
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

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. We find compelling evidence that teacher credentials, particularly licensure and certification, affects student achievement in systematic ways and that the magnitudes are large enough to be policy relevant. Our findings imply that the uneven distribution of teacher credentials by race and socioeconomic status of high school students—a pattern we also document—contributes to achievement gaps in high school. In addition, some troubling findings emerge related to the gender and race of the teachers.

I. Introduction

Nearly all observers of the education process, including scholars, school administrators, policymakers, and parents, point to teacher quality as the most significant institutional determinant of student achievement. Much less is known about how a teacher's quality is related to her credentials, or about the credential-related policy levers that might be used to raise the overall quality of teachers and to ensure an equitable distribution of high-quality teachers across schools and class-

The authors are all full professors at the Sanford School of Public Policy, Duke University. They thank Aaron Hedlund, L. Patten Priestley and Sarah Gordon for outstanding research assistance and the Spencer Foundation and the Center for the Analysis of Longitudinal Data in Education Research for financial support. An earlier, somewhat longer version of this paper is available as NBER working paper 13617. Researchers may acquire the data used in this article from the North Carolina Education Research Data Center <http://childandfamilypolicy.duke.edu/ep/nceddcenter/index.html>. For questions regarding the data, contact Helen Ladd, hladd@duke.edu.

[Submitted January 2008; accepted February 2009]

ISSN 022-166X E-ISSN 1548-8004 © 2010 by the Board of Regents of the University of Wisconsin System

rooms. Although the increasing availability of longitudinal administrative test data for students in Grades 3–8 has generated a number of studies investigating how teacher credentials affect student achievement in elementary schools (Clotfelter, Ladd, and Vigdor 2006, 2007a, 2007b; Goldhaber and Anthony 2007; Rockoff 2004), much less research has been done at the high school level.

Yet policymakers are increasingly turning their attention to high schools in the recognition that even minimal participation in the economic and political life of a knowledge-based world requires a high school diploma. New descriptive research documents that poor student performance in core high school courses is highly predictive of the failure to graduate (Allensworth and Easton 2007). Hence, more research is needed on the relationship between teacher credentials and student achievement in high schools, especially in the core courses taken early in a student's high school career.

Most of the existing knowledge about this relationship comes from studies based on national longitudinal surveys that are somewhat dated.¹ Although 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 nonrandom sorting of teachers and students that can bias the results (Goldhaber 2004; Goldhaber and Brewer 1997a). An alternative is to turn to an administrative data set, as done, for example, by Aronson, Barrow, and Sander (2007) in their study of teacher effects for ninth graders in Chicago public schools.

In this study, we use a rich data set on teachers and students from North Carolina to investigate in more depth than has heretofore been possible the effects of teacher credentials and qualifications on student achievement at the high school level. In contrast to most other states, North Carolina has long had a standard course of study for high school students that culminates in end-of-course (EOC) tests in each of a number of subjects, such as English, algebra, and biology. For the current study, 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. We match those test scores 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, and that the effects for subject-specific certification in math and English are large. Moreover the combined effects of all the credentials we measure are large enough to be relevant for policy, and the estimated effects for some credentials differ in some interesting ways from prior findings at the elementary level. 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-edu-

1. These surveys include 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, 2000).

cated 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 describe the data, explain and justify the empirical framework, and present the results. The paper concludes with a discussion of policy implications.

II. 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. In addition, many states, including North Carolina, encourage their teachers to apply for National Board Certification or provide for alternative forms of entry to the profession. Further, the types of salary schedules typically used in public schools explicitly reward two credentials—experience and graduate education. A better understanding of the relationship between teacher credentials and student achievement would help policymakers assess the value of policies designed to recruit teachers and to induce teachers to obtain those credentials. Some observers believe, however, that teacher credentials are such poor predictors of student achievement that much of the current apparatus for preparing and credentialing teachers should be jettisoned in favor of a new system in which teachers are hired (and fired) on the basis of their cognitive ability and their effectiveness in the classroom rather than their credentials (Walsh 2001).²

The research in this paper builds on our prior work on teacher credentials at the elementary school level in North Carolina. In that research, we document not only that teacher credentials matter for student achievement in Grades 4 and 5 but also that they are distributed in highly inequitable ways across schools (Clotfelter, Ladd, and Vigdor 2006, 2007a, and 2007b; Clotfelter, Ladd, Vigdor, and Wheeler 2007).

North Carolina serves as an excellent site for the study of teacher credentials at the high school level. Although many states now administer tests in high schools, most of those tests are in the form of comprehensive high school exit exams or minimum competency exams. Whatever the merits of such tests for assuring that students meet some specified level of achievement before they graduate, these tests are not very useful for examining the effectiveness of teachers. The material covered

2. The policy debate is lively and intense. See, for example, National Commission (1996); Walsh (2001); Darling-Hammond (2002); Kane, Rockoff and Staiger (2006); Staiger, Gordon and Kane (2006).

is usually broader than that covered in a specific course and the level of difficulty is typically more appropriate to middle school than to 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.³

III. 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 ninth and tenth grades: algebra; economic, legal and political systems (ELP)⁴; English I; geometry; and biology. The first three are typically taken in the ninth grade and the last two in the tenth grade, although many students take the relevant courses in other grades.⁵ The EOC test scores carry high stakes for students in that they count for 25 percent of the student's grade in the course.⁶

We analyze four cohorts of tenth graders—those who were in tenth grade in 1999/2000, 2000/01, 2001/02, and 2002/03.⁷ The final sample includes only those students for whom we could match at least three teachers to the EOC tests. That matching process involves a number of steps because the proctor identified on each student test record need not be the teacher of the course (see Clotfelter, Ladd, and Vigdor 2007c for the details). Appendix Table A1 provides information on the matched and unmatched students by subject and cohort. The percentages of all students with matched teachers taking at least one EOC test who meet the three-test criterion by cohort are 72.6, 77.3, 76.1, and 73.2.⁸ 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 implies that the estimated coefficients can be interpreted as fractions of a standard deviation.

3. For an overview of the use of comprehensive tests and end-of-course tests at the high school level in the South, see Southern Regional Education Board (2007).

4. 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.

5. 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 possibly be in the Career Prep track.

6. 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.

7. By selecting these cohorts, we allow each student in any of the cohorts the opportunity to take any one of the five tests. Because our data end in 2004/05, any student in tenth 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.

8. The comparable percentages for cohorts outside of our sample are 62.1 percent in 1999 and 68.7 percent in 2004.

IV. 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 students who possess unmeasured academic ability, 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 policymakers 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 could be subject to a negative bias. The statistical problems associated with this non-random sorting of teachers and students are 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.

In studies using state administrative data at the elementary level, researchers have addressed this problem by using longitudinal data that includes outcome measures, such as test scores in math, for each student across multiple years (Clotfelter, Ladd, and Vigdor 2007a, 2007b; Kane, Rockoff, and Staiger 2006). The availability of multiple observations 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. In the present study, we address the sorting problem once again with the use of student fixed effects, which means that we identify the relevant coefficients based on the within-student variation across subjects. As highlighted by Rothstein (2008) in the longitudinal context, however, the student fixed-effects method does not resolve bias when classroom assignments are associated with time-varying unobserved determinants of student achievement. The analogous concern in this cross-subject analysis is that classroom assignment may be correlated with unobserved determinants of student achievement that differ across subjects. We address that issue below.

At 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 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 in school k , our preferred model takes the following form:

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

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 also can include variables such as the size of the class and the characteristics of the students in the class. The other variables are as follows:

- λ_i is a set of student specific fixed effects;
- e_{ijsk} is a student-by-subject 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 individual 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, and Vigdor 2007b). 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.⁹

Equation 1 is equivalent to the following equation:

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

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 his achievement based on all his tests. Similarly, a teacher's credentials are measured relative to the average credentials of all of the teachers of that student. The term $(e_{ijsk} - e_i^*)$ is a student-specific error term that varies across subjects.

This model will generate unbiased estimates of β provided that the error term is uncorrelated with the relative (or demeaned) teacher credentials. That condition would be violated if, for example, students who are unobservably more able in some subjects (those for which $e_{ijsk} - e_i^*$ is positive) were systematically assigned to teachers with stronger credentials in that subject (those for whom $T_{ijsk} - T_i^*$ is posi-

9. An alternative would be to control for a prior test score in a related subject—for example, controlling for an eighth grade math score in a specification where Algebra or Geometry test scores serve as the dependent variable. We prefer our strategy in part because several of the tested high school subjects do not map neatly into reading or math, the two available eighth grade standardized test scores.

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				
Advanced algebra	0.131	0.182	0.305	2.33
Advanced English	0.294	0.446	0.688	2.34
Reading ability				
Advanced algebra	0.118	0.185	0.322	2.75
Advanced English	0.226	0.457	0.763	3.38

Notes: Based on data for all tenth 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.

tive).¹⁰ The common practice of tracking by ability level in high school could potentially create such a pattern, but the evidence indicates that outcome is not likely.

Assignment patterns for the 2002/2003 cohort of students show, for example, that only about 4 percent of the students who took an advanced Algebra 1 class took a regular (that is, nonadvanced) English class. Though not definitive, this assignment pattern is consistent with assignment to high school classes based on a univariate, rather than a subject-specific, concept of ability.¹¹ Table 1, which includes information on students' eighth grade tests scores in math and reading as proxy measures of subject-specific ability, explores this issue in more detail for the same cohort of students.¹² The question is the extent to which students end up in different types of classes by subject based on their subject-specific abilities. The more that they do so, the more reason there is to question the assumption of no correlation between the subject-specific error term in (Equation 1a) and the demeaned teacher credential term in that equation.

In Table 1, the students are divided into tertiles for each of the two ability measures and the entries are the probabilities that students of different ability levels in

10. It also would be biased if the teachers who are more effective based on unobservable characteristics, such as their motivation, are systemically assigned to students with stronger unobservable characteristics. Although we cannot test for this bias because it relates to the unobservable characteristics of the teachers, we have no reason to believe it would be large, especially if there is little or no bias based on the observable characteristics of the teachers.

11. The data indicate that smaller percentages of students are in advanced math classes than in advanced English classes. Hence a much higher percentage of students (31 percent) are in an advanced English class but a regular math class.

12. We are making the plausible assumption here that performance on the eighth grade math test is more closely related to performance in Algebra 1 than to that in English 1 and vice versa for performance in eighth grade reading. Subject-specific models similar to the type presented in Table 4, but with school fixed effects and various student-specific variables including their eighth grade test scores, (not reported here) provide support for this assumption.

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
Number of observations	30,010	35,369	36,598	35,620	137,597
R-squared	0.265	0.254	0.262	0.271	0.191

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.

math and reading are enrolled in advanced algebra and advanced English courses. The top panel shows that math ability is an equally strong predictor of assignment to advanced algebra sections as to advanced English sections. In both cases, students in the top math ability tertile were about 2.3 times more likely to enroll in an advanced section, relative to a student in the bottom ability tertile. With respect to reading ability, the positive correlations are again strong, but this time a bit stronger for being in an advanced English course. Note, however, that the association between reading ability and algebra placement is actually stronger than the association between math ability and algebra placement. These patterns are consistent with the view that when assigning students to classrooms, schools treat student ability as unidimensional.¹³

Table 2 provides a more direct test of the relationship between relative teacher credentials and relative student ability. For the purposes of this test, we focus on a single characteristic of teachers, namely the average test score on their licensure exams, a feature that previous research (as well as results 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. 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 subject-

13. Further support for this univariate concept of ability emerges from a recent study of Teach for America teachers in North Carolina that builds on the approach used in the present study (Xu, Hannaway, and Taylor 2008). The authors show, based on a principal components analysis of student test scores on eight end-of-course tests, that all eight test scores load predominantly on a single underlying dimension.

specific component of the error term in Equation 1a) and the relative qualifications of her high school algebra and English teachers (a proxy for the teacher term 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.¹⁴ Hence, the data provide little or no reason to question the basic assumption that the subject-related individual error term in a model with student fixed effects is uncorrelated with the explanatory variable of primary interest, (de-meaned) teacher credentials. At the same time, we are not able to prove conclusively the validity of this assumption.

V. Achievement Effects of Teacher Credentials

We include in the analysis several sets of teacher credentials such as years of experience and educational background, licensure test scores, type of licensure, and various forms of certification including National Board Certification. Of most interest here are the licensure and certification credentials because of their direct relevance to current policy debates. Certain teacher characteristics, such as race and gender, are also of interest, particularly as they interact with characteristics of the students.

Table 3 provides summary data for all the explanatory variables included in the basic regression. Given the large sample sizes, even small percentages of teachers within a specific category correspond to nontrivial numbers of teachers. Additional discussion of these variables appears below. The teacher credentials and characteristics are supplemented with classroom characteristics, including class size, whether the class is an advanced class, and the peer groups in each class. Each of the peer variables (percents nonwhite, male and average achievement) are based on all students in the specific classroom other than the subject student. The inclusion of these classroom characteristics assures that the estimated effects of teacher credentials are not confounded by any correlations with them.

The basic results for teacher credentials are reported in the left column of Table 4, with those for other variables in the right column. The model also includes student fixed effects for the reasons discussed above, and subject-by-grade fixed effects to control for the fact that not all students take a particular course in the typical grade

14. 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 two standard errors) but even that correlation is extremely small.

Table 3
Variable means and (standard deviations)

Dependent variable ^a	0.095 (0.944)	<i>Teacher characteristics</i>	
<i>Teacher credentials</i>		Gender (base = female)	
Years of experience (base = 0)		Male	0.328 (0.469)
1–2 years	0.117 (0.321)	Race (base = white)	
3–5 years	0.154 (0.361)	Black	0.112 (0.315)
6–12 years	0.209 (0.406)	Hispanic	0.002 (0.039)
13–20 years	0.168 (0.374)	Other race	0.010 (0.101)
21–27 years	0.179 (0.383)	<i>Classroom characteristics</i>	
> 27 years	0.098 (0.297)	Percent nonwhite	0.321 (0.257)
Graduate degree	0.331 (0.471)	Peer average	
Undergraduate institution		achievement	0.226 (0.599)
(base = not competitive)		Percent male	0.487 (0.124)
Very competitive	0.269 (0.444)	Advanced class	0.369 (0.483)
Competitive	0.514 (0.500)	Class size	23.1 (5.710)
Unranked	0.023 (0.150)		
Teacher test score			
(normalized)	0.216 (0.789)		
Type of license (base = regular)			
Lateral entry	0.049 (0.217)		
Lateral entry-prior	0.020 (0.140)		
Other license	0.015 (0.123)		
Certification (base = no certification)			
Certified in subject	0.969 (0.173)		
Certified in related subject	0.012 (0.111)		
National Board Certification status			
(base = never certified)			
Precertification	0.037 (0.190)		
Application year	0.021 (0.143)		
Has certification	0.052 (0.221)		

a. The dependent variable is normalized student achievement across five subjects for four cohorts of students. The number of observations is 857,548.

for that course, and the errors are clustered at the classroom level.¹⁵ Two asterisks signify that a coefficient is statistically significant from zero at the 0.01 level and one asterisk at the 0.05 level. We begin with the results for the most commonly measured credentials—those related to teachers' experience, general educational training, and licensure test scores—and compare our findings to those at the ele-

15. Clustering errors at the classroom level accounts for any correlation of errors associated with the common experience of students in a specific class. The inclusion of student fixed effects makes less compelling the argument for clustering errors at the student level.

Table 4
Achievement effects of teacher credentials—basic model^a

<i>Teacher credentials</i>		<i>Teacher characteristics</i>	
Years of experience (base = 0)		Gender (base = female)	
1–2 years	0.0487** (0.005)	Male	–0.0546** (0.003)
3–5 years	0.0608** (0.005)	Race (base = white)	
6–12 years	0.0614** (0.005)	Black	–0.0584** (0.004)
13–20 years	0.0600** (0.005)	Hispanic	–0.0088 (0.023)
21–27 years	0.0629** (0.005)	other race	–0.0419** (0.011)
> 27 years	0.0439** (0.006)	<i>Classroom characteristics</i>	
Graduate degree	0.0024 (0.002)	Percent nonwhite	–0.0219** (0.010)
Undergraduate institution (base = not competitive)	0.0188** (0.004)	Peer average achievement	0.1140** (0.004)
Very competitive	0.0049 (0.003)	Percent male	–0.0094 (0.009)
Competitive	0.0049 (0.003)	Advanced class	0.0278** (0.003)
Unranked	–0.0056 (0.007)	Class size	–0.0027** (0.000)
Teacher test score (normalized)	0.0071** (0.002)	<i>Student fixed effects?</i>	Yes
Type of license (base = regular)		<i>Regression information</i>	
Lateral entry	–0.0569** (0.007)	Number of observations	857,548
Lateral entry-Prior	0.0099 (0.010)	R-squared	0.784
Other license	–0.0737** (0.009)		
Certification (base = no cert.)			
certified in subject	0.0703** (0.009)		
certified in related subject	0.0511** (0.013)		
National Board Certification status (base = never certified)			
Precertification	0.0219** (0.006)		
Application year	0.0487** (0.008)		
Has certification	0.0494** (0.005)		

a. The dependent variable is normalized student achievement by subject for four cohorts of students. The model also includes 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.

mentary level. We then move to the licensure and certification credentials of primary interest for this study.¹⁶

A. Years of experience

Our measure of teaching experience includes all previous years of teaching, whether in North Carolina, or elsewhere. Because of our own prior research at the elementary level (Clotfelter, Ladd, and Vigdor 2006, 2007a, and 2007b) and that of others (for example, Hanushek, Kain, O'Brien, and Rivkin 2005), we expect the effect of an additional year 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 Table 4, all of the gains in achievement associated with teacher experience occur in the first five years of teaching. The coefficient for 1–2 years of experience is 0.0478 and for 3–5 years is 0.0608, with the difference being statistically significant. Though the coefficient rises to a peak of 0.0628 for a teacher with 21–27 years of experience, the difference between that and the one for 3–5 years of experience is not statistically significant. Thus we conclude that teachers with some experience are more effective than novice teachers, but, beyond the first five years, additional experience adds little to effectiveness.¹⁷ The magnitudes of the coefficients for the first five years of teaching reported here are generally comparable to those that emerged from our research on elementary school teachers. The flat pattern of coefficients for teachers with more than five years of experience at the

16. In addition to the basic model, we also estimated an alternative model with school, rather than student, fixed effects. Although the school fixed effects mitigate the bias in the estimates of the β coefficients associated with the nonrandom 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 the basic model. See Clotfelter, Ladd, and Vigdor (2007c) for a comparison of the basic and the alternative models based on a slightly earlier version of both models. The slight revisions to the basic model reported in the present paper do not alter the comparisons in any meaningful way. For most sets of the teacher credentials, the estimated coefficients from the alternative model are generally comparable to, but somewhat larger than, those reported in Table 4. This pattern suggests that the unmeasurable characteristics of students, such as their ability, and the strength of teacher credentials are positively correlated, and that student fixed effects are useful for reducing the resulting upward bias.

17. One interpretation of this pattern is that there is little or no additional learning on the job after the first five years of teaching in these high school courses. Another is that teachers continue to learn on the job, but more effective teachers have a higher propensity to opt out of the basic ninth and tenth grade courses than less effective teachers, or to stop teaching in North Carolina public schools entirely. To separate the effects of additional experience from other more permanent characteristics of teachers, we added teacher fixed effects to the basic model. Though the results depict a pattern of clearly rising coefficients on the experience variables, the coefficients are estimated very imprecisely, probably because of the inclusion of the student fixed effects. The estimation of an alternative model that includes school and teacher fixed effects, but no student fixed effects, generates a pattern of statistically significant rising coefficients (from 0.06 for 1–2 years of experience to 0.27 for more than 27). This pattern of rising coefficients echoes that of Kane, Rockoff, and Staiger (2006). Though such a pattern would support the case for trying to keep experienced teachers in the core high school courses, and possibly for retaining teachers who exit North Carolina public schools entirely, these results are biased upward because of the absence of the student fixed effects. In any case, as indicated by the results in Table 3, more than five years of experience is not predictive of higher achievement in practice.

high school level, however, contrasts with a rising pattern at the elementary level (Clotfelter, Ladd, and Vigdor 2007a, 2007b).¹⁸

B. Teacher education—advanced degrees and quality of undergraduate institution

In the basic model, 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 4 is the conclusion that having a graduate degree is not predictive of higher achievement compared to having a teacher without a graduate degree.

Further investigation at a more disaggregated level (not shown) generates a small positive coefficient of 0.004 for a master's degree (which is statistically significant at the 10 percent level) and an unexpected and surprisingly large 0.09 negative effect of having a teacher with a Ph.D. The latter coefficient is based on a very small number of teachers and may be spurious. With respect to master's degrees we find virtually no difference between teachers who received their master's degrees before entering the profession and those without master's degrees. However, teachers who received master's degrees after they began teaching appear to be somewhat more effective than those without a master's degree (with a statistically significant coefficient of about 0.008).¹⁹ This pattern differs quite markedly from the pattern that emerged in our previous research on elementary school teachers where the earning of a master's degree more than five years into teaching was associated with a negative effect on student achievement (Clotfelter, Ladd, and Vigdor 2007a and 2007b). In neither case can we separate the effects of the degree itself from the decision to obtain such a degree.

Following standard practice in the research literature, we assign to each teacher's undergraduate 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. We assign schools with a noncompetitive ranking to the base category.

Emerging from Table 4 is a positive and statistically significant coefficient of 0.0188 for teachers from a very competitive college and smaller and not significant

18. Our basic models in those papers differed somewhat from those presented here and we presented results for both lower and upper bound estimates. For math teachers, the lower bound estimates rose from 0.057 for teachers with 1–2 years of experience to 0.090 for those with 21–27 years of experience and for reading they rose from 0.023 to 0.067 over the same range. 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.

19. Actually, we included two separate variables, one for teachers who received their master's degrees between one and five years of starting teaching and the other for those who received them more than five years after starting teaching. Both coefficients are about 0.008.

coefficient of 0.0049 for teachers from a competitive college. These findings suggest 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. Nonetheless, the coefficients are quite small.

C. Teacher test scores

Teacher test scores are among the teacher credentials that most often emerge as statistically significant predictors of student achievement.²⁰ Most high school teachers in North Carolina have taken at least two Praxis II tests (one in content knowledge and one in content pedagogy) as part of their licensure requirements. Depending on when they were licensed they may have taken various other tests.²¹ 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 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 their normalized scores. Thus, in the basic regression, no attention is paid to the particular test or tests taken by each teacher.

Our basic specification for teacher test scores is linear. As shown in Table 4, teacher test scores enter with a positive coefficient of about 0.007, which is somewhat smaller than the coefficients of 0.011 to 0.015 that emerged in our prior research for teachers at the elementary grades (Clotfelter, Ladd, and Vigdor 2007a and 2007b). In an alternative specification (not shown), we allow for a more flexible form by using 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 relationship. In particular, the coefficient of the indicator for a high test score is small and insignificant while one for a low test score is -0.026 with a standard deviation of 0.005.

Beyond these specifications, we disaggregated the test scores by subject to determine the extent to which a teacher's knowledge of content and subject-specific pedagogy, as measured by her test results, affects her student's achievement in that specific subject. The standardized teacher test scores used in this specification refer only to the licensure tests that are relevant for that subject. Because there are no licensure tests specifically for algebra or geometry, we use the high school math test as the relevant test for both.²² The results of this disaggregated specification are

20. 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, and Ladd 2006a, 2007a, 2007b; Goldhaber, forthcoming).

21. Some teachers would have taken various forms of the old NTE tests in specialty areas and professional knowledge and some would have taken three Praxis II tests since the state required three for some specialty areas for a period in the 1990s.

22. The relevant tests are as follows: Biology: 0230 through 1993, 0231, 0233, and 0234 through 1999, 0234, and 0235 beginning in 2000; English: 0040 through 1993, 0041, 0042 and 0043 through 1999, 0041 and 0043 beginning 2000; math: 0060 through 1993, then 0061 and 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, and 0083 through 1999, 0081, and 0084 beginning in 2000. For teachers without the relevant test scores in the subjects they are teaching, we include an indicator variable that takes on the value one and set their test score equal to zero.

Table 5
Achievement effects by teacher test score and certification, by subject

	1	2	3
Teacher test scores by subject			
Math score	0.0472** (0.006)		0.0471** (0.006)
Biology score	0.0165** (0.003)		0.0162** (0.003)
ELP score	0.0002 (0.006)		0.0001 (0.006)
English test score	-0.0214** (0.006)		-0.0216** (0.006)
Certification by subject			
Math (for algebra and geometry)		0.1055** (0.014)	0.1100** (0.014)
Biology		-0.0132 (0.015)	-0.0087 (0.015)
ELP		0.0042 (0.009)	0.0054 (0.009)
English		0.1046** (0.012)	0.1033** (0.012)

Notes: a. All estimates come from variations of the full model reported in Table 3. See text for explanation. Each of the subject-specific test scores apply only to the relevant subject. Also included in the models are indicator variables by subject for teachers who have not taken a test 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.

shown in Columns 1 and 3 of Table 5. Those in Column 3 differ from those in Column 1 in that they are based on a model that also includes the subject-specific certification variables that are described below.

The reader should note that each of the subject-specific teacher test scores appears only in the student test-score observations for that subject. The clearest findings emerge for math and biology. A one standard deviation difference in a teacher's math test score is associated with a quite large and statistically significant 0.0472 standard deviation difference in student achievement in either algebra or geometry. The teacher test score in biology is also predictive of student achievement in biology but with a smaller coefficient. Our interpretation of the small and insignificant coefficient for Economics, Legal and Political systems (ELP) test variable is that the social studies test used for that subject measures quite imperfectly the knowledge

needed to teach ELP. The negative sign on the English test scores is unexpected. One possible, but at most speculative, explanation is that the English test is designed for a variety of courses that are more advanced than English I. A comparison of Columns 1 and 3 shows that the inclusion of subject-specific certification variables has little effect on the test score results.

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. As we show below, however, only the relatively large estimated effect of the math teacher test score is comparable in magnitude to the effects of many of the licensure and certification variables, to which we now turn.

D. Licensure type

Like other states, North Carolina requires that teachers be licensed in order 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). In response to concerns that licensure requirements impose unnecessary barriers to entry into teaching (Ballou and Podgursky 1998), many states have opened alternative routes into the teaching profession that require less up-front commitment of time. In North Carolina, the primary form of alternative entry 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 enroll in an approved teacher education program to complete the prescribed class work and must complete at least six semester hours of coursework each year. The first lateral entry license is issued for two years, and may be renewed for a third year.

We focus first on the three main categories of licensure, without reference to area of certification: regular, lateral entry, and "other." We further distinguish between teachers who have a lateral entry license at the time we observe them and those who had such a license in a prior year. "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 ten 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. Finally, the "other" category includes other forms of alternative entry, as well as provisional, temporary, and emergency licenses.²³

As shown in Table 4, students taught by teachers with a lateral entry license average 0.06 standard deviations lower than those taught by teachers with a regular license.²⁴ Prior lateral entrants, however, appear to be no less effective than teachers

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

24. Included in our category of lateral entrants are teachers provided through the Teach for America (TFA) Program. Such teachers differ from typical lateral entrants both in terms of the types of their undergraduate institutions and in terms of the support they receive from the Teach for America organization. A recent study based on the 23 North Carolina districts that have hired teachers shows positive outcomes for that subset of the lateral entry teachers at the high school level (Xu, Hannaway, and Taylor 2008).

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.²⁵ Students taught by teachers with “other” licenses have average test scores even further below those in the base category, but the coefficient does not differ from that for lateral entrants at standard levels of significance.

E. Certification by subject area

A second component of the licensing requirement is that teachers be certified by subject area. For the time period covered in this study, such certification required that a teacher both successfully complete an approved program of study in the subject area and earn passing grades on the appropriate Praxis II tests.²⁶

Table 4 shows the effects of certification by subject where results are aggregated across the five subject areas.²⁷ The coefficients 0.07 and 0.05, which are statistically distinct, indicate that being taught by a teacher who is certified in the subject she is teaching or in a related subject leads to higher test scores, and that the effects are large relative to those for the other teacher credentials. The estimated effects of certification, for example, are many times the size of those that are implied by a one standard deviation difference in test scores. We note, however, that the estimates are relative to a small group of teachers who are not certified.

These certification results are disaggregated by subject area in columns II and III of Table 5. The entries in these columns shed light on whether being certified in math contributes more to student achievement in algebra or geometry, for example, than does being certified in biology to achievement in biology. The entries in the two columns differ in that those in column III are based on models that also include subject-specific test scores. The similarity in the results across specifications highlights the fact that the subject-specific test scores, measured as continuous variables, exert effects on student achievement that are quite independent of that due to certification.

As was the case for the test score variable, the results for teachers of the two math courses, algebra and geometry, stand out. Being certified in math increases the achievement of a teacher’s students in a math course on average by about 0.11

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

26. In part because of the high pass rates on the PRAXIS tests for those teachers who successfully complete the required programs of study, the state has recently dropped the PRAXIS II tests for those teachers. In addition, it is now the case that a teacher who is already licensed in one area can become certified in another based on passing scores on the relevant PRAXIS tests alone.

27. These certification results, and also the subject-specific certification results in Table 5, are based on a relatively broad definition of certification. For example, certification in general science is included in the definition of certification for biology. The results in Table 4 are qualitatively similar when we use a stricter definition that would, for example, count general science as a related subject for biology. In the subject specific variations reported in Table 5, large standard errors make it impossible for us to sort out the effects of certification in the actual subject from those in a related subject.

standard deviations. This finding for math is fully consistent with earlier 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. The only other subject for which certification matters is English, where once again the estimated effects are large.

F. 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 (NBPTS), 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 and Sass 2007 and Goldhaber and Anthony 2007). A positive signaling effect emerges from Table 4 in the form of the positive coefficient of 0.0219 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 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. No evidence of a positive human capital effect emerged from our prior research on Board Certification at the elementary level.

G. Summary measure of the effects of credentials

We use these estimated coefficients to develop a summary measure of the effects of teacher credentials. Specifically, we compare the achievement effects of a teacher with weak credentials, defined as one at the tenth percentile in the predicted distribution of student achievement, where the predictions are based on teacher credentials alone, with those of a teacher with strong credentials, defined as one at the ninetieth percentile of the teacher distribution.²⁸ Based on the teachers in our sample, the difference in predicted student achievement between the two teachers is 0.23 standard deviations. Thus, by this metric a student with a weak teacher would be expected to perform 0.23 standard deviations lower than if she had a teacher with strong credentials. Though credentials may be bundled in various ways, it is clear from the estimated coefficients that novice or lateral entry teachers and those not certified in the field they are teaching or in a related field are most likely to be at the bottom of the distribution. We return in the conclusion to the question of whether this difference is large or small.

28. For this calculation the teachers are not weighted by the number of students they teach.

VI. Achievement Effects of Teacher and Classroom Characteristics

Also included in Table 4 are results for the characteristics of the teachers, such as their race and gender, and of their classrooms. With respect to classroom characteristics, classrooms with larger percentages of nonwhite students are associated with lower test scores for individual students in those classrooms. In contrast, classrooms with high average peer achievement and classes designated as advanced are associated with higher test scores. These coefficients could reflect true causal impacts of peer characteristics on outcomes, but we cannot rule out the possibility that they may be confounded by correlation between peer ability and unmeasured teacher characteristics. 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 indicates that being in a class with five fewer students than average would increase student achievement by only 0.0127 standard deviations.

Perhaps the most arresting results in the table are the large negative coefficients for black, “other” race and male teachers, coefficients that emerge even though we have controlled for their credentials. (Teachers in the “other” race category include those who are not identified as white, black or Hispanic.) We examine the achievement effects of race and gender in more detail in Table 6, which 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 Table 4. Since all the entries in the table are variations of the basic model, which includes student fixed effects, the race and gender of individual students are not included. Student characteristics only can be included in interactive form.

The first variation 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 -0.105 . 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/ethnicity. The student categories are white, black, and “other.” This “other” category includes all students who identify themselves as neither white nor black. It includes Hispanic and Native American students as well as those who identify themselves in some other way or for whom we have no racial or ethnic identification. Here the main findings are the large negative coefficients for a black teacher teaching a white student or a Hispanic teacher teaching a student in the “other” category. The latter effect may be spurious because of the small number of Hispanic teachers. The large negative effect associated with black teachers teaching white students, however, is cause for concern. In contrast to this large negative effect, black teachers appear to be more successful with black students and equally effective as white teachers are with white students.

Table 6
Teacher Characteristics with interactions^a

	Basic model	Variation 1	Variation 2
Gender (base = female)			
Male	-0.0546** (0.003)		-0.0542** (0.003)
Race (base = white)			
Black	-0.0584** (0.004)	-0.0583** (0.004)	
Hispanic	-0.0088 (0.023)	-0.0081 (0.023)	
Other	-0.0419** (0.011)	-0.0417** (0.011)	
Teacher and student gender (base = female <i>T</i> and female <i>S</i>)			
Female <i>T</i> and male <i>S</i>		0.0125 (0.042)	
Male <i>T</i> and female <i>S</i>		-0.1048** (0.003)	
Male <i>T</i> and male <i>S</i>		0.0120 (0.042)	
Teacher and student race (base = white teacher and white student)			
White <i>T</i> and black <i>S</i>			0.0357** (0.013)
White <i>T</i> and nonwhite or nonblack <i>S</i>			0.0548** (0.016)
Black <i>T</i> and black <i>S</i>			0.0098 (0.014)
Black <i>T</i> and white <i>S</i>			-0.0834** (0.005)
Black <i>T</i> and nonwhite or nonblack <i>S</i>			0.0028 (0.020)
Hispanic <i>T</i> and nonwhite or nonblack <i>S</i>			-0.0815 (0.056)
Hispanic <i>T</i> and white <i>S</i>			0.043 (0.031)
Hispanic <i>T</i> and black <i>S</i>			-0.0177 (0.034)

a. Entries in first column are the same as those in Table 3. In the variations, the indicated variables replace the race and gender variables in the full model as indicated. ** 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.

VII. Policy Implications and Conclusions

As we reported above, a reasonable estimate of the difference in achievement effects of having a weak rather than a strong teacher is about 0.23 standard deviations of the current test score distribution. This figure is somewhat larger than the 0.20 effect size often deemed small or moderate in the education literature. Two considerations suggest, however, that the figure may significantly understate the size of the true impact of teacher credentials on student achievement. First, as noted by Boyd et al. (2008), because education is a cumulative process, a case can be made that the estimate should be interpreted relative to the standard deviation of gain scores rather than of levels. The second consideration relates to measurement error. According to these authors, a correct interpretation of the estimated coefficients should account for the measurement error in the reported test scores by comparing them to the dispersion in “true” achievement gains rather than in the measured achievement gains. In their empirical analysis based on teacher credentials and fifth grade test scores in New York, they found that the true effects of teacher credentials were about four times larger than estimates that emerged from their regressions (Boyd et al. 2008). Based on that logic, the 0.23 estimate reported in this study would be deemed a large effect.

In terms of the specific credentials of high school teachers, the most important new findings to emerge from this study relate to teacher licensure and certification. In particular, this study provides new evidence that subject-specific certification, particularly in math and English, generates higher student achievement, that National Board Certification generates positive effects, and that, at least during their initial years of teaching, lateral entry teachers on average are less effective than teachers with regular licenses.

Also of potential policy interest is the proportion of the total variation in overall teacher quality, defined in terms of teachers’ success in raising student test scores, accounted for by the variation in teacher credentials. 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 of overall teacher quality is harder to pin down.

Typically, researchers estimate that figure from the distribution of the teacher fixed effects that emerge from achievement models in which all teacher credentials and time-invariant teacher characteristics are replaced by indicator variables for every teacher in the sample. Using that method with our data, we obtain a rough estimate of the standard deviation of overall teacher quality of about 0.51.²⁹ That estimate, however, undoubtedly overstates the variation because we have made no adjustment for the measurement error highlighted in previous studies, such as Hanushek, Kain, O’Brien, and Rivkin (2005), which is based on Texas data, and because the inclusion of the student fixed effects in our model is likely to generate significant noise in the estimates of the teacher fixed effects. Adjusting the figure down by the percentage

29. 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.

adjustment for measurement error that emerged in the Texas study, we conclude that 0.34 is a reasonable upper bound estimate of the standard deviation of the distribution of overall teacher quality in North Carolina. Based on that figure, the variation in teacher credentials would account for at least a fifth of the overall distribution in teacher quality.

The discrepancy between the overall variation in teacher quality and that predicted by credentials alone implies that it would be a mistake for policymakers to put so much weight on measurable credentials in determining teacher quality that they ignore other contributors to teacher effectiveness, many of which only can be determined by observation at the school or classroom level³⁰. 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 sufficiently important that they can be used as the basis for policies to improve 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 and 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.³¹ Table 7 summarizes the patterns for five sets of credentials from that study.

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, those from less competitive institutions, and those 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 (defined in terms of standard deviations around a mean of zero) and the lowest percentages of Board-certified teachers.

Using the current data set, we provide a more detailed look at how teacher characteristics are distributed by type of student among Algebra I courses. Table 8, which is based on the data for the 2002–2003 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

30. 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.

31. 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 by school.

Table 7
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).

students, those who are still in high school and are taking Algebra I. Thus a disproportionate number of disadvantaged members of the age cohort may be excluded from the sample. The credentials are all defined to represent weaker qualifications. Hence, in all cases, a larger number signifies 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 noncollege-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 four percentage point difference in the probability of having a lateral entry teacher translates into only a 0.0024 difference in predicted achievement (.04 times the 0.06 estimate from Table 4). 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, which is clearly tiny relative to the overall achievement gap between black and white students. Following

Table 8
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 Undergraduate 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
Noncollege 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

the same logic as we used above, this difference looms larger 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 I.³² 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.

In sum, the systematic differences in the distribution of teacher credentials by their students' race, gender, and education level of parents combined with the evidence presented in this paper that teachers' credentials are predictive of student achievement should be cause for serious policy concern. In addition, the findings of unexpectedly large negative coefficients for black teachers teaching white students and male teachers teaching female students are troubling and worth additional attention. We leave for future research by both ourselves and others the investigation of the benefits and costs of various policy mechanisms for promoting a more equitable distribution of high school teachers, for promoting credentials-based teacher policies that have the promise of raising student achievement, and for exploring further the teacher-related gender and race results that emerge from this study.

References

- Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25(11):95–135.
- Allensworth, Elaine, and John Easton. 2007. What Matters for Staying On-Track and Graduating in Chicago Public Schools. Research Report. Chicago: Consortium on Chicago School Research at the University of Chicago.
- Ballou, Dale, and Michael Podgursky 1998. "The Case Against Teacher Certification." *Public Interest* 132:17–29.
- Boyd, Dan, Pam Grossman, Hamp Lankford, Susanna Loeb, and Jim Wyckoff. 2008. "Overview of Measuring Effect Sizes: The Effect of Measurement Error." Brief 2. National Center for the Analysis of Longitudinal Data in Education Research. November.
- Clotfelter, Charles, Helen Ladd, and Jacob Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources* 41(4):778–820.
- . 2007a. "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects." *Economics of Education Review* 26(6):673–82. December.
- . 2007b. "How and Why Do Teacher Credentials Matter for Student Achievement?" Working Paper No. 12828, NBER.
- . 2007c. "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." Working Paper No. 13617, NBER.
- Clotfelter, Charles, Helen Ladd, Jacob Vigdor, and Justin Wheeler. 2007. "High Poverty Schools and the Distribution of Principals and Teachers." *North Carolina Law Review* 85(5):1345–79.

32. For students who took Algebra I in eighth grade, we use their seventh grade math scores.

- Darling-Hammond, Linda. 2002. "Research and Rhetoric on Teacher Certification: A Response to 'Teacher Certification Reconsidered.'" *Education Policy Analysis Archives* 10(36).
- Ehrenberg, Ronald, and Dominic Brewer. 1994. "Do School and Teacher Characteristics Matter? Evidence from High School and Beyond." *Economics of Education Review* 13(1):1–17.
- Goldhaber, Dan. 2004. "Why Do We License Teachers?" In *A Qualified Teacher in Every Classroom: Appraising Old Answers and New Ideas*, ed. Frederick Hess, Andrew Rotherham, and Kate Walsh, 81–100. Cambridge: Harvard Education Press.
- . 2007. "Everyone's Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness?" *Journal of Human Resources* 42(4):765–94.
- Goldhaber, Dan, and Dominic 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):505–23.
- . 1997b. "Evaluating the Effect of Teacher Degree Level on Educational Performance." In *Developments in School Finance 1996*, ed. William Fowler, 197–210. Washington, D.C.: National Center for Education Statistics.
- . 2000. "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." *Educational Evaluation and Policy Analysis* 22(2):129–45.
- Goldhaber, Dan, and Emily Anthony. 2007. "Can Teacher Quality be Effectively Assessed? National Board Certification as a Signal of Effective Teaching." *Review of Economics and Statistics* 89(1):134–50.
- Gordon, Robert, Thomas Kane, and Douglas Staiger. 2006. "Identifying Effective Teachers Using Performance on the Job." The Hamilton Project, Discussion Paper 2006–01. Washington, D.C.: Brookings Institution.
- Hanushek, Eric. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis* 19(2):141–64.
- Hanushek, Eric, John Kain, Daniel O'Brien, and Steven Rivkin. 2005. "The Market for Teacher Quality." Working Paper No. 11154, NBER.
- Harris, Douglas, and Tim Sass. 2007. "The Effects of NBPTS-Certified Teachers on Student Achievement." *Journal of Policy Analysis and Management* 28(1):55–80.
- Kane, Thomas, Jonah Rockoff, and Douglas Staiger. 2006. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27(6):615–31.
- Monk, David, and Jennifer King. 1994. "Multi-level Teacher Resource Effects on Pupil Performance in Secondary Mathematics and Science: The Case of Teacher Subject-Matter Preparation." In *Contemporary Policy Issues: Choices and Consequences in Education*, ed. Ronald Ehrenberg, 29–58. Ithaca: ILR Press.
- Monk, David. 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.
- Rockoff, Jonah. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review Papers and Proceedings* 94(2):247–52.
- Rothstein, Jesse. 2008. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." Working Paper No. 14442, NBER.
- 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, Kate. 2001. *Teacher Certification Reconsidered: Stumbling for Quality*. Baltimore: Abell Foundation.

Xu, Zeyu, Jane Hannaway, and Colin Taylor. 2007. "Making a Difference? The Effects of Teach for America in High School." Working Paper No. 17, National Center for Analysis of Longitudinal Data in Education Research. April.

Appendix

Table A1

Unmatched and matched students by student cohort and by subject

Tenth 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 (ELP)						
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