided students into tertiles based on those scores. The relevant question is the extent to which students with high abilities – both absolute and relative – end up in the more advanced high school classes. We distinguish between advanced and regular algebra and English courses based on whether the course is designated as one of several types of advanced course or whether it is a regular course, and look at the probabilities of being in an advanced section.⁴ Our expectation is that the patterns across absolute ability tertiles will be much clearer than those across tertiles based on relative ability, where a student's relative ability in math or reading is measured as the difference between her test score in that subject and her average test score in the two subjects.

The table entries are the probabilities that students of different absolute and relative ability levels in math and reading are in advanced algebra and advanced English

³ See below for additional discussion of the data. If a student took one of the end-of-course tests in eight grade, the math and reading scores are based on seventh grade end-of-grade tests.

⁴ A course is classified as an advanced class if it is designated as an honors course, an advanced course or a course for academically gifted students.

course. In line with our expectation that advanced class assignments are based on average – not relative – ability, the top two panels indicate a strong positive correlation between absolute ability, as measured either by math or by reading scores, and the probability of being in an advanced algebra or English class. Moreover the patterns (although not the levels) across students grouped by their math ability are strikingly similar for algebra and English. As shown in the final column, the probability that a high-ability math student will enroll in either an advanced algebra course or an advanced English course is about 2 1/3 times the probability that a student with low math ability will enroll in such courses. With respect to reading ability, the positive correlations are again strong, but this time a bit stronger for being in an advanced English course. At the same time, reading scores are even better predictors than math scores of algebra placement. Hence, the data support the notion that schools consider student ability to be single dimensional.

Consistent with the data in the top panel, the bottom two panels of Table I show that the correlations are far less evident when students are grouped by their relative abilities. In particular, those with high ability in math relative to reading or high ability in reading relative to math are no more likely to be in an advanced algebra class than those with low math or reading ability. We cannot rule out, however, some selection by relative ability into advanced English classes. Students with higher relative ability in reading are slightly more likely to be in an advanced English class (compared with those with low relative ability) and those with higher relative ability in math are slightly less likely to be in an advanced English class. Even this limited evidence of sorting by relative ability into advanced English courses would create a problem for our analysis only if teachers were sorted across classrooms in a systematic way.

A more direct test would look directly at the relationship between relative teacher credentials and relative student ability. The results of such a test are reported on Table 2. For the purposes of this test, we focus on a single characteristic of teachers, namely the average test score on their licensure exams that research, including research reported below, shows to be predictive of student achievement. The table reports results for four regressions, one for each of the four cohorts of students included in our analysis as well as for our entire sample (which is described below). The dependent variable in each regression is the difference between the average licensure test score of the i th student's high school algebra and English teachers. The explanatory variable of primary interest is the student's relative ability as measured by the difference between her eighth grade math score and her eighth grade reading score. Also included in each regression are a constant term and school fixed effects. Thus we are testing the null hypothesis of no relationship between the student's relative ability in math and reading (as a proxy for the first component of the error term in equation 1a) and the relative qualifications of her high school algebra and English teachers (a proxy for the dependent variable in equation 1a).

Because the regression reported in the final column is based on the largest of the five samples, it generates the smallest standard error for the key coefficient and hence is the most likely to generate a statistically significant coefficient that would allow us to reject the hypothesis of no relationship. As reported in the table, in none of the five regressions does a statistically significant relationship emerge between the relative credentials of the teacher by subject and the student's ability in math relative to reading.⁵

⁵ If school fixed effects are excluded, one coefficient, that for the key explanatory variable in cohort 3 is significant, but only at the 0.10 level. Note that even the results in the final column of the table do not permit us to rule out a relationship between student and teacher relative ability as high as 0.014 (= the estimated coefficient plus 2 standard errors) but even that correlation is extremely small.

Hence, the data provide little or no reason to question the basic assumption that the subject-related individual error term is uncorrelated with teacher credentials. .

The other error term in equation 1a , ($e_{ij|sk}-e_{is}$), accounts for the effects on student achievement of unmeasured characteristics of teachers, such as their motivation or effort. This term will not bias the coefficients of interest if teacher effort is randomly distributed among teachers with any given set of credentials. It only creates a problem if, for any given set of teacher credentials, teacher effort varies in a systematic way with the unmeasured characteristics of the students in the class. Once again the presence of the student fixed effects goes a long way toward mitigating any potential bias since any problematic correlations must be between unmeasured subject-specific student characteristics and unmeasured subject-specific teacher characteristics.

Thus, although we cannot prove conclusively that our analysis is completely free from bias, the logic and evidence presented here gives us confidence in the approach. The actual situation in North Carolina high schools, of course, is more complicated than suggested by equations 1 or 1a . As a result in the empirical work additional variables are needed to control for unusual situations such as students choosing to take standard courses earlier or later in their educational career than is typical.

One consequence of estimating a model with student fixed effects is that we are not able to include in the model any characteristics of students that do not vary across subjects such as their gender or race, or their prior test scores in basic skills such as math or reading. Any such subject invariant student characteristics disappear from the model when the variables are demeaned. Nonetheless, for the purposes of comparison of effect

sizes it may be useful to have rough estimates of how student characteristics affect achievement. Hence, in addition to model 1, we estimate model 2:

$$A_{ijsk} = \alpha + T_{ijsk}\beta + X_i \delta + \eta_k + e_{ijsk} \quad (2)$$

where X refers to student characteristics and

η_k refers to school, rather than student, fixed effects.

Although the school fixed effects mitigate the bias in the estimates of the β coefficients associated with the non-random matching of students and teachers across schools (provided that the unmeasured effects enter the equation linearly), they do not address the nonrandom matching of teachers and students across classrooms within schools, which is why we prefer model 1. Nonetheless, there are advantages to including student characteristics. As we document below, by far the most important of the student-specific variables are the student's eighth grade test scores in math and reading, which serve as proxies of student ability and motivation.

The North Carolina Data

North Carolina has long had a standard course of study for students in all grades, including those in high school. Moreover, since the early 1990s it has administered statewide end-of-course (EOC) tests at the high school level. Though EOC tests are given in multiple subjects, we restrict our analysis here to the five subjects that are typically taken by students in the 9th and 10th grades. These include (algebra; economic, legal and political systems (ELP)⁶; and English I which are typically taken in the 9th grade and geometry and biology which are typically taken in the 10th grade. Many

⁶ The ELP course has recently been restructured and renamed Economics and Civics. No EOC test was given either for ELP or for Economics and Civics in 2005.

students however, take the relevant courses in other grades, either before or after the typical year.⁷ The EOC test scores are high stakes for students in that they count for 25 percent of the student's grade in the course.⁸

We are working with four cohorts of 10th graders – those who were in tenth grade in 1999/2000; in 2000/01 in 2001/02 and in 2002/03. By selecting these cohorts, we allow each student in any of the cohorts the opportunity to take any one of the five tests. Since our data end in 2004/05, any student in 10th grade in 2002/03 would still have two more years to take the test. For the same reason, our earliest cohort allows us to go back in time so that we can include the students within the cohort who took EOC tests in middle school. Furthermore, by restricting the analysis to 9th and 10th grade tests, we minimize attrition related to dropping out of school in grades 11 and 12 and also keep to a minimum any confounding factors related to the selection by students into advanced courses.

The final sample includes only those students for whom we could match at least three teachers to the EOC tests. The percentages of all students with matched teachers taking at least one EOC test who meet this criterion by cohort are 72.6; 77.3, 76.1, and 73.2. (The comparable percentages for cohorts outside of our sample is 62.1 percent in 1999 and 68.7 percent in 2004.) The appendix describes our method of linking students to teachers and includes information on the samples.

⁷ North Carolina has four courses of study: Career Prep, College Tech Prep, College/University Prep, and Occupational. We believe that most of the students in our sample are in either of the two college tracks, although some could be possibly be in the Career Prep track.

⁸ Currently, students are not required to pass the exams to graduate. Beginning with the class of 2010, North Carolina students will be required to pass end-of-course exams in Algebra I, biology, civics and economics, English I and U. S. history to graduate.

In all cases, we have normalized the EOC test scores by grade and by year, with mean zero and standard deviation equal to one. This normalization means that the coefficients can be interpreted as fractions of a standard deviation.

Achievement Effects of Teacher Credentials

The main results for teacher credentials for models 1 and 2 are reported in Table 3. These results are a subset of the full set of results, which for model 1 include those reported in Table 4, and for model 2 include those reported in both Tables 4 and 5. Both models also include subject-by-grade fixed effects and model 2 includes cohort fixed effects. The subject-by-grade effects control for the fact that not all students take a particular course in the typical grade for that course. The cohort effects in model 2 control for changes over time, such as accountability pressures, not captured by other variables.

The entries in the table are the estimated coefficients of seven sets of teacher credentials, with standard errors in parentheses. Two asterisks signify that the coefficient is statistically significant from zero at the 0.01 level. As discussed above, Model 1 generates the preferred results. The observation that most of the estimated coefficients of the teacher credentials in that model are slightly smaller than those in model 2 is consistent with the more advantaged teachers being matched with the more able students. In the following discussion we refer mainly to the preferred results in the first column.

Years of 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 is based on all the years of teaching, whether in North Carolina, or elsewhere, for which the state has given the teacher credit.

Because of our own prior research at the elementary level and that of others (e.g. Hanushek, Kain, O'Brien and Rivkin, 2005), we expect the effects of 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.

As reported in column 1, most of the gains in achievement associated with teacher experience occur in the first two years of teaching with an effect size of 0.0503. Though the estimated coefficients rise to a peak of 0.0617 for a teacher with 21-27 years of experience, none of the coefficients for additional years of experience differ statistically from the coefficient for 1-2 years. Thus we conclude that novice teachers in the sample are less effective than teachers in the sample with some experience, but beyond the first couple of years, more experienced teachers are no more effective than those with a couple of years of experience.

One interpretation of this pattern is that there is little or no additional learning on the job after the first few years in teaching. Another is that teachers continue to learn on the job but, because the more effective teachers either leave the profession, or in this case stop teaching the basic ninth and tenth grade courses, at higher rates than less effective teachers, the typical teacher with more experience who continues to teach a core high school course is no more effective than one with a few years of experience. We examine these two interpretations in Table 3A. For purposes of comparison, the first two columns replicate the experience results from models 1 and 2 of the previous table. The second two columns differ in that they are based on comparable models that include teacher fixed effects. Thus, the coefficients in these columns factor out the losses in

average effectiveness that occur because of the departure of the more effective teachers from these courses. Consequently, they should be interpreted as the expected achievement effects of additional experience for a teacher relative to her own effectiveness at a previous period.

The finding that these coefficients rise quite dramatically with years of experience supports the conclusion that teachers who stay on the job continue to become more effective. The 0.0563 difference between the coefficients for a teacher with 6-12 years of experience and one with 13-20 years of experience in column 3 indicates that a teacher with about 16 years of experience raises student achievement by about 0.06 standard deviations more than that very same teacher would have done had she had only about 9 years of experience. Though the patterns in columns 3 and 4 support the case for trying to keep experienced teachers in the core courses, these estimates should not be interpreted as implying that, in general, a very experienced teacher is significantly more effective than a typical teacher with limited experience. As we have already noted, the results in columns 1 and 2 show that is not the case; the departure of the more effective teachers from the core courses largely offsets the salutary effects of experience on teacher effectiveness.

Similar patterns, in which the coefficients rise more steeply in models with teacher fixed effects than in the models without them, emerge at the elementary level for teachers in New York City (Kane, Rockoff, and Staiger, 2006, Table 10). In our own prior research on teacher credentials at the elementary level, we found rising coefficients related to teacher experience even in the absence of teacher fixed effects. For math achievement in grades four and five, for example, our basic estimates (based on models

without teacher fixed effects) relative to the base of no experience range from 0.057 to 0.072 for 1-2 years of experience to peaks of 0.092 to 0.110 for the 21-27 year category (Clotfelter, Ladd & Vigdor, 2007).⁹ Since those estimates do not account for the differential attrition of more effective teachers, one interpretation is that such attrition is less prevalent at the elementary level where teacher attrition typically means leaving the profession than at the high school level, where attrition in the context of this analysis also includes shifting away from the teaching of core courses to higher level courses. We plan to investigate these patterns further in future work.

Teacher test scores

Teacher test scores are among the teacher credentials that most often emerge as statistically significant predictors of student achievement. In his 1997 meta analysis, for example, Hanushek found far more consistently positive results for teacher test scores than for credentials such as years of experience and master's degree (Hanushek, 1997). Positive effects also emerge from more recent studies based on state administrative data for elementary schools (Clotfelter, Vigdor & Ladd 2006a; Goldhaber, forthcoming).

Most high school teachers in North Carolina have taken Praxis II tests as part of their licensure requirements. These tests include subject tests that measure knowledge of specific subjects that educators will teach, as well as general and subject-specific teaching skills and knowledge. We normalized test scores on each of the tests separately for each year that 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

⁹ Our basic models in that paper differed somewhat from those presented here and we presented results for both lower and upper bound estimates.

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 the normalized scores.

Our basic specification for teacher test scores is linear. As shown in the Table 3, teacher test scores enter Model 1 with a coefficient of about 0.010, which is only slightly smaller than the coefficients of 0.011 to 0.015 that emerged in our prior research for teachers at the elementary grades. Table 3B reports two alternative specifications of the teacher variable, both of which are embedded in the full model 1 specification. In the second column, we report the coefficients for a more flexible form based on indicator variables for average test scores that are more than one standard deviation above or below the mean, with the base category being test scores within one standard deviation. The results suggest a nonlinear effect of test scores. In particular, the negative effect of having a teacher with an average licensure test score more than one standard deviation below the mean is more than twice (in absolute value terms) the positive effect of having a teacher with an average test score more than one standard deviation above the mean.

In the third and fourth columns of Table 3A, we disaggregate the test scores by subject so that we can determine the extent to which a teacher's knowledge of content and subject-specific pedagogy, as measured by her test results, affects her students' achievement in that specific subject. The fourth column differs from the third in that the equation also includes subject-specific certification variables that are described below.

The reader should note that each of the subject-specific test scores appears only in the observations associated with that subject. Because not all teachers of a specific subject have taken a test in that subject, we include an extra control variable for each subject indicating that the teacher has no test score for that subject. Of most interest are

the coefficients on the normalized scores on the relevant test, and for the average of the teacher's other normalized test scores. Because there is no specific test for algebra or geometry, we use the high school math test as the relevant test for both those subjects.¹⁰

The clearest findings emerge for math. A one standard deviation difference in a teacher's math test score is associated on average with a 0.03 standard deviation difference in student achievement in either algebra or geometry. The teacher test scores in biology are predictive of student achievement in biology, but the coefficients are smaller than those for math, and teachers' abilities or knowledge as measured by their nonbiology test scores enter positively as well. The negative signs for the ELP and English test scores are unexpected. For ELP, the negative sign could potentially reflect the fact that the licensure test we used for this analysis is a general social studies test that applies to a broad range of social sciences rather than one specifically related to the course material. A related explanation may apply to English given that the English test is designed for a variety of courses that are more advanced than English I.

We find for all subjects that teachers who have no reported test score in the subject they are teaching are slightly less effective than those who did not take the relevant subject-specific test (although not all the coefficients are statistically significant). The negative coefficients do not mean that taking the test makes a teacher more effective; more likely, it suggests that taking the test provides a signal of interest or training in the subject.

¹⁰ The relevant tests are as follows: Biology: 0230 through 1993, 0231,0233 &0234 through 1999,0234&0235 beginning in 2000; English: 0040 through 1993,0041,0042 &0043 through 1999,0041&0043 beginning 2000; math: 0060 through 1993, then 0061 & 0065. There is no specific test for Economic, Legal and Political Systems (ELP). Instead we used the Social Studies tests: 0080 through 1993, 0081,0082 &0083 through 1999, 0081 &0084 beginning 2000.

To summarize, our findings indicate that teacher test scores are predictive of student achievement and that teacher test scores in math are particularly important for student achievement in algebra and geometry. This latter finding is consistent with studies by Monk (1994) and Monk and King (1994) who find, using national survey data, that teacher preparation in math has positive effects on student test scores in math, but that preparation in other subjects does not translate into higher student achievement in those subjects.

Licensure type and certification by subject

Like other states, North Carolina requires that teachers be licensed to teach in public schools. Such licensing is presumably intended to protect the public from poor hiring decisions, but does not by itself assure a high quality teaching force (Goldhaber and Brewer, 2000). We have divided teacher licensure in North Carolina into three categories: regular, lateral entry, and “other.” Regular includes both initial and continuing licenses and represents the base, or left-out, category. Teachers are granted an initial license after completing a state-wide approved teacher preparation program, performing at least 10 weeks of student teaching, and earning passing scores on applicable Praxis II tests. Teachers are granted a continuing license after three years of successful teaching as an initially licensed teacher. Though they are a traditional component of state teacher policies, such licensure requirements are under attack nationally from some quarters either for not being predictive of effective teachers or for imposing unnecessary barriers to entry into teaching (Walsh, 2001; Ballou and Podgursky, 1998).

Many states have responded to such criticisms, or simply to the need for more teachers, by offering alternative routes into the teaching profession that require less up-

front commitment of time. The primary form of alternative entry in North Carolina is the lateral entry program. Lateral entry licenses are issued to individuals who have at least a bachelor's degree and the equivalent of a college major in the area in which they are assigned to teach. Such teachers must affiliate with colleges and universities with approved teacher education programs to complete prescribed class work and must complete at least 6 semester hours of coursework each year. The first lateral entry license is issued for two years, and may be renewed for a third year. In the empirical models, we distinguish between teachers who at the time we observe them have a lateral entry license and those who had such a license in a prior year.

As shown in Table 3, having a teacher with a lateral entry license reduces student achievement by about 0.06 standard deviations compared to having a teacher with a regular license. Prior lateral entrants, however, appear to be no less effective than teachers with a regular license. Though this finding may reflect in part the training that lateral entrants receive during the two years of their license, it also reflects selection. Lateral entrants have high departure rates and it is reasonable to assume that the ones who remain in teaching are more effective than those who depart. The students in our most recent sample cohort were taught by 804 lateral entrants, but by only 155 former lateral entrants.¹¹

Finally, the "other" category includes other forms of alternative entry, as well as provisional, temporary, and emergency licenses.¹² Table 3 indicates that such licenses

¹¹ The 804 lateral entrants were distributed by subject as follows: 226 in algebra, 195 in biology, 132 in ELP, 164 in English and 87 for geometry.

¹² None of these licenses are available in cores grades/subjects after June 2006 due to the regulations under the federal No Child Left Behind Act of 2001.

are associated with a negative achievement effect of -0.0466 which is significantly smaller (at the 5 percent level) than the coefficient for lateral entrants.

Table 3 also shows the effects of certification by subject. Relative to the base case of no certification, being certified in the subject (regardless of the specific subject) is predictive of higher achievement than being certified in a related subject and both of the coefficients are statistically different not only from the base, but also from certification in some other subject and from each other. These results are disaggregated in Table 3C, which in addition to the basic results in column 1, reports the effects of certification by subject area to determine whether the effects of being certified in, for example, math and teaching algebra or geometry differ from being certified in, for example, biology and teaching biology. Columns 2 and 3 both include variables for certification in the specific subject that is being taught, in a related subject, or in some other subject. The results in column 3 differ from those in column 2 in that they are based on models that also include subject-specific test scores. The fact that the entries in columns 2 and 3 are so similar suggests that controlling for subject-specific teacher test scores has little effect on the certification estimates.

Once again the results for teachers of the two math courses, algebra and geometry, stand out. Being certified in math increases student achievement in a math course on average by about 0.12 standard deviations, and being certified in any field also raises achievement in algebra or geometry but by the smaller amount of about 0.05 standard deviations. Certification also matters for biology and for ELP, but interestingly, the relevant certification apparently does not need to be in the specific field. Being certified in English or a related subject is not predictive of student achievement in

English. Interestingly, being taught English I by a teacher who is certified in some unrelated subject has a large negative effect on student performance in English.

National Board Certification

North Carolina has been a leader in the national movement to have teachers certified by the National Board for Professional Teaching Standards (NPPTS), and provides incentives in the form of a 12 percent boost in pay for teachers to do so. Such certification, which requires teachers to put together a portfolio and to complete a variety of exercises and activities designed to test their knowledge of material in their particular field, takes well over a year and is far more difficult to obtain than state licensure.

Following other researchers, we test both for the signaling effect of Board Certification and a human capital effect (Harris & Sass, 2007 and Goldhaber and Anthony, 2007). A positive signaling effect emerges in the form of the positive coefficient of 0.0215 on the variable denoted pre-certification. This variable takes on the value 1 for any teacher who ultimately will become Board Certified. The second Board Certification variable takes on the value 1 in the year in which the candidate for Board Certification is going through the process, and the third variable indicates that the teacher is Board Certified. The finding that the coefficients on the two latter variables are statistically significantly larger than the pre-certification coefficient provides evidence of a positive human capital effect. That is, teachers appear to become better teachers as a result of the Board Certification process. This positive human capital effect did not emerge in our prior research on Board Certification at the elementary level.

Teacher education – advanced degrees and quality of undergraduate institution

In the basic models, we include a single variable to indicate whether a teacher has a graduate degree of any type such as a master's that leads to a higher salary, a Ph.D., or another "advanced" degree including those that do not affect the teacher's salary.

Emerging from Table 3 is the conclusion that having a graduate degree is not predictive of higher achievement compared to having a teacher without a graduate degree.

That finding is examined further in Table 3D, which also reports two variations of the teacher education variable. Variation 1 disaggregates the effects into those for master's, "advanced" and Ph.D degrees. The results indicate a small positive effect of having a teacher with a master's degree and an unexpected and surprisingly large negative effect of having a teacher with a Ph.D. The latter finding is based on a very small number of teachers and may say more about the characteristics of the teachers in this particular sample who have a Ph.D. than the potential effectiveness of teachers with Ph.Ds.

Variation 2 focuses on the teachers with master's degrees. The results indicate virtually no difference between teachers without master's degrees and those who received their master's before entering teaching. However, teachers who received master's degrees after they began teaching appear to be somewhat more effective than those without a master's degree. This pattern differs quite markedly from the pattern that emerged in our previous research on elementary school teachers (Clotfelter, Ladd & Vigdor, 2007a). For teachers in the earlier grades, the earning of a master's degree more than five years into teaching was associated with a negative effect on student achievement. We interpreted that finding to mean that it was the less effective teachers who chose to pursue master's

degrees later in their careers. At the high school level, in contrast, for whatever reason, having a teacher with a master's degree is predictive of higher achievement.

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

Emerging from model 1 in Table 3 is a positive and statistically significant coefficient of 0.0169 for teachers from a very competitive college and a marginally significant smaller coefficient of 0.0047 for teachers from a competitive college. These findings suggests that the quality of a teacher's undergraduate institution is somewhat more predictive of student achievement at the high school level than at the elementary level.

Summary of effects of credentials

To illustrate the cumulative effects of credentials, we compare the predicted achievement effects of teachers with different bundles of credentials. The first three columns of Table 3E describe two set of credentials, a very weak set and a very strong set. In the final column we use the estimated coefficients from prior tables to determine the average differential effect of having a teacher with one bundle rather than the other.

Based on our characterization of a teacher with very weak credentials (one with no experience, low licensure test scores, a lateral entry license, certified but not in the specific subject or a related field, not Board Certified, no graduate degree and from an uncompetitive college), we find that students exposed to such a teacher would be expected to achieve close to 0.30 standard deviations lower than if they had a teacher with the strong set of credentials described in the table. The largest negative effects are associated with the lack of experience, the fact that the teacher is a lateral entrant, and her inappropriate certification

While the entries in Table 3E are useful for demonstrating the relative contributions of the various credentials to differences in teacher effectiveness, the cumulative measure might well be viewed as a misleadingly large estimate of the differential effect of having a teacher with weak credentials. Based on the distribution of teachers as characterized by their predicted effects on student achievement, the teacher with weak credentials in Table 3E would be in the bottom three percent of the EOC teachers in our sample, while the one with very strong credentials would be in the 95th percentile.¹³

An alternative approach is to define a teacher with weak credentials as one at the 10th percentile in the predicted distribution of student achievement and a teacher with strong credentials as one at the 90th percentile. Based on the teachers in our sample, the difference in predicted student achievement between the two teachers is 0.18 standard deviations. Thus, by this metric a student with a weak teacher would be expected to perform 0.18 standard deviations lower than if she had a teacher with strong credentials.

¹³ This statement is based on teachers in the sample for the 2001/02 academic year, but would not differ much for sample teachers in other years.

This smaller figure provides a more conservative, and probably more useful, estimate of the average effect of having a teacher with weak credentials rather than one with strong credentials. We return in the conclusion to the question of whether this difference is large or small.

Achievement Effects of Teacher and Classroom Characteristics

Also included in the full models are characteristics of the teachers such as their race and gender, and characteristics of their classrooms. The estimated coefficients of these variables are reported in Table 4 for both models 1 and 2. The table shows that, controlling for their credentials, male teachers are less effective than female teachers and that black and “other” race teachers are less effective than white teachers.

In terms of classroom characteristics, we find that classrooms with larger percentages of nonwhite students are associated with lower test scores and that those with high average peer achievement or are designated as advanced classes are associated with higher test scores. Consistent with a growing literature on class size – most of which relates to elementary schools -- we find that smaller class sizes are associated with higher student achievement. The effect, however, is small. The coefficient of -0.0026 from model 1 indicates that being in a class with five fewer students than average would increase student achievement by 0.0125 standard deviations.

Perhaps most arresting in this table are the large negative coefficients for black teachers and for male teachers. We examine the achievement effects of race and gender in more detail in Table 4A. That table includes various interactions between the gender or race of the teacher (T) and the gender and race of the student (S). The first column replicates the teacher results for gender and race from the previous table. Since all the

entries in the table are variations of model 1, none of the models on which they are based include the race or gender of individual students. Instead those characteristics are captured by the student fixed effects.

The first variation in Table 4A includes interactions between student and teacher genders. Compared to the base case of a female teacher and a female student, the combination of a male teacher with a female student generates a large negative effect of -0.1069 . In contrast, female teachers appear to be equally effective in teaching male students as they are in teaching female students. Further, male teachers teaching male students are equally effective as female teachers teaching female students. Thus, the large overall negative coefficient for male teachers is driven entirely by the negative interactions between male teachers and female students.

Variation 2 focuses on race. Here the main findings are the large negative coefficients for a black teacher teaching a white student or a Hispanic teacher teaching a non-white or non black student. The latter effect may be spurious because of the small number of Hispanic teachers. The large negative effect associated with black teachers and white students, however, is cause for concern. In contrast to this large negative effect, black teachers appear to be more successful with black students. Although the relevant coefficient for that combination is negative, it is far smaller and not statistically different from the effects of a white teacher/student pair.

Student Characteristics

Table 5 reports the final set of variables that are included in model 2, but not in model 1, namely the characteristics of students and indicator variables for the different cohorts of students. Not surprisingly, eighth grade math and reading scores enter with

large and positive coefficients, signifying that achievement in those subjects is highly predictive of high school performance.

In interpreting the coefficients of the gender and race variables, it is important to bear in mind that we are dealing with a select sample, namely those students who have taken at least three high school end-of-course tests. That means they have not dropped out of school and are most likely in a college track. Within this sample, all else constant, boys perform better than girls, black students perform less well than white students, and students with highly educated parents do better than those with less highly educated parents.¹⁴ Of interest is that, controlling for the other characteristics, Hispanic students achieve at slightly higher levels than white students. This finding is consistent with our previous research on achievement gaps in grades three to eight in which we found that, controlling for various measures of socio-economic status, the achievement of Hispanics who remained in the North Carolina system for all six years of elementary and middle school outpaced that of white students (Clotfelter, Ladd & Vigdor, 2007b). Once again, though, we emphasize that this new finding for Hispanics does not apply to a random sample of Hispanic students.

Policy Implications and Conclusions

For purposes of policy, it would be useful to know whether the estimated effects of the teacher credentials are large or small. As discussed above, a reasonable estimate of the difference in achievement effects of having a weak rather than a strong teacher is about 0.18 standard deviations. One approach to evaluating the policy significance of this

¹⁴ We have not included an income variable because the only income variable available in the data set is whether the student is eligible for free and reduced price lunch. Because many high school students are reluctant to sign up for a subsidized lunch, this variable is not a good measure of family income at the high school level.

magnitude is to compare it to the achievement effects that emerge from other variables in the analysis, such as those for class size or for student characteristics. Given the very small estimated effects for the class size variable in Table 4 (0.0026), having a strong, rather than a weak, teacher appears to be far better for student achievement than being in a classroom with five fewer students rather than one of average size.

Further, the effects of teacher credentials are larger than those of student characteristics (as estimated imperfectly in model 2 and reported in Table 5). For example, consider the effects on student achievement, controlling for eighth grade reading and math scores, of being a black student with parents who are high school graduates with no college compared to being a white student with college educated parents. Using the relevant coefficients from Table 5 (-0.0593 for being black and 0.0571 for having parents who are college graduates), we find a difference in achievement of -0.1164. Thus, having a teacher with strong rather than weak credentials would, on average, more than offset the adverse effect of racial and socio-economic differences as defined in this way.

We conclude that teacher credentials matter in a systematic way for student achievement at the high school level and that the magnitudes are large enough to be policy relevant. Also of potential policy interest, however, is the extent to which the variation in teacher credentials alone explains the variation in overall teacher quality, with overall quality defined in terms of how effective they are in raising student test scores. Based on our estimated equations, the standard deviation of the predicted distribution in student achievement associated with differences in teacher credentials alone is about 0.075. The standard deviation for the distribution of overall teacher quality

is harder to pin down. The typical approach for examining overall teacher quality is to examine the distribution of the teacher fixed effects that emerge from achievement models that replace all teacher credentials with indicator variables for every teacher in the sample. One careful study of the variation in teacher quality at the elementary level in Texas generated estimates of the standard deviations of that distribution that ranged from 0.22 to 0.27. (Hanushek, Kain, O'Brien, & Rivkin, 2005, Table 1 and related discussion). Based on this range, the standard deviation in North Carolina high schools for predicted achievement based only on teacher credentials would be about a quarter to a third of the standard deviation for the overall distribution. Our own rough estimate of the standard deviation of overall teacher quality in North Carolina is closer to 0.51, but is undoubtedly an overestimate because we have made no adjustment for the measurement error in such estimates highlighted by the Texas research team and because the inclusion of the student fixed effects is likely to generate significant noise in the estimates of the teacher fixed effects.¹⁵ Given that the adjustments for measurement error reduce the Texas estimates by about one third, a reasonable upper bound estimate of the standard deviation of the distribution of overall teacher quality in North Carolina is 0.34. Based on that figure, the variation in teacher credentials would account for at least a fifth of the overall distribution in teacher quality.

This discrepancy between the overall variation in teacher quality and that predicted by credentials alone implies that it would be a mistake for policy makers to put so much weight on measurable credentials in determining teacher quality that they ignore other contributors to teacher effectiveness, many of which can only be determined by

¹⁵ Given the technical challenges of estimating models that include both student and teacher fixed effects, we estimated the model with teacher fixed effects for a random subsample of 10 percent of the high schools in our sample.

observation at the school or classroom level¹⁶ Clearly, not all teachers with weak credentials are poor teachers, and, analogously, not all teachers with strong credentials are effective teachers. All the same, the point remains: teacher credentials are important policy levers that are clearly predictive of student achievement.

In light of this conclusion, another policy question relates to how credentials are distributed across schools and students. An uneven distribution indicates that, on average, some types of schools or groups of students are disadvantaged relative to others. In a previous paper (Clotfelter, Ladd, Vigdor & Wheeler, 2007), we grouped all North Carolina high schools into quartiles based on the percentage of low income students they serve and compared various characteristics of teacher across the quartiles.¹⁷ Table 6 summarizes the patterns for five sets of credentials.

The patterns across quartiles of schools depict a consistently disadvantageous situation for students in the high poverty (quartile 1) schools. The first three credentials in the table are defined so that higher percentages indicate weaker qualifications. Thus, the table shows that high poverty schools have higher proportions of inexperienced teachers, of teachers from less competitive institutions, and with non-regular licenses. The final two credentials are defined in the opposite direction. Thus, the fourth and fifth rows show that the high poverty schools have the teachers with the lowest teacher test scores

¹⁶ Some people may want to go further to argue that the best way to evaluate the effectiveness of an individual teacher at the high school level is to look at that teacher's ability to raise test scores. We would not support that policy recommendation. First, measuring value added at the high school level is difficult because of the absence of pre-test scores by subject area. Second, it would put too much emphasis on test scores relative to other components of high school courses, including various skills important for future success in higher education such as ability to work in teams and to solve problems. Finally, it is not very feasible since most high schools do not require state-wide (or even district wide) end of course tests and even when they do, obtaining unbiased estimates of teacher effectiveness requires attention to the differential sorting of teachers among classrooms and schools.

¹⁷ For this purpose we used the percent of students eligible for free and reduced price lunch. Though an imperfect measure of income status at the high school level, this is the best measure of income available at the school level.

(defined in terms of standard deviations around a mean of zero) and the lowest percentage of Board-certified teachers.

A more detailed look at how teacher characteristics are distributed by type of student among algebra I courses is shown in Table 7. This table, which is based on the data for the 2002-03 cohort of students in our sample, depicts the probabilities that a student of each type will be in classroom with the specified type of teacher. We remind the reader that this sample includes a selected group of students, those who are still in high school and are taking algebra I. The credentials are all defined to represent weaker qualifications. Hence, in all cases, a larger number means that a student has a higher probability of having a teacher with relatively weak qualifications along the specified dimension.

The table's first column indicates that the probability of having a novice teacher for algebra I is higher for black students than for white students, for males than for females, and (slightly so) for students with non-college-educated parents compared to students with college educated parents. The difference of 4.5 percentage points between black males and white females means that black males are about 22 percent more likely than white females to have a novice teacher. Similar patterns are evident for the other seven measures shown. Particularly striking are the differences in the probabilities of having a teacher with test scores more than one standard deviation below the average. The probability for a black male is about 10 percent, while that for a white female is about 4.6 percent. Thus, black males are more than twice as likely as white females to have a teacher with low test scores.

Despite the remarkably consistent patterns of differences by race across credentials, the differences between the teacher credentials for black and white students translate into what may at first appear to be very small differences in student achievement. For example, a 4 percentage point difference in the probability of having a lateral entry teacher translates into only a 0.0024 difference in predicted achievement (=0.04 times the 0.06 estimate from Table 3). The effects summed across all the credentials in the table leads to a predicted achievement difference between black and white students that is less than 0.02 standard deviations. This difference looms larger, however, in light of the *changes* in the black-white test score gap in math as students progress from middle school to high school. For the most recent cohort of students in our sample, the black- white difference in test scores in eighth grade math was 0.7060 standard deviations and it increased slightly to 0.7069 standard deviations in algebra 1.¹⁸ Thus, the predicted achievement effects of the uneven distribution of teachers across students of different races is not only large enough to account for this increase in the achievement gap but also could have reduced it somewhat had teachers been more evenly distributed.¹⁹

Thus, the combination of the systematic differences by race, gender, and education level of the parents in the distribution of teacher credentials and the evidence presented in this paper that credentials are predictive of student achievement should be cause for serious policy concern.

¹⁸ For students who took algebra 1 in 8th grade, we use their 7th grade math scores.

¹⁹ Another way to interpret the 0.02 impact is to compare it to the estimated coefficient of 0.0593 for black students that emerges from a model specification that omits student fixed effects (see footnote 9 above). The assignment of teachers with poorer credentials increases the black-white test score gap by a significant amount relative to this black-white difference. .

Appendix

This appendix briefly describes the steps we took to match students' test scores by subject to the students' teachers, and reports the distributions of test taking for matched students.

Matching students to teachers by subject.

In the end-of-course (EOC) data file, we have a unique identifier for each student and a unique identifier for the proctor of the test. The problem is that the proctor may not have been the student's teacher in that subject. We used the following steps to distinguish the valid proctors (those for whom we are quite certain that the proctor was the student's teacher for that subject) from the invalid ones.

Starting with the EOC data, we divided students into possible classrooms by school year, semester²⁰, subject, class period and proctor for the EOC exam. We then characterized each of those possible classrooms by variables such as group size, minimum and maximum grade level of students represented in the class, and indicator of a class size of less than 5 or more than 40 so that we could compare the possible classrooms to actual classrooms as reported in school activity reports (SAR). Because a variety of course names and numbers are used for each subject in the SAR data, we had to specify which courses were relevant for which subject. We used the following course codes: For Algebra 1, courses with subject codes 2021, 2011, and 2023; for Biology, course codes between and including 3020-3035; for ELP, course codes between and including 4005 and 4095; for English 1021; and for Geometry 2030 and 2031. At this initial stage we eliminated a significant number of possible classrooms based on inconsistencies in the grade spans between the possible classrooms in the EOC data and classrooms in the SAR data.

We then matched the remaining possible classrooms from the EOC data to classrooms in SAR by year, school semester, teacher ID and subject. For each potential match at this stage, we constructed a fitness statistic based on the sum of the squared deviations of class size, males and whites, and rejected any matches for which this fit statistic exceeded 0.5.

The final step was to check that the teachers (= proctors) in the matched classrooms were reported as teaching the specified subject in the right semester at the right time.

This strategy allows us to match about 70 to 75 percent of all students to their teachers in each subject for each cohort as shown in Appendix Table A.1. The final

²⁰ Since the EOC data do not have a semester variable, we created one based on the date the test was taken.

columns of that table compare the matched to all students by subject and year on two dimensions, percent minority and average normalized test scores. The matched students are slightly less likely to be minority, but the differences in the shares are small and are less than 0.02 in most cases. With respect to test scores, the averages for all students are, by construction, 0.000. In each cell of the table, the average test scores of the matched students are slightly above average, on the order of about 0.02 to 0.05 standard deviations, with no clear patterns over time within subjects.

Distribution of students by number of tests.

Recall that we chose to work with 10th grade cohorts in years from 1999/2000 to 2002/2003 so that we would have data on as many of the ninth and tenth grade tests as possible for each student. Table A.2 provides information on the number of students in each cohort who have 3, 4, or 5 EOC test scores. The table indicates that the percentage of students with matched teachers with at least three tests varies across cohorts from 73 to about 77 percent, with no clear pattern over time. The analysis in this paper is based on this subset of the matched students.

References

- Ballou, D., & M. Podgursky (1998). "The case against teacher certification." *Public Interest*, 132: 17-29.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, James. (2006). "How changes in entry requirements alter the teacher workforce and affect student achievement." *Education Finance and Policy*, 1 (2), 176-216.
- Cavalluzzo, L. C. (2004). "Is National Board Certification an Effective Signal of Teacher Quality?" The CNA Corporation. Available online at <http://www.cna.org/documents/CavaluzzoStudy.pdf>.
- Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor. (Forthcoming). "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects." *Economics of Education Review*.
- Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2006a). "Teacher-student matching and the assessment of teacher effectiveness." *Journal of Human Resources*, XLI (4), 778-820.
- Clotfelter, C. T., Ladd, H.F. & Vigdor, J.L. (2006b) "The Academic Achievement Gap in Grades 3 to 8." National Bureau of Economic Research Working Paper.
- Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2007). "How and why teacher credentials matter for student achievement." National Bureau of Economic Research Working Paper.12828. Also available on the CALDERCenter.org web site.
- Clotfelter, C.T., Ladd, H.F., Vigdor, J.L. & Wheeler, J. (2007). "High poverty schools and the distribution of principals and teachers." *North Carolina Law Review*. Vol 85, no. 5 (June), pp.1345-1379
- Darling-Hammond, L. (2002). *Research and rhetoric on teacher certification: A response to "Teacher Certification Reconsidered," Education Policy Analysis Archives*, 10(36). Retrieved [9-5-06] from <http://epaa.asu.edu/epaa/v10n36.html>.
- Ehrenberg, R. G. and D. J. Brewer (1994). "Do school and teacher characteristics matter? Evidence from High School and Beyond." *Economics of Education Review* 13(1): 1 - 17.
- Glazerman, S., D. P. Mayer, and P.T. Decker. (2005). "Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes." *Journal of Policy Analysis and Management* 25(1): 75-96.

- Gleason, Philip M., "Participation in the National School Lunch Program and the School Breakfast Program," *Am J. Clinical Nutrition* 61 (suppl) (1995): pp. 213S-220S.
- Goldhaber, D. (forthcoming). "Everyone's Doing It, but What Does Teacher Testing Tell Us About Teacher Effectiveness?" *Journal of Human Resources*.
- Goldhaber, D. (2006). "National Board Teachers Are More Effective, But Are They In The Classrooms Where They're Needed The Most?" *Education Finance and Policy*, 1(3).
- Goldhaber, D. (forthcoming). "Everyone's Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness?" *Journal of Human Resources*.
- Goldhaber, D. (In press, 2006). "Teacher Licensure Tests and Student Achievement: Is Teacher Testing an Effective Policy?" In *Learning from Longitudinal Data in Education*. Edited by Duncan Chaplin and Jane Hannaway. Washington, DC: UI Press.
- Goldhaber, D. 2004. "Why Do We License Teachers?" In Frederick Hess, editor, *A Qualified Teacher in Every Classroom: Appraising Old Answers and New Ideas*. Edited by Frederick Hess, Andrew Rotherham, and Kate Walsh. Cambridge, MA: Harvard Education Press, pp. 81-100. (CHECK this citation)
- Goldhaber, D. and D. J. Brewer (1997a). "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity." *Journal of Human Resources* 32(3): 505-523.
- Goldhaber, D. and D. J. Brewer (1997b). Evaluating the Effect of Teacher Degree Level on Educational Performance. *Developments in School Finance 1996*. J. William Fowler. Washington, DC, National Center for Education Statistics: 197 - 210.
- Goldhaber, D. and D. J. Brewer (2000). "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." *Educational Evaluation and Policy Analysis* 22(2): 129-145.
- Goldhaber, D. and E. Anthony (2007). "Can Teacher Quality be Effectively Assessed? National Board Certification as a Signal of Effective Teaching." *Review of Economics and Statistics* 89(1): 134-150.
- Goldhaber, D. Forthcoming. "Teachers Matter, But Effective Teacher Quality Policies Are Elusive: Hints from Research for Creating a More Productive Teacher Workforce." In Helen F. Ladd and Edward B. Fiske, editors, *Handbook of Research on Education Finance and Policy*. Lawrence Erlbaum/Routledge Press.
- Gordon, R., T.J. Kane, and D.O. Staiger. (2006). "Identifying Effective Teachers Using Performance on the Job." *The Hamilton Project: Discussion Paper 2006-01*.

- Washington, DC: Brookings Institution. Available online at http://www.brook.edu/views/papers/200604hamilton_1.pdf.
- Gundersen, Craig, Rosanna Mentzer Morrison, and Linda M. Ghelfi, "Certifying Eligibility in the National School Lunch Program," Food Assistance Research Brief, Food Assistance and Nutrition Research Report Number 34-4, USDA, July 2003.
- Hanushek, E. A., Kain, J. F., O'Brien, D. M. and Rivkin, S.G. (2005). "The market for teacher quality.National Bureau of Economic Research." Working paper 11154. .
- Hanushek, E.A. (1997). "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis*, 19(2), 141-164.
- Harris, D.N and Sass, T.R. (2007) "The Effects of NBPTX-Certified Teachers on Student Achievement". CALDER working paper (caldercenter.org)
- Kane, T.J., J.E. Rockoff, and D.O. Staiger. (2006). "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." Working Paper.
- Ladd, H.F., T.R. Sass and D.N. Harris,. 2007. " The Impact of National Board Certified Teachers on Student Achievement in Florida and North Carolina: A Summary of the Evidence." Prepared for the National Academies Committee on the Evaluation of the Impact of Teacher Certification by NBPTS. Available at CALDERCENTER.ORG.
- Monk, D. and J. King (1994). "Multi-level Teacher Resource Effects on Pupil Performance in Secondary Mathematics and Science: The role of teacher subject matter preparation." *Contemporary Policy Issues: Choices and Consequences in Education*. R. G. Ehrenberg (ed). Ithaca, NY, ILR Press.
- Monk, D. H. (1994). "Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement." *Economics of Education Review* 13(2): 125-145.
- National Commission on Teaching and America's Future (NCTAF). (1996) *What Matters Most: Teaching for America's Future*. New York: Author.
- Rivkin, S.G., Hanushek, E.A. & Kain, J.F. (2005). "Teachers, schools and academic achievement." *Econometrica*, 79, 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.

Clotfelter, Ladd, and Vigdor, March, 2008

Southern Regional Education Board (SREB) 2007. *The Changing Role of Statewide High School Exams*. A Focus Report in the Challenge to Lead Series.
http://www.sreb.org/main/Goals/Publications/07E03_Statewide_Exams.pdf
(Accessed 09/23/07)

Walsh, K. (2001). *Teacher certification reconsidered: Stumbling for quality*. Baltimore, Abell Foundation.

| Table 1. Probabilities of enrolling in advanced high school courses by absolute and relative measures of student ability in math and reading.* | | | | |
|--|-----------------------------------|---------------|-------------|-------------------------------|
| | Low | Medium | High | Ratio: High to low |
| | Math ability (absolute) | | | |
| Advanced algebra | 0.131 | 0.182 | 0.305 | 2.33 |
| Advanced English | 0.294 | 0.446 | 0.688 | 2.34 |
| | Reading ability (absolute) | | | |
| Advanced algebra | 0.118 | 0.185 | 0.322 | 2.75 |
| Advanced English | 0.226 | 0.457 | 0.763 | 3.38 |
| | Math ability (relative) | | | |
| Advanced algebra | 0.205 | 0.206 | 0.204 | 1.00 |
| Advanced English | 0.517 | 0.472 | 0.432 | 0.84 |
| | Reading ability (relative) | | | |
| Advanced algebra | 0.204 | 0.206 | 0.205 | 1.01 |
| Advanced English | 0.432 | 0.472 | 0.537 | 1.22 |
| <p>Based on data for all 10th graders in 2002/03 for whom we have data on all the relevant variables. Ability in math and reading are measured by normalized scores on end-of-course tests in eighth grade. Relative measures are defined as the student's normalized test score in one subject minus her average test scores in the two subjects. The columns low, medium, and high refer to tertiles of the distributions of the various measures of ability. All entries except for those in the final column are probabilities. (Calculated from memo 6.08.07)</p> | | | | |

| Table 2. Regression- based test of assumptions, by cohorts and full sample | | | | | |
|--|--------------------|--------------------|-------------------|--------------------|-------------------|
| | Cohort 1 - 2000 | Cohort 2 - 2001 | Cohort 3- 2002 | Cohort 4 - 2003 | All 4 cohorts |
| Student ability difference | 0.0121 (0.011) | 0.0017 (0.010) | 0.0061 (0.009) | -0.0059 (0.010) | 0.0023 (0.006) |
| Constant ? | Yes | Yes | Yes | Yes | Yes |
| School fixed effects? | Yes | Yes | Yes | Yes | Yes |
| No. of obser- vations | 30,010 | 35,369 | 36,598 | 35,620 | 137,597 |
| R- squared | 0.265 | 0.254 | 0.262 | 0.271 | 0.191 |
| <p>Notes. Dependent variable is the difference between the average licensure test scores of the student's algebra and English teachers. Student ability difference is the difference between the student's eighth grade test score in math and in reading. Standard errors are in parentheses. (From memo 6.13.07 and 6.21 update)</p> | | | | | |

| Table 3. Achievement effects of teacher credentials (from full models)^a | | |
|---|---|---|
| | Model 1 (with student fixed effects) (preferred model) | Model 2 (with school fixed effects; no student fixed effects) |
| Teacher credentials | | |
| Years of experience (base =0) | | |
| 1-2 years | 0.0503** (0.004) | 0.0535** (0.005) |
| 3-5 years | 0.0611** (0.004) | 0.0682** (0.005) |
| 6-12 years | 0.0611** (0.004) | 0.0662** (0.005) |
| 13-20 years | 0.0594** (0.004) | 0.0674** (0.005) |
| 21-27 years | 0.0617** (0.004) | 0.0673** (0.005) |
| more than 27 years | 0.0429** (0.005) | 0.0566** (0.006) |
| Teacher test score (normalized) | 0.0105** (0.001) | 0.0125** (0.002) |
| Type of license (base = regular) | | |
| lateral entry | -0.0609** (0.005) | -0.0554** (0.006) |
| lateral entry- prior | 0.0171 (0.033) | 0.0488 (0.040) |
| other license | -0.0466** (0.004) | -0.0429** (0.004) |
| Certification (base= no cert.) | | |
| certified in subject | 0.0808** (0.012) | 0.0537** (0.014) |
| certified in related subject | 0.0744** (0.012) | 0.0545** (0.014) |
| certified to teach | 0.0116 (0.014) | 0.0407* (0.016) |
| National Board Certification status (base = never certified) | | |
| pre-certification | 0.0215** (0.005) | 0.0233** (0.006) |
| certification app. year | 0.0483** (0.007) | 0.0509** (0.008) |
| has certification | 0.0509** (0.004) | 0.0528** (0.005) |
| Graduate degree | 0.0003 (0.002) | 0.0015 (0.003) |
| Undergraduate institution (base = not competitive) | | |
| very competitive | 0.0169** (0.003) | 0.0209** (0.004) |
| competitive | 0.0047+ (0.003) | 0.0069* (0.003) |
| unranked | -0.0057 (0.006) | 0.0001 (0.007) |
| Regression information | | |
| No. of observations | 857,548 | 856,929 |
| R-squared | 0.783 | 0.636 |

a. The dependent variable is normalized student achievement by subject for four cohorts of students. Model 1 also includes all the variables in Table 4, as well as cohort and subject-by-grade fixed effects. Model 2 includes all the variables in both Tables 4 and 5 as well as cohort fixed effects. It differs from Model 1 in that it includes school, rather than student, fixed effects, and also by the inclusion of the student characteristics in Table 5. ** signifies statistical significance at the 0.01 level; * at the 0.05 level; and + at the 0.10 level. All errors are clustered at the classroom level. (Memo 7/04/07, update of 5/16/07)

| Table 3A. Achievement effects of teacher experience with variations | | | | |
|--|-------------------------|------------------------|--|--|
| Years of experience (base =0) | Model 1 (student FE) | Model 2 (school FE) | Model 1 with teacher fixed effects | Model 2 with teacher fixed effects |
| 1-2 | 0.0503** | 0.0535** | <i>0.0539*</i> | <i>0.0588**</i> |
| 3-5 | 0.0611** | 0.0682** | <i>0.0862**</i> | <i>0.1210**</i> |
| 6-12 | 0.0611** | 0.0662** | <i>0.1409**</i> | <i>0.1685**</i> |
| 13-20 | 0.0594** | 0.0674** | <i>0.1972**</i> | <i>0.2099**</i> |
| 21-27 | 0.0617** | 0.0673** | <i>0.2957**</i> | <i>0.2280**</i> |
| >27 | 0.0429** | 0.0566** | <i>0.3524 **</i> | <i>0.2572**</i> |

Notes. The first two columns replicate columns 1 and 2 from Table 3. The third column is similar to model 1 except that it includes teacher fixed effects as well as student fixed effects and excludes any teacher variables that do not vary by subject. The third column is similar to model 2 except that it includes teacher fixed effects as well as student fixed effects and excludes any teacher variables that do not vary by subject. ** signifies statistical significance at the 0.01 level; * at the 0.05 level; and + at the 0.10 level. All errors are clustered at the classroom level. (memo 7/05/07 update of 6/25/07)

| Table 3B. Achievement effects of teacher test scores with variations | | | | |
|--|----------|------------|------------------------------|--|
| | Model 1 | Non-linear | Subject specific test scores | Subject specific test scores and subject certification |
| Teacher test score (normalized) | 0.0105** | | | |
| Teacher test scores (Base = > -1 sd and < 1 sd) | | | | |
| > 1 s.d. | | 0.0098** | | |
| < - 1 s.d. | | -0.0266** | | |
| Teacher test scores by subject | | | | |
| Math score | | | 0.0309** | 0.0310** |
| Other scores | | | 0.0025 | 0.0014 |
| No math test | | | -0.0148** | -0.0123* |
| Biology score | | | 0.0125* | 0.0138* |
| Other scores | | | 0.0088* | 0.0091** |
| No biology test | | | -0.0021 | -0.0080 |
| ELP score | | | -0.0133** | -0.0121* |
| Other scores | | | 0.0094 | 0.0077 |
| No ELP test | | | -0.0231** | -0.0263** |
| English test score | | | -0.0151** | -0.0139* |
| Other scores | | | -0.0043 | -0.0028 |
| No Engl. test | | | -0.0091** | -0.0018 |
| Notes. The entry in the first column comes from model 1 in Table 3. The entries in the second column come from model 1 with two indicator variables substituted for the linear form of the teacher test score variable. The entries in the third column come from model 1, with subject specific test scores substituted for the average test score variables. The entries in column 4 with subject specific test scores and subject specific certification variables. Each of the subject specific test scores apply only to the relevant subject. ** signifies statistical significance at the 0.01 level; * at the 0.05 level; and + at the 0.10 level. All errors are clustered at the classroom level. (memo 7/13/07 update to 5/25/07) | | | | |

| Table 3C. Achievement effects of teacher certification with variations | | | |
|---|----------|--------------------------------|---|
| | Model 1 | subject specific certification | subject specific certification, with subject specific test scores |
| Certified in the subject | 0.0808** | | |
| Certified in a related subject | 0.0744** | | |
| Certified | 0.0116 | | |
| Algebra and Geometry -- certified in math | | 0.1266** | 0.1199** |
| Algebra and Geometry – certified (but not in math) | | 0.0488* | 0.0525* |
| biology - certified in biology | | 0.0293 | 0.0271 |
| biology – certified in related subject | | 0.0172 | 0.0145 |
| biology – certified in some other subject | | 0.0632* | 0.0661** |
| ELP – certified in ELP | | 0.0237 | 0.0036 |
| ELP-certified in related subject | | 0.1022** | 0.0817** |
| ELP-certified in some other subject | | 0.1057** | 0.1000* |
| English – certified in English | | -0.0250 | -0.0154 |
| English – certified in related subject | | 0.0150 | 0.0241 |
| English –certified in some other subject | | -0.1538** | -0.1448** |

Notes. The entries in the first column comes from model 1 in Table 3. The entries in the second column come from model 1 with subject-specific certification variables substituted for the general certification variables. The entries in the third column come from model 1 with subject specific certification variables substituted for the general certification variables and subject specific test scores substituted for the average test score variables. Each of the subject-specific variables apply only to student test scores in the specified subject.. ** signifies statistical significance at the 0.01 level; * at the 0.05 level; and + at the 0.10 level. All errors are clustered at the classroom level. (memo 7/13/07 update to 5/25/07)

| Table 3D. Achievement effects of graduate degrees with variations | | | |
|--|---------|--------------------------|---------------------------|
| | Model 1 | Model 1 with variation 1 | Model 1 with variation 2. |
| Graduate degree (base = no graduate degree) | | | |
| Any graduate degree | 0.0003 | | |
| Master's | | 0.0046* | |
| “advanced” | | 0.0012 | - 0.0039 |
| Ph.D. | | -0.1001** | -0.1001** |
| Masters before teaching | | | -0.0055 |
| Masters after 1 year before 5 years | | | 0.0091** |
| Master's after 5 years | | | 0.0090** |
| Note. Entry in column 1 is from Model 1 in Table 3. Variation 1 substitutes three graduate degrees for the single graduate degree. Variation 2 substitutes time varying master's degrees for the single master's variable. | | | |

| Table 3E. Comparisons of Achievement Effects by Credential sets | | | |
|--|---|----------------------|---|
| Credentials | Very weak set | Very strong set | Estimated differential achievement effect (weak – strong) |
| Experience (years) | 0 | 6-12 | - 0.0611 |
| Teacher test score (SD) | <-1 | >1 | - 0.0364 |
| Type of license | Lateral entry | Regular | - 0.0609 |
| Certification by subject | Certified but not in subject or related subject | Certified in subject | - 0.0692 |
| National Board Certification | Not Board Certified | Board Certified | - 0.0509 |
| Graduate degree | No graduate degree | Master's degree | - 0.0030 |
| Undergraduate institution | Uncompetitive | Very competitive. | - 0.0169 |
| Total difference | | | - 0.2984 |
| Based on coefficients from prior tables. | | | |

| Table 4. Achievement effects of teacher and class characteristics from full models^a | | |
|--|---|---|
| | Model 1 (with student fixed effects) (preferred model) | Model 2 (with school fixed effects; no student fixed effects) |
| Teacher characteristics | | |
| Gender (base = female) | | |
| Male | -0.0566** (0.002) | -0.0562** (0.003) |
| Race (base = white) | | |
| Black | -0.0592** (0.003) | -0.0559** (0.004) |
| Hispanic | -0.0150 (0.020) | -0.0078 (0.024) |
| Other | -0.0420** (0.010) | -0.0276* (0.012) |
| Classroom characteristics | | |
| Percent nonwhite | -0.0228* (0.008) | -0.0434** (0.010) |
| Percent male | -0.0121 (0.008) | -0.0382** (0.009) |
| Peer average achievement | 0.1123** (0.003) | 0.1988** (0.003) |
| Advanced class | 0.0281** (0.003) | 0.0498** (0.003) |
| Class size | -0.0026** (0.000) | -0.0030** (0.000) |
| <p>a. The dependent variable is (normalized) student achievement by subject for four cohorts of students. Model 1 also includes all the variables in Table 3 as well as cohort and subject-by-grade fixed effects. Model 2 includes all the variables in Tables 3 and 5 as well as cohort fixed effects. It differs from Model 1 in that it includes school, rather than student, fixed effects, and by the inclusion of the student characteristics in Table 5 ** signifies statistical significance at the 0.01 level; * at the 0.05 level; and + at the 0.10 level. All errors are clustered at the classroom level. See regression summary information in Table 3.</p> | | |

| Table 4A. Teacher Characteristics with interactions | | | |
|---|-----------|-------------|-------------|
| | Model 1 | Variation 1 | Variation 2 |
| Gender (base = female) | | | |
| Male | -0.0566** | | -0.0563** |
| Race (base=white) | | | |
| Black | -0.0592** | -0.0591** | |
| Hispanic | -0.0150 | -0.0144 | |
| Other | -0.0420** | -0.0418** | |
| Teacher and student gender (base= female T and female S) | | | |
| Female T and male S | | 0.0139 | |
| Male T and female S | | - 0.1069** | |
| Male T and male S | | 0.0115 | |
| Teacher and student race (base = white teacher and white student) | | | |
| White T and black S | | | 0.0067 |
| White T non white or non black S | | | 0.0327* |
| Black T and black S | | | -0.0199 |
| Black T and white S | | | -0.0848** |
| Black T and non white or non black S | | | -0.0210 |
| Hispanic T non white or non black S | | | -0.1053* |
| Hispanic T white S | | | 0.0334 |
| Hispanic T Black S | | | -0.0499+ |
| Notes. Entries in first column are the same as those in Table 4. In the variations, the indicated variables replace the race and gender variables as indicated. | | | |

| Table 5. Achievement effects of student characteristics from full model with school fixed effects ^a | |
|--|--|
| | Model 2 (School fixed effects; no student fixed effects) |
| Student characteristics | |
| 8 th grade math score (normalized) | 0.4057** (0.001) |
| 8 th grade reading score (normalized) | 0.3444** (0.002) |
| Gender (base = female) | |
| male | 0.0522** (0.002) |
| Race (base = white) | |
| black | - 0.0593** (0.002) |
| Hispanic | 0.0222** (0.005) |
| other race | 0.0292** (0.004) |
| Handicapped | - 0.0001 (0.004) |
| Limited English | 0.0279* (0.012) |
| Parental education (base = high school graduate) | |
| high school drop out | -0.0102** (0.003) |
| some college | 0.0346** (0.002) |
| college graduate | 0.0571** (0.002) |
| Age (in months) | -0.0029** (0.000) |
| Repeat test | 0.0713** (0.005) |
| Student cohort (base = 10 th grade in 2000) | |
| 10 th grade in 2001 | 0.0135** (0.003) |
| 10 th grade in 2002 | 0.0198** (0.003) |
| 10 th grade in 2003 | 0.0312** (0.003) |
| a. The dependent variable is normalized student achievement by subject for four cohorts of students. Also included in the full model are the variables in Tables 3 and 4 as well as subject-by-grade fixed effects. ** signifies statistical significance at the 0.01 level; * at the 0.05 level; and + at the 0.10 level. All errors are clustered at the classroom level. See regression summary information in Table 3. | |

| Table 6. Credentials of High School Teachers by Poverty Quartile, 2004 (Averages weighted by number of teachers in each school; percent except where noted) | | | | |
|--|---|------------|------------|--|
| | Quartile 1 (high poverty schools) | Quartile 2 | Quartile 3 | Quartile 4 (low-poverty schools) |
| Less than three years experience | 17.3 | 15.2 | 13.4 | 14.6 |
| Less competitive undergraduate institution | 27.4 | 19.6 | 15.4 | 14.2 |
| Nonregular license | 20.5 | 17.7 | 14.1 | 13.3 |
| Licensure test scores (average, in standard deviations) | -0.057 | 0.032 | 0.105 | 0.117 |
| Board Certified | 4.1 | 7.9 | 9.4 | 9.9 |
| Source. Calculated by the authors. See Table 3 in Clotfelter, Ladd, Vigdor & Wheeler (2007) | | | | |

| Table 7. Probabilities that a student of a particular type will have a teacher of a specific type., Algebra I, students in 2002/03 tenth grade cohort | | | | | | | | |
|--|--------|---------------|---------------|---------------------------------|---------------------------|--------------------|-----------------------|-----------------------|
| Student | Novice | Other License | Lateral Entry | Uncompetitive Undergrad College | Teacher Test Score <-1 SD | No Advanced Degree | Not Certified in Math | Never Board Certified |
| Blacks | 0.247 | 0.138 | 0.094 | 0.275 | 0.100 | 0.722 | 0.061 | 0.940 |
| Hispanics | 0.205 | 0.107 | 0.083 | 0.244 | 0.067 | 0.684 | 0.048 | 0.915 |
| Whites | 0.212 | 0.107 | 0.054 | 0.211 | 0.048 | 0.674 | 0.039 | 0.903 |
| Black Males | 0.253 | 0.141 | 0.102 | 0.275 | 0.104 | 0.730 | 0.071 | 0.939 |
| Black Females | 0.242 | 0.135 | 0.089 | 0.275 | 0.096 | 0.716 | 0.053 | 0.941 |
| White Males | 0.216 | 0.112 | 0.055 | 0.206 | 0.046 | 0.674 | 0.043 | 0.907 |
| White Females | 0.208 | 0.103 | 0.053 | 0.215 | 0.050 | 0.673 | 0.035 | 0.900 |
| Non College Educated Parent | 0.222 | 0.121 | 0.068 | 0.243 | 0.064 | 0.691 | 0.048 | 0.918 |
| College Educated Parent | 0.220 | 0.104 | 0.060 | 0.216 | 0.062 | 0.676 | 0.040 | 0.908 |

| Table A-1. Unmatched and matched students by student cohort and by subject | | | | | | |
|---|--------------|--|------------------|------------------|-----------------------|------------------|
| 10 th grade cohort * | All students | Students matched with their teachers (percent of all students) | Percent minority | | Normalized test score | |
| | | | All students | Matched students | All students | Matched students |
| Algebra I | | | | | | |
| 1999/2000 | 91,102 | 64,648 (71.0) | 0.336 | 0.334 | 0.000 | 0.007 |
| 2000/2001 | 94,085 | 67,337 (71.6) | 0.353 | 0.348 | 0.000 | 0.006 |
| 2001/2002 | 100,048 | 70,424 (70.4) | 0.367 | 0.359 | 0.000 | 0.013 |
| 2002/2003 | 107,362 | 73,587 (68.5) | 0.387 | 0.380 | 0.000 | 0.003 |
| English I | | | | | | |
| 1999/2000 | 95,772 | 72,790 (76.0) | 0.348 | 0.340 | 0.000 | 0.056 |
| 2000/2001 | 96,907 | 71,698 (74.0) | 0.351 | 0.349 | 0.000 | 0.050 |
| 2001/2002 | 99,480 | 73,332 (73.7) | 0.351 | 0.346 | 0.000 | 0.040 |
| 2002/2003 | 101,157 | 74,496 (73.6) | 0.358 | 0.343 | 0.000 | 0.038 |
| Biology | | | | | | |
| 1999/2000 | 82,072 | 62,288 (75.9) | 0.329 | 0.331 | 0.000 | 0.030 |
| 2000/2001 | 83,301 | 61,844 (74.2) | 0.352 | 0.340 | 0.000 | 0.023 |
| 2001/2002 | 85,570 | 66,372 (77.6) | 0.354 | 0.351 | 0.000 | 0.023 |
| 2002/2003 | 88,106 | 64,450 (73.2) | 0.369 | 0.352 | 0.000 | 0.035 |
| Econ/Legal/Political | | | | | | |
| 1999/2000 | 81,038 | 61,388 (75.8) | 0.349 | 0.343 | 0.000 | 0.019 |
| 2000/2001 | 92,228 | 68,435 (74.2) | 0.350 | 0.337 | 0.000 | 0.027 |
| 2001/2002 | 97,624 | 74,856 (76.7) | 0.371 | 0.361 | 0.000 | 0.025 |
| 2002/2003 | 91,710 | 68,769 (75.0) | 0.384 | 0.375 | 0.000 | 0.020 |
| Geometry | | | | | | |
| 1999/2000 | 64,821 | 48,914 (75.5) | 0.313 | 0.313 | 0.000 | 0.025 |
| 2000/2001 | 65,716 | 50,564 (76.9) | 0.310 | 0.289 | 0.000 | 0.052 |
| 2001/2002 | 69,065 | 50,615 (73.3) | 0.328 | 0.304 | 0.000 | 0.065 |
| 2002/2003 | 71,962 | 52,995 (73.6) | 0.348 | 0.334 | 0.000 | 0.038 |
| .Source. Calculated by the authors from End-of-Course files and School Activity Reports * refers to all students who were in 10 th grade in the specified year who took the test, regardless of the grade or year in which they took it. (ah 1/12/07) | | | | | | |

| Table A.2. Distribution of test taking by matched students, by cohort | | | | | |
|--|-----------------|---------|---------|---------|-----------------|
| 10 th grade cohort* | no. of students | 3 tests | 4 tests | 5 tests | 3 or more tests |
| 1999/2000 | 80,240 | 25.9 | 29.6 | 17.1 | 72.6 |
| 2000/2001 | 83,581 | 23.2 | 31.0 | 23.1 | 77.3 |
| 2001/2002 | 86,338 | 23.9 | 31.3 | 20.9 | 76.1 |
| 2002/2003 | 88,444 | 24.0 | 29.6 | 19.6 | 73.2 |

Source. Calculated by the authors.

* Refers to all students who were in 10th grade in the specified year who took any of the tests, regardless of the grade or year in which they took it. (ah 1/23/07)