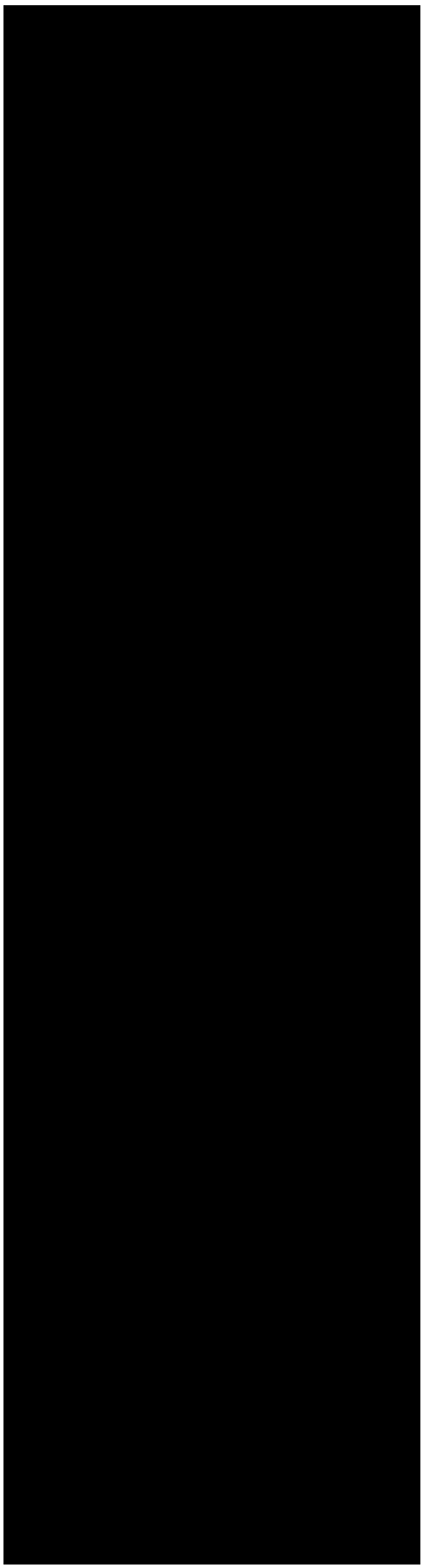


TERRY
SANFORD INSTITUTE
OF PUBLIC POLICY
DUKE

Teacher Quality and Minority Achievement Gaps

Charles Clotfelter
Helen F. Ladd
Jacob Vigdor

Working Paper Series
San04-04

October 2004

TERRY
SANFORD INSTITUTE
OF PUBLIC POLICY
D U K E

Teacher Quality and Minority Achievement Gaps

Charles Clotfelter, Helen F. Ladd, and Jacob Vigdor
Sanford Institute of Public Policy
P.O. Box 90243
Duke University
Durham, N.C. 27708

Charles.clotfelter@duke.edu
hladd@pps.duke.edu
jvigdor@pps.duke.edu

October 2004

This paper was prepared for the 2004 annual research conference of the Association for Public Policy Analysis and Management in Atlanta, Ga. The paper summarizes the results of our recent research on the contribution of teacher quality, as measured by the characteristics of teachers, to the gap in achievement between minority and white students. We are grateful to the Spencer Foundation for financial support for this research, to the North Carolina Department of Public Instruction for making available the administrative data, to the North Carolina Education Research Data Center for preparing the data for our use, and to the many able research assistants who worked with us on the various articles on which this summary is based.

Teacher Quality and Minority Achievement Gaps

Abstract

In this paper we summarize the findings from five separate papers, with the goal of providing a more complete picture than is normally possible in a single paper, of the extent to which the variation in teacher characteristics contributes to minority achievement gaps. All five papers are based on a rich administrative data set that includes information on all students and all teachers in North Carolina. We document the extent to which students in North Carolina are segregated by race not only across schools but also across classrooms within schools; the extent to which teachers with stronger qualifications are overrepresented not only in schools serving more advantaged students, but also in the classrooms serving such students within such schools; the extent to which various teacher characteristics affect student achievement, and finally, the extent to which North Carolina's relatively sophisticated school-based accountability program has exacerbated the challenges that low-performing schools face in retaining teachers. Together our findings clearly implicate the unequal distribution of teachers, as defined by their characteristics, as one of many factors that contribute to minority achievement gaps. Of particular importance is the uneven distribution of novice teachers across schools and classrooms.

Keywords: achievement gap, race, education

Introduction

The purpose of this paper is to summarize the results of a series of papers that use newly available administrative data for North Carolina to examine how the distribution of teachers, as described by their qualifications, across schools and classrooms contributes to achievement gaps, particularly between minority students and white students. The basic idea is that if black students are more likely than white students to be exposed to teachers with a specific characteristic, for example, having no experience, then the distribution of teachers defined by that characteristic contributes to the black-white test score gap, assuming, of course, that that characteristic affects student achievement. Our rationale for focusing on measurable teacher characteristics is that they are more amenable to systemic policy interventions than are other measures of teacher quality.

For this research we have had access to data from the North Carolina Department of Public Instruction made available to us through the North Carolina Education Research Data Center at Duke University (henceforth, the Data Center). The Data Center, which has been supported by initial and follow-up funding from the Spencer Foundation, receives administrative data on all student test scores, teachers, and schools from the North Carolina Department of Public Instruction. Before providing that information to researchers, the Data Center removes all identifying information and replaces it with encrypted identifiers, cleans the data, and matches data at the level of the individual student and teacher from year to year. The result is an extremely rich data set that permits us to follow both teacher and students over time and, in may cases, to match teachers to students.

This summary is based on the five papers included in the reference. The short names in bold for each paper are the names by which we refer to each of them in the rest of this summary paper. In the following sections of this paper, we describe our research and draw on selected findings from these papers to support our conclusion that the uneven distribution of teacher characteristics is clearly implicated as one of many factors that contribute to minority achievement gaps. Of particular importance is the uneven distribution of novice teachers across classrooms and schools.

Our goal here is to provide a readable summary of a large number of interrelated findings. That inevitably means that we must ignore many details along the way. Those details can be found in the specific papers.

Segregation of Students by Race[1]

Given our interest in the extent to which teachers, as defined by their qualifications, contribute to minority achievement gaps, it was logical to begin our research by examining the extent to which students are segregated across classrooms by

[1] See our **Segregation** and **Segregation (Short)** papers.

their race.[2] If there were no racial segregation, there would be no average differences in the types of teachers to which students of different races were exposed. Thus racial segregation at the classroom level is a necessary, but not sufficient, condition for students of different races to be exposed to teachers with differing characteristics. It is not sufficient because, even in the presence of racially segregated classrooms, students of different races could face teachers with similar characteristics. That would occur if school administrators chose to distribute equally qualified teachers evenly across the classrooms. Thus the examination of racial segregation is only a first step, albeit an important one, in the examination of whether students of different races are exposed to teachers with different characteristics.

In contrast to many administrative or other data sets that, at best, allow the researcher to measure racial segregation at the school level, our rich North Carolina data set permitted us to measure segregation at the classroom level. By using school activity reports that provided information on the racial breakdown of students in each "activity" during the day, we were able to construct measures of racial segregation at the classroom level for students in grades 1 and 4 (where most classrooms are "self-contained"), and for all English classes in grades 7 and 10. Our gap-based measure of classroom segregation is based on an exposure rate (E_k) of white students to nonwhite students in district k, where the exposure rate can be interpreted as the proportion of nonwhite students in the typical white student's classroom. Thus:

$$E_k = (\Sigma W_i \, \%NW_i)/\Sigma W_i \,, \qquad (1)$$

where k refers to the district and i to the classroom. W is white students and NW is nonwhite students.

The segregation index (S_k) is the gap between the maximum possible exposure rate in the district and the actual exposure rate expressed as a fraction of the maximum. For these calculations the maximum exposure rate is given by the percent of nonwhite students in the district and would be obtained if all nonwhite students were distributed evenly across classrooms in the district.[3] Thus:

$$S_k = (\%NW_k - E_k)/ \%NW_k \,. \qquad (2)$$

The segregation index ranges from 0, which indicates no segregation, to 1, which indicates full segregation.

---

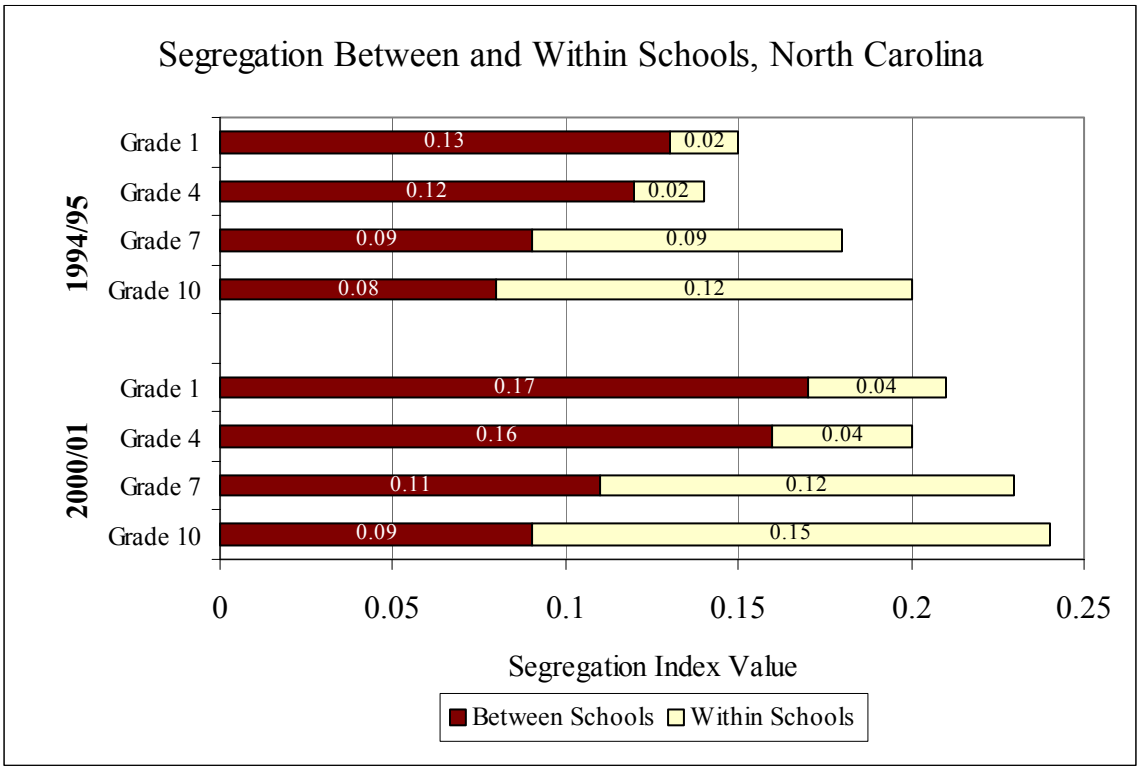
[2] We began our research with this issue not only because it is a logical first step but also because the data files necessary for this investigation were available from the North Carolina Education Research Data Center (NCERDC) far earlier than were the data files for teachers. In fact it took close to two years from the start of the (NCERDC), which also corresponded to the start of this project, to get the teacher licensure files in shape to be used. That delay reflected the difficulty of reading huge files that had been stored using outdated software along with some of the standard difficulties of converting administrative data that were collected and used for a different purpose to their use for research.

[3] In current ongoing research we are modifying the expression for schools in which there are very few students of a particular race to take account of the fact that it may not be possible to distribute students evenly across classrooms.

In North Carolina, nonwhite students are mainly African American. Although Hispanic students still account for only a small share of the total, they are growing rapidly. In 2000/01 the share of African Americans was 31.1 percent, of Hispanics was 4.4 percent and other nonwhite 3.3 percent. The results in the following table are based on the white-nonwhite distinction. In our published work we also calculate segregation indices for other pairings, such as white-black and white-Hispanic.

One nice feature of this gap-based measure of segregation is that it can be decomposed into the racial segregation that occurs because of an uneven distribution of non-white students across schools within each district (*between* school segregation) and that which occurs because of uneven distributions of nonwhite students across classrooms within schools (*within* school segregation). The following figure shows our segregation results for the entire state (i.e. summed across the state's 117 districts) in 2000/01 for each of the four grades and for two years, 1994/95 and 2000/01.

Figure 1.



Based on the data in **Segregation,** Table 4.

Several points are worth noting. First, in 1994/95 the between-school segregation (as depicted by the dark portion of each bar) in North Carolina was quite low both compared to other parts of the country, and, most obviously, relative to the state in the pre-1968 period when segregation indices would have been close to 1.00. In addition between-school segregation was higher in the elementary grades than at the middle and

high school level. That pattern presumably reflects the fact that elementary schools often serve neighborhoods that are themselves racially segregated while middle and high schools are larger and typically draw students from a broader and more diverse geographic area. The picture changes, however, once one includes as part of the segregation measure the within-school component as depicted by the light component of each horizontal bar. Because within-school segregation is higher in the upper grades, the total amount of segregation is higher in the upper grades than at the lower grades. Stated differently, focusing attention only on the segregation between schools misses more than half of the story at the middle and high school level.

A final observation is that total segregation at all four grade levels increased during the 1994/95 to 2001/02 period. Segregation of students in 10th grade English, for example, increased from 0.20 to 0.24, or by 20 percent. Though not shown here, a closer look at the changes in segregation separately for each of the five largest districts in the state as well as for six other groups of smaller districts in various parts of the state shows a similar rise in segregation over time.

The bottom line is that though still very low compared to the 1960s, segregation at the classroom level in North Carolina is still sufficiently high to warrant concern about the possibility of differential access by race to teachers with strong qualifications. Even more important is that segregation has been growing over time, a phenomenon that has been noted in other southern states as well, and that opens up the possibility of even greater differential access to high-quality teachers in the future.

The Matching of Teachers and Students [4]

We next turned to the empirical question of whether black and white students are differentially exposed to teachers with particular qualifications. We restricted our analysis initially to a single characteristic, whether or not a teacher was a novice, that is, one with no prior experience. Our decision to highlight the distribution of novice teachers was based in part on the type of data that was available at that point in the project and on our reading of the existing literature on teacher effectiveness. Several studies, including most notably the careful work by Hanushek, Kain and Rivkin for Texas, provided evidence that teachers with no, or very little, experience were less effective in the classroom that were those with more experience. Our own subsequent work, described below, confirmed that, regardless of how effective they may eventually become, novice teachers in North Carolina are less effective in raising student achievement than teachers with more experience.

Preliminary descriptive regressions based on district-level data indicated that novice teachers were overrepresented in districts with higher proportions of minority students, and that was true even when we controlled for other district characteristics, such as total enrollment and the percent of students eligible for a free or reduced price lunch. This cross-district pattern most likely reflects not only lower salaries in many high

[4] This section is based on our papers, **Who Teaches Whom** and **Teacher Sorting**.

minority districts in North Carolina but also the tendency of teachers to prefer to teach in districts with more advantaged students.

But those regressions represent an incomplete picture in that they provide no information on the extent to which a typical minority student is more or less likely than a white student to have a novice teacher within a district. To motivate our within-district empirical analysis we developed a simple conceptual model of the behavior of school administrators, interpreted either as district superintendents or school principals, in which the matching of students to teachers is the outcome of a complex set of policies relating to the distributions of both students and teachers. One possibility is that school administrators consciously design policies to put minority students in classrooms with less-experienced teachers because of their own racial prejudice. The purpose of our model is to show that under various sets of conditions, that same outcome might well occur even in the absence of any racial bias on the part of school administrators.

Consider, for example, administrators who are unbiased and whose only goal is to allocate students and teachers in the way that maximizes student learning. For simplicity, we refer here to a principal who is distributing students of only two types – hard-to-educate and easy-to-educate – and teachers of only two types – low quality and high-quality -- among equal sized classrooms within her school. If the principal is unconstrained by pressures from teachers and parents, the way for her to maximize student learning depends only on the technology of learning. Under a reasonable set of assumptions about that technology, including, for example, that high-quality teachers produce more learning than low-quality teachers for any mix of students, that the learning of any student is lower the higher is the proportion of hard-to-educate students in the class, and that high-quality teachers are more effective than low-quality teachers in classrooms with large proportions of hard-to-teach students, the principal would have an incentive to place a disproportionate number of the hard-to-teach students in classrooms with the high-quality teachers. If all the teachers were the same quality, whether they be high or low quality, the principal would maximize student learning by mixing students of different types in each classroom.

School principals, however, are not unconstrained. Instead they have to respond to a variety of pressures, including those from teachers and parents. High-quality teachers, who are likely to prefer to be in classrooms with large numbers of easy-to-teach students, may exert pressure on the principal by threatening to move to another school or school district if they are given undesirable teaching assignments. Parents of middle-class students, who lobby school principals to have their children assigned to high-quality teachers, can exert pressure in the form of vocal complaints about the administrator or by the threat of moving to another public, charter or private school. As we show with the help of various simplified variations of the model, the effects of these pressures could well be sufficient to induce the principal to depart from the policies she would otherwise pursue and, instead, assign the high-quality teachers disproportionately to the students who are easier to teach.

Because we did not develop the model with reference to particular racial groups, we can apply it to such groups only if we are willing to posit that students from different racial groups are disproportionately represented in the two categories of students. To the extent that African American students in North Carolina are more likely to come from households with lower income and lower parental education than white students, they could well be harder to educate on average than their white counterparts. If that were the case, this model would provide an explanation for differential exposure of black and white students to novice teachers within districts.

With one exception, we have done no empirical testing of the assumptions and predictions of the model. (The exception is discussed below.) Nor have we specified the parameters that we would need to make specific predictions about the outcomes of the complex process of assigning students and teachers. At this point, we simply use the model to illustrate the plausibility of within-district differences in the types of teachers to which black and white students are exposed and, thereby, to motivate our descriptive analysis of the extent to which black and white students are disproportionately exposed to novice teachers across classrooms in North Carolina. The research focuses on 7th graders in math and English classes. In effect we make calculations analogous to the exposure rates we calculated above (equation 1), but this time for the probability that a typical black student and a typical white student is exposed to a novice teacher. The basic findings are shown in Table 1.

As we have done in much of our research, we report results here for the state as a whole, for the five largest districts, and for six sets of other districts: urban and rural districts in the three main areas of the state, the mountain, piedmont and coastal areas. The table shows that the probability that a typical black seventh grader in North Carolina faces a novice teacher in math is 0.128, which exceeds the probability for a typical white seventh grader by 54 percent. In 7th grade English classes the black disadvantage is somewhat smaller, at 38 percent. Though the magnitudes of the racial differences differ across individual districts and sets of districts within the state, in all but two cases black students are at a disadvantage relative to white students, and in some cases by a substantial amount.

The state-wide differences in exposure rates between black and white students shown in Table 1 can be decomposed into differences between districts, between schools within districts, and between classrooms within schools, as shown in Table 2. The top two rows of the table indicate that all three levels are implicated. Somewhat more than a third of the differences for both math and English reflect different average proportions of novice teachers in seventh grade classrooms across districts in the state, a similar share reflects differences across schools within districts, and about a quarter reflects differences between classrooms within schools. These decompositions imply that even if policy makers could find a way to equalize the proportions of novice teachers across districts, more than 60 percent of the disadvantage faced by black students would still remain, unless some way were found to change the way novice teachers are distributed within districts.

A comparison of the results for Mecklenburg and Wake school districts indicates the potential for significant variation across districts in the extent to which the differential exposure is between or within schools. In Mecklenburg, the state's largest district, about 70 percent of the difference arises because the schools that black students attend tend to have higher proportions of novice teachers than do the schools that white students attend, and only 30 percent is attributable to differences within schools. In Wake, the state's second largest district, the situation is reversed with almost two-thirds of the racial differential occurring across classrooms within schools rather than across schools.

Table 1. Exposure to a Novice Teacher, 7th Grade Math and English, 2001

	Math			English		
	Black	White	% Diff.	Black	White	% Diff.
State of NC	0.128	0.083	54*	0.106	0.077	38*
Five Largest Districts						
Mecklenburg (Charlotte)	0.182	0.137	33*	0.167	0.119	40*
Wake (Raleigh)	0.122	0.076	67*	0.167	0.110	52*
Guilford (Greensboro)	0.094	0.077	22*	0.119	0.119	0
Cumberland (Fayetteville)	0.245	0.156	57*	0.090	0.084	13
Forsyth (Winston-Salem)	0.153	0.144	6	0.119	0.060	98*
Urban districts						
Coastal	0.140	0.063	122*	0.062	0.050	24*
Piedmont	0.083	0.080	4	0.131	0.065	100*
Mountain	0.060	0.057	5	0.080	0.079	1
Rural districts						
Coastal	0.159	0.060	165*	0.027	0.038	-29
Piedmont	0.091	0.064	42*	0.108	0.079	37*
Mountain	0.104	0.096	8	0.105	0.077	36*

Source: Calculations by the authors based on data from the North Carolina Department of Public Instruction. * denotes statistical significance at the 5 percent level.
[See **Who Teaches Whom?** Table 3.]

Table 2. Decomposition into District, School, and Classroom Effects, 7th Grade Math and English, 2001

	Total black-white difference	District effect	School effect	Classroom effect
NC State				
Math	0.0451	0.0171 38%	0.0165 37%	0.0114 25%
English	0.0295	0.0106 36%	0.0104 35%	0.0085 29%
5 largest districts				
Mecklenburg				
Math	0.0447	--	0.0313 70%	0.0135 30%
English	0.0567	--	0.0391 69%	0.0176 31%
Wake				
Math	0.0464	--	0.0170 37%	0.0294 63%
English	0.0202	--	0.0075 37%	0.0126 63%
Guilford				
Math	0.0170	--	-.0002 0%	0.0171 100%
English	-0.0007	--	-0.0085 < 0 %	0.0078 >100 %
Cumberland				
Math	0.0887	--	0.0579 65%	0.0309 35%
English	0.0057	--	-0.0086 < 0 %	0.0143 >100 %
Forsyth				
Math	0.0084	--	0.0090 >100 %	-0.0005 < 0 %
English	0.0590	--	0.0485 82%	0.0104 18%

Source: **Who Teachers Whom?** Table 4. A negative number signifies that the probability of exposure to a novice teacher is higher for white students than for black students. .

This pattern reflects Wake's ongoing efforts to balance its schools, initially through the use of magnet programs and, more recently, through school assignment programs specifically designed to distribute low-income and low performing students relatively evenly across schools.

Not shown here is additional analysis that explores the extent to which the greater exposure of black students to novice teachers is the result of tracking of students into remedial, standard, or advanced courses. The data indicate that black students are disproportionately represented in remedial classes and underrepresented in advanced courses and, further, that few novice teachers are used to teach advanced courses. Thus, tracking is clearly one of the mechanisms through which black and white students end up with differential exposure to a novice teacher. It is clearly not the only mechanism, however. Even when the analysis is limited to standard-level courses, black students are still far more likely than white students to have a novice teacher.

We present an alternative, and more comprehensive, picture of how teachers, as defined by their measurable qualifications, are distributed relative to students in our **Teacher Sorting** paper. Our focus there is the matching of fifth grade students with their teachers.[5] In that paper we refer to two processes that lead to what we refer to as positive matching of teachers and students. Positive matching occurs when the teachers with stronger observable qualifications end up in classrooms and schools with the more advantaged and higher performing students. One process is teacher sorting, by which we mean the tendency of teachers to prefer to teach in schools and classrooms with more advantaged students. The other process is teacher shopping, by which we mean the tendency of middle class parents to lobby to get the best teachers for their children within schools. The outcomes, however, are not determined by these two processes alone. Instead outcomes are also affected by the willingness of school administrators to accede to the pressures from teachers and parents. Hence the extent to which these processes lead to positive matching in practice is an empirical question.

Table 3 summarizes how fifth grade teachers end up being sorted across schools in North Carolina. Down the left hand column are various qualifications of teachers: their experience, the selectivity of their undergraduate college as measured by Barrons's, their scores on the teacher licensure test (normalized to mean 0 and standard deviation 1 to make the results comparable over time), whether or not the teacher was national board certified, and whether the teacher has an advanced degree. The entries in the table are the average characteristics of the students in the schools in which teachers of each type were teaching in 2000/01. Each category of student characteristics is defined so that a higher entry represents more advantaged students, whether defined by race, eligibility for free and reduced price lunch, education level of the students' parents, or student achievement (as measured by a student's prior year test score).

[5] We focus on fifth grade students in this analysis because the ultimate purpose of the paper is to examine how teacher sorting and teacher sorting affect the assessment of teacher effectiveness. For that purpose, we needed to match teachers with specific students, something we are able to do more successfully at the elementary than at the middle and high-school level. .

Table 3. Teacher sorting: Characteristics of students taught by the typical teacher having specified qualification, North Carolina schools offering 5th grade

Teacher qualification	Percent White	Percent not receiving subsidized lunch	Percent with parents who are college graduate parents	Mean prior year test score (Z)
Teacher experience:				
0 to 1 year	58.0	51.8	22.9	-0.134
2 to 5 years	58.2	54.4	23.8	-0.072
6 or more years	62.8	54.5	23.5	0.000
Barron's College Rank:				
Less competitive	53.7	49.8	20.3	-0.206
Competitive	64.4	57.1	24.4	0.118
Very competitive	59.3	58.2	30.4	0.126
Not Ranked	58.8	53.5	24.9	-0.047
Licensure test score:				
Z-score < -1	51.2	46.4	18.2	-0.306
-1 < Z-score < 1	62.9	56.0	24.3	0.054
Z-score > 1	66.2	58.4	26.8	0.158
Nat'l Board Certification				
No	61.0	54.4	23.4	0.000
Yes	65.0	57.6	23.8	-0.002
Advanced Degree				
No	60.0	53.5	22.9	-0.043
Yes	64.9	57.8	25.2	13.8
Overall mean	61.1	54.5	23.5	0.000

Note: For teachers with a given qualification, table entries are averages of school-wide figures computed over those schools with at least one such teacher. Using *F*-tests, the hypothesis of student characteristic equality across teacher qualification categories is rejected in all except the following cases: teacher experience and percent of students with parents who are college graduate parents; teacher national Board Certification and all four student characteristics. Source: **Teacher Sorting, Table 2.**

Consider, for example, teachers as described by their scores on the state's licensure tests. The table shows that teachers with the highest test scores (those more than one standard deviation above the mean) were in schools with higher proportions of white students, higher percentages of non-poor students, higher percentages of students whose parents were college graduates, and students whose prior year test scores were well above the mean. Analogously, teachers with low test scores taught in schools with lower proportions of white students, non-poor students, and students whose parents had college degrees, and in schools with low average achievement as measured by prior year test scores. This pattern of positive matching occurs quite consistently for most of the other teacher characteristics, although not all the relationships are monotonic and, as noted in the table, a few are not statistically significant. Overall, though, the table provides strong evidence of positive matching of teachers to students across schools.

In Table 4, we switch the focus to the matching of teachers and students across classrooms within schools. The sets of teacher qualifications are the same as those in the previous table. For this table, however, the student characteristics are for classrooms relative to the average for each school. Thus, for the first three columns, if teachers described by a particular qualification were evenly distributed across classrooms in each school, the entries would all be 1.00; for the final column the entries would be 0.000.

Consider, once again, teachers as described by their licensure test scores. The table shows that in all cases the teachers with the highest test scores are teaching the students within each school who are more advantaged along each dimension. Analogously, the teachers with the lowest test scores are teaching in the classrooms within each school with below-average percentages of white students, non poor students, and students whose parents have college degrees, and students with below-average achievement as measured by their prior year test scores. Of particular note are the patterns for National Board certified teachers. The large positive entries for the "yes" category indicate that such teachers are far more likely to be in the classrooms with the more advantaged and higher achieving students.

We note that most of the entries in the column labeled "Percent white" are quite close to the mean of one. This pattern is consistent with our previous finding of little racial segregation of students across classrooms within elementary schools. As we noted earlier, a prerequisite for students of different races to be taught by teachers with different qualifications is for them to be segregated by race into different classrooms.

Of interest are the larger differences that emerge for the other three student characteristics, particularly those in the final two columns – having parents who are college graduates and students' prior year test scores. The patterns are consistent with the notion that middle class parents are relatively successful in getting their children into the classrooms with the more qualified teachers, and also with the observation that the more highly qualified teachers have more bargaining power within a school than do their colleagues with weaker qualifications. Particularly noteworthy are the different mixes of students in the classrooms taught by teachers with and without National Board certification. The teachers with that prestigious certification end up in classrooms with

Table 4. Evidence of teacher shopping: classroom characteristics for teachers with varying qualifications, relative to school, North Carolina schools with more than one 5th grade class

Teacher characteristic	Percent White	Percent not receiving subsidized lunch	Percent with parents who are college graduates	Mean prior year test score (Z)
Teacher experience				
0 to 1 year	0.99	0.97	0.94*	-0.050
2 to 5 years	1.01	1.00	1.00	0.004
6 or more years	1.00	1.00	1.00	0.009
Barron's College Rank:				
Less competitive	1.00	1.00	0.98	-0.052*
Competitive	0.99	0.99	1.00	0.017
Very competitive	1.04	1.00	1.07	0.052*
Not Ranked	0.97	0.87	1.08	-0.184
Licensure test score:				
Z-score < -1	0.98***	0.98	0.94*	-0.133***
-1 < Z-score < 1	1.01	1.01	1.00	0.023
Z-score > 1	1.01	1.00	1.08	0.075**
Nat'l Board Certification				
No	1.00	1.00	0.99	-0.006
Yes	1.06	1.11*	1.23**	0.182**
Advanced Degree				
No	1.00	1.00	1.00	0.004
Yes	0.99	0.98	1.00	-0.011

Note: For teachers with a given qualification, table entries in the first three columns are ratios of classroom characteristics to school-wide averages. Table entries in the last column are mean differences between classroom and school-average test scores. *** denotes a ratio or mean difference significantly different from one at the 1% level; ** the 5% level; * the 10% level. Source: **Teacher Sorting,** Table 3.

disproportionate shares of students from highly educated families and who scored well above average on prior year tests. In contrast, teachers with the weakest qualifications along a number of dimensions tend to teach the less advantaged and lower-achieving

students within each school. The table shows, for example, that the teachers with the lowest licensure test scores (those with test scores more than one standard deviation below the mean) teach in classrooms occupied by students whose parents are less likely college educated and students with prior year test scores below average for all fifth graders in the school.

The observation that, on average, schools appear to place the more highly qualified teachers in classrooms with more advantaged and higher achieving students need not mean that all schools behave that way. Indeed as part of our informal case studies of districts and schools, we visited one elementary school in which the principal told us that she randomly assigns teachers to classrooms. The process for assigning teachers and students to fifth grade classrooms in that school worked as follows. The teachers with knowledge of the rising fifth grade students (for example, 4^{th} grade teachers) are asked to develop lists of students to be placed in each of, say, three fifth grade classrooms, paying attention to racial balance and the known counterproductive interactions between certain individual students. The principal would then randomly assign each of the three fifth grade teachers to one of the three groups of students. The goal was specifically to forestall any efforts by parents to choose their child's teachers.

In order to determine how widespread such policies were, and also to develop a subsample of schools to be used in the subsequent stage of our analysis, we performed a large number of chi-squared tests to determine which, if any, schools appeared to randomly distribute students across classrooms and hence, teachers to classrooms. For this analysis, we focused on six measurable characteristics of students: gender, race, free and reduced price lunch status, prior year standardized test score, parental education (as reported by the student's prior year teacher), and whether the student attended the same school the prior year. Thus for 1205 schools having at least two fifth grade classrooms, we did more than 6000 chi-squared tests (1 test for each dimension for each of the 1205 schools) to test the null hypotheses that students were randomly distributed among classrooms within the school with respect to each dimension. To increase the power of the tests, we used up to three grades in each school and to lower the probability of false positives (that is accepting the hypothesis of random assignments when it is not true), we accepted a 10 percent risk of incorrectly rejecting the hypothesis of random assignment when it indeed true.

Based on this exercise, we identified more than 500 schools for which we were unable to reject the null hypothesis of random assignment along all six of the measurable dimensions tested. Thus, we concluded that in those schools, which represent about 45 percent of all schools in our sample, students appear to be randomly assigned to classrooms, and hence to teachers as well. We make use of this subsample of schools in some of our analysis described in the next section.

Teacher Qualifications and Student Achievement[6]

Though striking, the patterns that emerge in Tables 1-4 help explain gaps in achievement between various groups of students only to the extent that the relevant qualifications are causally related to student achievement. We now turn to an examination of that issue.

Our basic approach is to estimate a relatively standard education production function, but with particular attention to the processes of teacher sorting and shopping that interfere with the estimation of the causal effects of teachers on student achievement. The basic equation takes the form:

$$y_{ijt} = \delta y_{ijt-1} + \beta_1 X_{it} + \beta_2 X_{jt} + \varepsilon_{ijt}$$

where y is the ith student's fifth grade test score in school j and year t,

X_{it} is a vector of student characteristics,

X_{jt} is a vector of school characteristics, which includes the demographic characteristics and qualifications of individual teachers.

and ε_{ijt} is an error term.

Because of the processes of teacher sorting and shopping, it is quite likely that the explanatory variables will be correlated with the error term. Even after we have controlled for the measurable characteristics of students, there could still be reverse causation in higher achieving students may still be matched with teachers with stronger qualifications.

We use three strategies to address this potential statistical problem of reverse causation. First, in addition to a standard set of student demographic variables, we include an extended set of student variables based on survey responses collected at the time the students were tested. These variables include information on time spent on homework, use of computers, and time spent watching television. The inclusion of these variables is helpful in that it reduces the magnitude of the error term, thereby reducing the room for reverse causation. Second, we add school-level fixed effects, which we are able to do because of our ability to match students to teachers at the classroom level. The inclusion of fixed effects means that we are identifying the effects of teacher qualifications based only on the variation across teachers within each school. As a result, we are eliminating the statistical problems that emerge because of the sorting of teachers across schools that emerged so clearly in Table 3 above. Third, we address any remaining problems associated with the nonrandom assignment of students and teachers

[6] See our **Teacher Sorting** paper.

to classrooms within schools by restricting the analysis to the subsample of schools that appear to assign students randomly.[7]

The main substantive results, which are based on fifth grade students in 2000/02, are shown in Table 5. As indicated at the bottom of the table, all the reported regressions include an extended set of demographic control variables (including lagged student achievement) as well as school fixed effects. The first two equations (one for math and one for reading) differ from the final two only in terms of the sample on which they are based. The sample for the first two is all fifth grade students in schools that offer two or more fifth grades and for whom complete data were available. The far smaller sample for the final two regressions is similar except that it is limited to students in the smaller set of schools that met the criteria specified above to be treated as if they randomly assign students to classrooms within each school. Which set of regressions is preferred is a debatable question. While the first set could yield slightly biased results because they do not address the possibility of a nonrandom distribution of students across classrooms within each school, they have the advantage of being based on a larger sample. The second set has they advantage of ruling out any confounding effects related to teacher shopping across classrooms but suffers from being based on a slightly smaller, and possibly somewhat unrepresentative, sample. Fortunately, the results for both math and science are remarkably similar across the two samples.

For simplicity, we refer here mainly to the results based on the full sample in the first two columns. Emerging for both math and reading achievement is a strong and consistent effect of teacher experience. All the entries for teacher experience are statistically significantly different from the base case of no experience and rise almost monotonically reaching a peak at 20-27 years of experience. The estimates suggest that, controlling for other teacher characteristics, the presence of a highly experienced teacher increases student achievement in math by close to a tenth of a standard deviation relative to a novice teacher and by a bit less in reading. Moreover, in both cases almost half of the achievement effect is attributable to the first few years of experience. Thus we find strong support for our early decision to focus on novice teachers. Regardless of how effective they may eventually become, during their first year of teaching they are clearly less effective than more experienced teachers.

In addition, we find that a higher score on the state licensure test, all other factors held constant, also generates higher test scores, but only in math. There is also some weak evidence to suggest that being certified by the National Board has a small positive impact on reading scores. The most surprising result is the negative and statistically significant coefficient on having an advanced degrees relative to not having such a degree. One interpretation is that the possession of an advanced degree reduces a

[7] In our original proposal, we indicated we would try to address the statistical problem associated with the way that teachers sort themselves among schools by trying to find an appropriate exogenous instrumental variable. Our interviews with district and schools officials about how the teacher assignment process worked made it clear, however, that an appropriate instrument would be difficult, if not impossible, to find. Once we confirmed that we could match students to teachers at the classroom level, we developed the much cleaner method described in the text.

Table 5. Effects of teacher qualifications, with school fixed effects, full sample

	Full Sample		Apparent Random Sample	
Independent Variable	Math	Reading	Math	Reading
Black teacher	-0.016	-0.007	-0.008	0.005
	[0.011]	[0.010]	[0.018]	[0.016]
Hispanic teacher	0.026	0.052	-0.084	0.057
	[0.069]	[0.045]	[0.094]	[0.059]
Other race teacher	0.018	0.022	-0.054	0.042
	[0.034]	[0.030]	[0.057]	[0.042]
Male teacher	0.016	-0.023**	-0.006	-0.011
	[0.011]	[0.009]	[0.018]	[0.013]
Teacher experience (base=0 years)				
1-2 years experience	0.051***	0.035***	0.066***	0.017
	[0.014]	[0.013]	[0.020]	[0.017]
3-5 years experience	0.078***	0.046***	0.080***	0.035*
	[0.014]	[0.013]	[0.021]	[0.018]
6-12 years experience	0.076***	0.051***	0.085***	0.064***
	[0.014]	[0.013]	[0.020]	[0.018]
13-20 years experience	0.089***	0.065***	0.113***	0.073***
	[0.015]	[0.014]	[0.022]	[0.019]
20-27 years experience	0.096***	0.079***	0.101***	0.080***
	[0.014]	[0.013]	[0.021]	[0.018]
>27 years experience	0.090***	0.067***	0.130***	0.095***
	[0.016]	[0.014]	[0.023]	[0.020]
Quality of teacher's college (base=less competitive)				
competitive college	0.004	0.008	-0.013	0.006
	[0.008]	[0.007]	[0.012]	[0.010]
very competitive college	0.013	0.002	-0.005	0.009
	[0.012]	[0.011]	[0.020]	[0.014]
unranked college	0.000	0.011	-0.067*	0.027
	[0.027]	[0.032]	[0.039]	[0.041]
Teacher with advanced degree	-0.016**	-0.018***	-0.023**	-0.007
	[0.008]	[0.007]	[0.012]	[0.010]
Teacher Nat'l Board Certified	-0.004	0.030*	-0.035	0.005
	[0.018]	[0.016]	[0.028]	[0.023]
Teacher's licensure test score	0.012***	0.005	0.012*	0.002
	0.004]	[0.004]	[0.006]	[0.006]
Class size	0.002	0.001	0.006	0.002
	[0.002]	[0.002]	[0.004]	[0.003]
Student demographic controls	Yes	Yes	Yes	Yes
Extended set of student controls	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes
Lagged student achievement controls	Yes	Yes	Yes	Yes
Observations	60,656	60,502	24,768	24,711
R^2	0.756	0.707	0.766	0.708

Note: standard errors, in square brackets, have been corrected for within-classroom clustering. *, **, and *** denote significance at the 10%, 5% and 1% levels. Demographic controls include gender, race, and free/reduced price lunch status. Extended set of controls includes categorical measures of computer use, time spent free reading, time spent watching TV, parental education, and time spent on homework. Sample is restricted to the 521 elementary schools for which chi-square tests fail to reject the hypothesis of random assignment along six dimensions: race, gender, parent education, prior year test score, whether a student attended the same school in the previous year, and free/reduced price lunch receipt. Source: **Teacher Sorting**, from Tables 7 and 8.

teacher's effectiveness in the classroom. An alternative explanation, however, is the possibility that, controlling for all the other variables in the model, those who seek master's degrees are generally less effective than those who do not seek such degrees. Further investigation of this issue is needed with the use of longitudinal data. By itself, the negative coefficient would suggest that school districts might be wasting money by paying higher salaries to teachers with advanced degrees. That conclusion could change, however, if the higher compensation induced teachers to remain in the profession and thereby led to a more experienced teaching staff. More work on that issue is needed.

Among the teacher qualifications that are not significant are the selectivity of the teachers' colleges (as measured by Barron's ratings) and teachers' demographic characteristics. We note that, in regressions that exclude all student demographic controls and school fixed effects, it appears as if black teachers and male teachers reduce student achievement relative to white and female teachers. The more complete analysis reported here, however, indicates that those negative associations reflect the processes of teacher sorting and shopping rather than causal relationships between those characteristics and student achievement. Finally, we note that the small and insignificant finding for the class size variable at the bottom of the table should not be interpreted as the absence of a class size effect. Given that we have estimated the equations with school fixed effects, our methodology is far better suited to estimating the effects of teacher qualifications, which vary quite significantly across classrooms within schools. than it is for estimating the effects of class size. There simply is not sufficient variation in class sizes for fifth grades within schools to estimate the effect of class size on student achievement.

We undertook one final exercise as part of this project on assessing teacher effects. By interacting teacher qualifications with various student characteristics, we were able to determine whether a particular teacher characteristic, such as years of experience, generated greater gains in achievement for some types of students than for others. The results were surprising in that we found that more experienced teachers were even more effective in raising the achievement of advantaged students than of disadvantaged students, as defined by family income, parents' education level, and students' prior year test scores. This finding, which emerged only for math and not for reading, deserves further empirical investigation and verification, and suggests that we may want to revisit at least one aspect of the basic learning technology that we posited in our simple conceptual model of how a school administrator assigns teachers and students. The direction of the differential effect provides an additional explanation, namely the potential to increase average achievement, of why school administrators may match the more qualified teachers with the more advantaged and higher-performing students.

Accountability and the Distribution of Teachers [8]

Finally we examined whether North Carolina's relatively sophisticated and established school-based accountability program has exacerbated or alleviated the

[8] See our **Accountability** paper.

tendency for the higher quality teachers to end up in the schools with more advantaged and higher performing students. Our research on this topic did not have a specific racial focus. Instead we were interested in examining the hypothesis that the accountability program made it more difficult for schools with disproportionate shares of low-performing students to retain and attract high quality teachers. Since African American or other minority students tend to be overrepresented in those schools, any conclusions about those schools have direct implications for such students.

From a theoretical perspective that outcome appeared likely but not inevitable. On the one hand, unless a school-based accountability system is so carefully designed that it does not favor the schools serving the more advantaged students it will exacerbate existing incentives for high quality teachers to leave schools serving low-performing students in favor of those serving more advantaged students. By making such a move, teachers significantly increase the probability of earning a bonus, which in the North Carolina program is $1500 if the school is deemed exemplary and $750 if it meets its achievement expectations. Conversely, teachers in schools that are specifically labeled as "low-performing" under the program face increased scrutiny from the state and the humiliation of being part of a school that is publicly identified as failing. In fact, the North Carolina accountability program is quite well designed in that the teacher bonuses are given based on school-wide gains in achievement rather than on achievement levels (or, analogously, on percentages of students who are proficient). This focus on gains in test scores reduces the correlation between school performance and the socio-economic background of the students but does not eliminate it. As a result, even North Carolina's relatively sophisticated accountability program gives teachers additional incentives to move away from schools serving disadvantaged and low-performing students in favor of those serving more advantaged students.

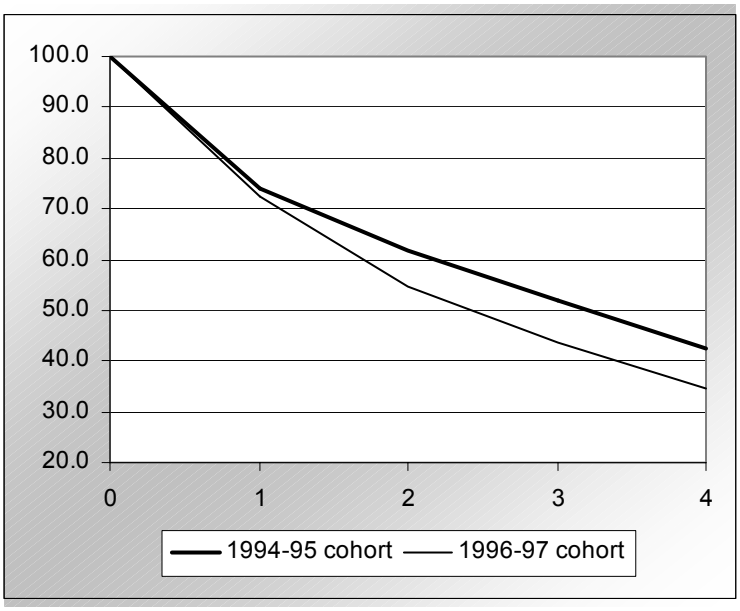
On the other hand, and working in the other direction is the possibility that the new emphasis on achievement for all students, including historically low-performing students, could exert pressure on school administrators to intervene more forcefully in the teacher-student matching process on behalf of the low-performing students. That might entail raising salaries in districts serving large numbers of such students, increasing salaries in struggling schools (an action that would most likely have to be implemented at the state level because of the tradition of uniform salary schedules within districts), or altering internal transfer policies to make it more difficult for high- quality teachers to leave low-performing schools. Thus, the effect of the accountability program on the ability of low-performing schools to retain and attract teachers is theoretically ambiguous and is an empirical question. Note that we are using the term low-performing school here as shorthand to signify a school with low-performing students. In fact such a school could be doing a good job in the sense of adding value to students who enter with very low levels of achievement.

Figure 2 depicts our initial descriptive analysis of the effects of the accountability program on the ability of low-performing elementary schools to retain teachers. We look at two cohorts of teachers – all the teachers teaching in low performing schools as of 1995 before the introduction of the accountability program (the 1995 cohort) and all the

teachers in low-performing schools in 1997, after the introduction of the accountability program (the 1997 cohort) . For the purposes of this figure, we have defined a low-performing school as one in which fewer than half the students are at grade level. Alternative definitions of low-performing, such as schools in the bottom 10 or 20 percent of the either the state or the district distribution, generate comparable results.

The horizontal axis represents the number of years from the initial period. The vertical axis indicates the percentage of the original cohort of teachers who remain in the school after the specified number of years. Emerging from the figure is the conclusion that the retention rate drops off more rapidly over time for the post-accountability cohort than for the pre-accountability cohort of teachers. Though not shown here, this pattern remains even when we depict the retention rates for each cohort relative to the aggregate retention rates for all teachers in each of the two years to account for any changes in labor market contexts over time.

Figure 2. Comparison of retention rates in low-performing schools, 1995 and 1997 cohorts (1995 cohort is pre-NC accountability program; 1997 cohort is post-accountability)

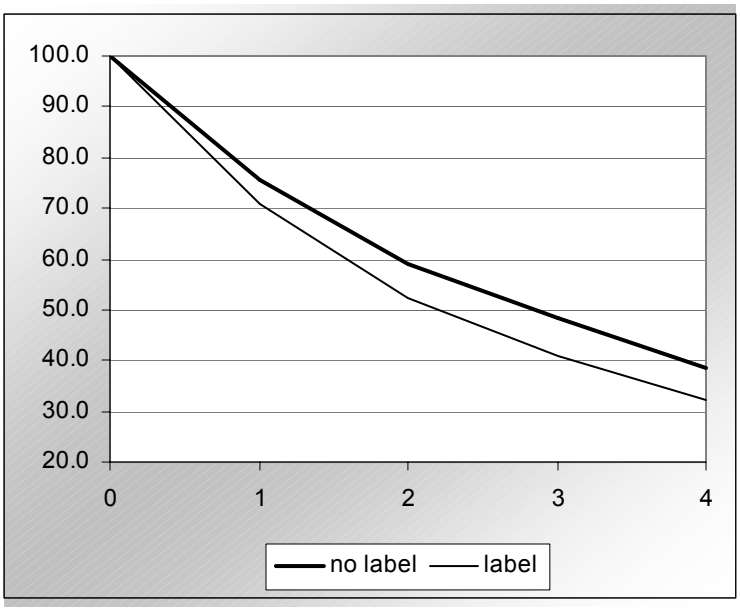


Note. Low- performing schools are defined here as schools with more than half of the students below grade level in the initial year. The horizontal axis refers to the number of years since the initial year for each cohort.
Source. **Accountability**, Figure 1.

In figure 3, we present a comparable analysis of retention rates but in this case the figure refers to teachers only in the 1997 cohort and the comparison is between retention rates in schools officially labeled as low-performing and other low-performing schools (as defined above). Here we see that the retention rates declined more rapidly in the labeled than the non-labeled low-performing schools, suggesting that the labeling has an

adverse effect on the ability of schools to retain teachers. Once again, there could be alternative explanations for the results, but more complex analyses yield the same basic conclusion.

Figure 3. Comparison of retention rates in low-performing schools, labeled versus non-labeled schools, 1997 cohort.



Note. Retention rate is the percentage of teachers who remained in their original schools after the indicated number of years. Only teachers from low-performing schools are included, where a low performing school is one with over half of its students below grade level in math or reading scores in the initial year. The horizontal axis refers to the number of years since the 1996-97 year.

Source: **Accountability,** Figure 2.

Shifting the analysis from the cohort level to the level of the individual teacher, we used a duration model to estimate exit rates for individual teachers. In effect we have estimated the probability that in any specific year a teacher will leave a school, conditional on her remaining in the school until that point. Of particular interest are interaction terms between low-performing schools and the post-accountability time period. A positive sign on the interaction term means that the accountability program increased the probability that a teacher would leave the school and provides evidence that the accountability system has exerted an adverse effect on the ability of low-performing schools to retain teachers.

In Table 6, we summarize the results of the model for five different definitions of low-performing schools, as indicated in the footnote of the table. For this table we have translated the estimated coefficients into probabilities that a typical teacher with either ten years or one year of experience would leave the school in which she was teaching.. Focusing only on the first column, we find that a typical teacher with 10 years of experience who is teaching is a low-performing school prior to the accountability system

had a 0.176 probability of leaving that school during the year (see the third row). That probability rises to 0.191 in the post-accountability period and to 0.209 if the school is labeled as a low-performing schools Similarly a typical new teacher (see the bottom panel) in a low performing school had a 0.338 probability of departure before accountability. That probability then rises to 0.383 in the post accountability period and to 0.403 if the school is officially labeled as low performing. Thus, the accountability program apparently increased the departure rates of teachers from low-performing schools by about 20 percent. We conclude with some confidence, therefore, that North Carolina's accountability system made it more difficult for low-performing schools to retain teachers.

Table 6. Estimated probabilities of departure from a school, typical teachers with 10 years of experience and 1 year of experience.

	Models that differ by the definition of a low- performing school				
	(1)	(2)	(3)	(4)	(5)
10 years					
Typical individual	0.150	0.152	0.150	0.150	0.151
+ Post	0.155	0.153	0.151	0.154	0.152
+ low performing	0.176	0.158	0.160	0.173	0.161
+ Low × Post	0.191	0.171	0.167	0.170	0.169
+ Label × Post	0.209	0.206	0.201	0.193	0.205
1 year (new teacher)					
Typical individual	0.320	0.322	0.320	0.318	0.316
+ Post	0.328	0.323	0.321	0.324	0.319
+ low performing	0.338	0.313	0.317	0.338	0.332
+ Low × Post	0.389	0.365	0.359	0.377	0.370
+ Label × Post	0.403	0.399	0.395	0.395	0.397

Note. Calculated by authors based on the coefficients of the model in Table 2. Models 1-5 differ from each other only in how a low-performing school is defined. The definitions are: (1) a school in which more than half of its students are below grade level on math or reading test scores; (2) a school that is in the bottom 10 percent of the district distribution of schools ranked by percentage of students at grade level in math or reading test scores; (3) same as (2) but in the bottom 20 percent of the district distribution (4) a school that is in the bottom 10 percent of the statewide distribution of school ranked by percentage of students at grade level in math or reading test scores; and (5) same as (4) but in the bottom 20 percent of the state-wide distribution.

Source. **Accountability,** Table 3

These lower retention rates, and hence higher rates of teacher turnover, are undesirable because they increase the challenges that low-performing schools face in providing a stable and productive educational environment. Another potential adverse outcome of the lower retention rates is that, because they have more vacancies to fill, the low-performing schools will end up relying to an even greater extent on novice, and hence less-effective, teachers than they would have in the absence of the accountability system.

Somewhat to our surprise, however, we found no statistically significant evidence that the state's accountability system increased the percentage of novice teachers (or of teachers from unselective colleges) in low-performing schools. Comparisons of the shares of novice teachers in the low-performing schools before and after the introduction of the accountability program provided no support whatsoever for the hypothesis of a decline in the quality of their teachers. Shifting the focus to changes in changes over time rather than changes in levels, we found patterns consistent with the hypothesis that the accountability adversely affected teacher quality in the low- performing schools but none of the differences-in-differences were statistically significant. These results appear to be attributable, at least in part, to a number of other policy changes related to the labor market for teachers in North Carolina during the relevant period. These policies include the increasing of teacher salaries and an increased reliance on lateral entry teachers and teachers from other states rather than on new graduates from the state's colleges of education.

Our finding that North Carolina's relatively sophisticated accountability system exacerbated the difficulties that low-performing schools face in retaining teachers highlights the importance of taking a systemic view in designing accountability programs. Though putting additional pressure on the teachers in low-performing schools to increase student achievement may have some desirable effects, those positive effects could well be offset through the negative effects that operate through the labor market for teachers. Thus, if the policy goal is to reduce achievement gaps, policy makers need to pay attention to unintended side effects of this type.

Conclusion

The research summarized here, which represents three years of work with newly available administrative data on North Carolina students, teachers and schools, sheds new light on a number of policy issues related to minority gaps, not only in North Carolina but also for other states. First it highlights the need for policy makers to pay attention to the resegregation of schools, not only as traditionally measured as the segregation between schools but also as the segregation that occurs across classrooms within schools. Second, it highlights the need for policies designed to distribute novice teachers more evenly across schools and classrooms so as to minimize the disadvantage currently experienced by minority students relative to white students. Third it documents one unintended side effect of accountability systems, one that works to the disadvantage of the very students that accountability systems are intended to help.

Although all our research to date is policy relevant, more specific work is needed to determine the power of specific policy levers to address the issues we have identified. For example, to what extent would the offer of higher salaries for teachers to work and remain in struggling schools reduce the number of novice teachers in those schools or what is the likely effect on minority achievement gaps of alternative licensure programs? In addition, while our research to date has shed light on a number of important issues, it has also raised a number of new questions. For example, do the greater observed returns to years of experience reflect the greater effectiveness of experienced teachers or, alternatively, the departure from the profession of the less effective teachers. Similarly, the negative effect on student achievement that emerges for teachers with advanced degrees cries out for more analysis. The NC administrative data set is sufficiently rich to examine these and other issues, particularly by making greater use of the longitudinal aspects of the database. Thus, we still have much work to do.

References

1.Clotfelter, C.T., H.F. Ladd, and J.L. Vigdor. 2003. "Segregation and Resegregation in North Carolina's Public School Classrooms." *North Carolina Law Review*. Vol. 81, no. 4, May. **Segregation.**

2. Clotfelter, C.T., H. F. Ladd, and J.L. Vigdor. Forthcoming. "Classroom-Level Segregation and Resegregation in North Carolina" In C. J. Boger and G. Orfield (eds). *Resegregation of the American South* (University of North Carolina Press). **Segregation (short). T**his chapter is a shortened and more accessible version of the previous paper.

3. Clotfelter, C.T., H.F. Ladd, J.L. Vigdor, and R. Aliaga Diaz. 2004. "Do School Accountability Systems Make it More Difficult for Low-Performing Schools to Attract and Retain High-quality Teachers? *Journal of Policy Analysis and Management.* Vol. 23, no. 2, pp. 251-271. **Accountability**

4. Clotfelter, C.T., H.F. Ladd, and J. L. Vigdor. Forthcoming. "Who Teaches Whom? Race and the Distribution of Novice Teachers." *Economics of Education Review*. **Who Teaches Whom?**

5. Clotfelter, C.T., H.F. Ladd, and J.L. Vigdor. 2004. "Teacher Sorting, Teacher Shopping and the Assessment of Teacher Effectiveness." Under review for publication. **Teacher Sorting.**