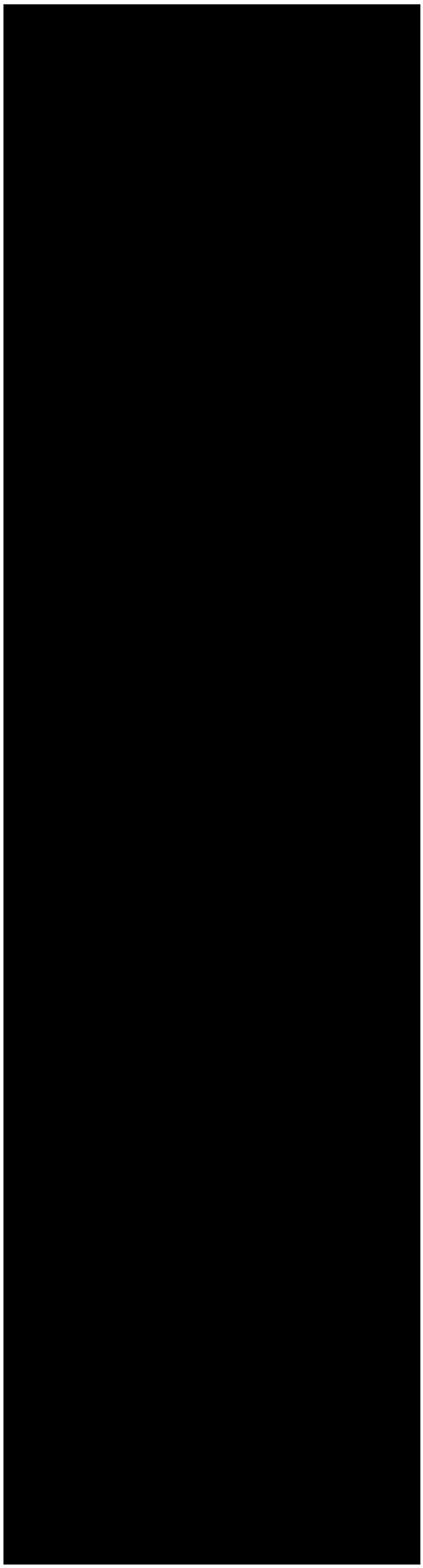

School Policies and the Black-White Test Score Gap

Helen F. Ladd

Working Papers Series
SAN08-03

March 24, 2008

TERRY
SANFORD INSTITUTE
OF PUBLIC POLICY
D U K E

School Policies and the Black-White Test Score Gap

Helen F. Ladd

Abstract

This paper examines school-related policies and strategies that have been proposed or justified, at least in part, on the basis of their potential for reducing black-white test score gaps. These include strategies, one of which is greater integration, to reduce differences in the quality of teachers faced by black and white students; school and classroom policies designed to improve the achievement of low-performing students; and the strategies of school accountability and parental choice designed to change incentives throughout the education system. While none of these strategies is likely to be sufficiently powerful to offset the powerful non-school social forces that contribute to the racial achievement gap, the failure of education policy makers to be vigilant about the aspects of the problem over which they do have some control could well lead to even greater gaps in the future or to lost opportunities to reduce them.

Keywords: education, achievement gap, integration, school accountability

Forthcoming in Katherine Magnuson and Jane Waldfogel, eds. *Steady Gains and Stalled Progress: Inequality and the Black-White Test Score Gap.* Russell Sage. The author thanks Charles Clotfelter, Jens Ludwig, Jacob Vigdor and the editors of this volume for their helpful suggestions.

Contact: Helen F. Ladd
Professor of Public Policy and Economics
Terry Sanford Institute of Public Policy
Duke University
Durham, NC 27708-0245
helen.ladd@duke.edu

I. Introduction

Black students in the U.S. achieve on average at lower levels than do white students. Recent evidence from the National Assessment of Educational Progress (NAEP) indicates, for example, that the gap between 13- year-old black and white students was about 0.6 standard deviations in reading and about 0.8 standard deviations in math as of 2004. To be sure, such gaps were far larger in the 1970s when they exceeded a full standard deviation in both subjects. The gaps fell quite dramatically during the 1970s and 1980s, increased during the early 1990s and then fell again in the latest five year period. These ups and downs, notwithstanding, the persistence of these gaps is cause for significant policy concern for reasons discussed elsewhere in this book and in Jencks and Phillips, 1998.

The chapters by Vigdor and Ludwig and by Corcoran and Evans in this volume have drawn attention to school-related trends such as in the racial segregation of the schools and the widening disparities in teacher qualifications between black and white students, especially at the elementary level, that may have stalled the convergence of the black and white test scores in the 1990s. This chapter picks up from their analysis and asks what educational policies might be pursued moving forward to help reduce the black-white test score gap, or at least to offset some of the other trends that may tend to widen it, such as rising income and social inequality. Of particular interest for this review are school policies and strategies that have been proposed or justified -- at least in part -- on the basis of their potential for reducing black-white test score gaps. As will

become apparent, not all the proposed strategies are likely to be effective in that regard and their net effect on the size of the gap is likely to be relatively small.

The following discussion is divided into five sets of policy strategies. The first two focus on teachers, but from quite different perspectives. One perspective relates to the assignment of students to schools, with attention to how racial segregation of students affects the quality of teachers for black students relative to white students. The other perspective focuses on more direct interventions designed to improve the quality of the teachers of black students. The third set includes the non-teacher strategies of reducing class size and implementing whole school reform. The fourth and fifth sets of strategies emerge from a more systemic view of the educational challenge and are designed to change the incentives throughout the education system. Included here are both top-down accountability strategies designed to hold schools accountable for the performance of their students and bottom up strategies such as increased parental choice and competition designed either to improve schooling options for certain groups of students or to make use of market type pressures to improve educational outcomes.

The main thrust of this chapter is that while none of the strategies discussed here is likely to be sufficiently powerful to offset the powerful non-school social forces that contribute to the racial achievement gap, school related strategies are a necessary component of any overall effort to reduce such gaps. Moreover, the failure of education policy makers to be vigilant about the aspects of the problem over which they do have some control could well lead to even greater gaps in the future or to lost opportunities to reduce them.

II. Student assignment policies

Vigdor and Ludwig (this volume) document that progress in reducing the black-white test gap stalled at about the same time that efforts to desegregate schools, as measured by the trend in the segregation of schools relative to the segregation of neighborhoods, slowed down. The authors conclude that if school desegregation had proceeded at the same rate as neighborhood desegregation, the black white gap might have narrowed somewhat, but only slightly because of the relatively small change in the racial composition of neighborhoods during the relevant period and the small effect sizes that emerge from the literature they review. At the same time, the authors emphasize that any retreat from the goal of racially integrated schools would exacerbate black-white test score differences in the future.

Racial integration could reduce the black-white achievement gap through two main mechanisms. The first is through the potential for positive spillover effects from one group of students to another. That is the mechanism emphasized in the "peer effects" literature described by Vigdor and Ludwig (this volume).[1] The second mechanism works through the teacher labor market. As discussed by Evans and Corcoran (this volume) , the evidence is increasingly compelling that certain credentials are predictive of student achievement and that teachers with the weaker credentials are more likely to teach the more advantaged students. More specifically, teachers tend to sort themselves among schools in ways that work to the disadvantage of students in schools disproportionately

[1] Having more advantaged peers need not always lead to more positive outcomes. As emphasized by Jencks and Mayer (1990), another possibility is that children in schools with more advantaged peers may become discouraged by their relative deprivation. In that case, having more advantaged peers could reduce student achievement.

serving minority students, including black students. The following discussion focuses on this second mechanism.[2]

Racial segregation and teacher disparities by race

Assuming that access to teachers is measured at the school level, racial balancing of students across schools would assure that students of each race would have access to similar teachers on average. To be sure, racially balanced schools need not mean that all classrooms within schools are racially integrated. Nonetheless the more racially integrated are the schools, the more likely it is that students of different races will have teachers with similar qualifications.[3]

New evidence from North Carolina documents not only that teachers in high-minority schools have weaker qualifications on average than those in schools serving white students – an observation that emerges in many states -- but also that the black-white differences in teacher qualification have been growing over time as minority students have become more concentrated in high minority schools. Table 1 provides information on the racial composition of students in two groups of North Carolina schools, those in the quartile with the highest percentages of minority students (Quartile I) and those in the quartile with the lowest percentages (Quartile IV), separately by level of school and by year.[4] The table indicates that that the high minority schools at each level of schooling are becoming more racially concentrated over time, and in the process are becoming increasingly different from the low minority schools. The patterns and

[2] As noted by Vigdor and Ludwig, the methodology used in the peer effects literature typically does not incorporate the effects of this second mechanism. In particular, the inclusion of school fixed effects in the regression models holds constant the time-invariant characteristics of schools, including the mix of their teachers.

[3] Clotfelter, Ladd and Vigdor (2003 and 2008) document for North Carolina that at the elementary level, most of the racial segregation is between, not within, schools. At the high school level, within-school segregation plays a far larger role. .

[4] The quartiles are redefined for each year.

trends in this table provide the context for Table 2 which, based on the same groupings of schools, reports the average percentages of teachers with fewer than three years of experience. The research literature indicates that inexperience has a clear adverse causal impact on student learning (Clotfelter, Ladd and Vigdor, 2006 and 2007; Goldhaber summary, 2008). Other characteristics that are also predictive of student achievement, such as teacher test scores, exhibit similar patterns but are not shown.

For each year and each level of schooling, it is clear that the students (most of whom are black in North Carolina) in the high minority schools are more likely to have an inexperienced teacher than those in low minority schools. In addition, however, the differences in percentages between the Quartile I and IV schools have been rising over time. Thus at the same time that minority students are increasingly concentrated in the Quartile I schools, the proportions of inexperienced teachers in those schools has been rising both absolutely and relative to those in the low-minority schools.

Additional and more precise evidence of the link between changes in racial segregation and teacher credentials emerges from Table 3, which highlights changes in the credentials of teachers faced by the typical black and typical white student in Charlotte-Mecklenburg between 2000/01 and 2005/06. This district is of interest because of the precipitous shift in its student assignment policy in 2002 as it moved away from court-induced efforts to maintain racially balanced schools to a choice-based neighborhood approach that greatly increased racial segregation. As documented by Clotflelter, Ladd, and Vigdor, 2008, the percent of nonwhite students in that district enrolled in schools with 90-100 percent nonwhite students increased from 6.9 percent in 2000/01 to 38.5 percent in 2005/06, which far exceeds the increase in any other large NC

district. [5] Table 3 includes information on three credentials of teachers at the school level, all of which have been shown to be predictive of student achievement : the percent of teachers (1) with three or more years of experience, (2) who scored in the top quartile on standardized teacher tests and (3) who were fully certified as teachers (Clotfelter, Ladd and Vigdor, 2007a and b). The entries in the table are the weighted averages of the percentages of a school's teachers in each category where the weights are, successively, the number of white and black students in each school. Each of the credentials is defined in a positive way, so that higher proportions indicate teachers with stronger qualifications.

Consistent with the evidence for all schools in North Carolina the patterns for 2000/01 favored white students, although not so dramatically as in some other North Carolina urban school districts at that time.[6] Of particular interest here is how those disparities increased in the wake of the district's increase in segregation. For exposure to experienced teachers, the disparity between black and white students rose from 2.9 to 4.2 percentage points; for high-scoring teachers it rose from 8.4 to 8.6 percentage points; and for certified teachers it rose from 2.2 to 3.8 percentage points. This example provides the clearest evidence to date that increases in racial segregation are likely to bring with them greater black-white disparities in teacher credentials.

Policy strategy – balancing schools by SES

Though the recent rise in segregation in Charlotte-Mecklenburg exceeds that in other North Carolina districts and possibly in most other districts as well, it may well be

[5] A more nuanced measure of segregation indicates a similar increase. That measure increase from 0.20 to 0.33 over the same period, again a huge increase relative to that of other districts.

[6] Data not shown. The one exception to the statement in the text is Wake County which explicitly promoted racial balance across schools prior to 1999 and since then has promoted economic balance. See discussion below.

suggestive of future trends because of the recent backtracking of the federal court on the issue of school desegregation. In the early 1990s, the courts ruled that school districts declared "unitary" have no obligation to offset de facto segregation in schools resulting from residential segregation.[7] Subsequently, in a series of decisions meant to apply to districts not under court order, the Fourth Circuit Court of Appeals ruled that race could not be used in assigning students to schools.[8] Most recently, in Parents Involved in the Community Schools v. Seattle School District No. 1, the Supreme Court made a similar ruling by declaring unconstitutional school assignment plans that were based on the race of individual students. Hence, any efforts to promote racially integrated schools at the district level from now on will have to be done indirectly. Among the policies that may pass court muster are the selective siting of schools and rezoning of school catchment areas.

One of the most commonly advanced indirect strategies is to integrate schools by the socio economic (SES) characteristics of their students (Kahlenberg, 2001 and Century Foundation Task force, 2002). The SES based school assignment strategy pursued by Wake County, North Carolina, exemplifies this approach.[9] Designed to ensure that all schools are middle class schools, the district limits the percentage of low income students in each school to 40 percent and the percentage of students scoring below grade level to 25 percent. Some supporters of SES balancing would prefer such a strategy to a race-based strategy in any case. Because black students within a district are likely to be

[7] Board of Education of Oklahoma (1991) and Freeman v. Pitts (1972).
[8] *Capacchione v. Charlotte-Mecklenberg Schools*, 57 F. Supp. 2d 228 (W .D.N.C. 1999); *Eisenberg v. Montgomery County Public Schools*, 197 F.3d 123 (4th Cir. 1999); *Tuttle v. Arlington County School Board*, 195 F.3d 698 (4th Cir. 1999). For an analysis of these decisions, see Boger (2000).
[9] Other districts that have implemented socioeconomic integration plans include LaCrosse, Wisconsin, Cambridge, Massachusetts, and San Francisco. See Reardon, Yin and Kurlaender (2006) for descriptions of the plans in these three districts.

overrepresented among students from low income families, however, the SES strategy has also been justified in part on its potential to reduce racial segregation and in some areas has succeeded (Kahlenberg, 2001, Chaplin 2002). In 20005/06, for example, Wake County had only 2.3 percent of its nonwhite students enrolled in 90-100 nonwhite schools, far below the 38.5 percent already noted for Charlotte and also well below that for other big districts in North Carolina.[10]

A 2002 study based on national data on the distributions of black and white students provides additional support for this race-based rationale for SES balancing (Chaplin, 2002). More recent research by Reardon, Tun, and Kurlaender (2006), however, highlights the limitations of an income-based balancing strategy for reducing racial disparities across schools. At one extreme, if all black students were poor and no white students were poor, distributing poor students equally among schools would be tantamount to distributing black students evenly among schools. The authors show, however, that within large urban areas in the U.S. the income distributions of blacks and whites are not sufficiently different in practice to guarantee much racial integration even with a strictly defined income integration scheme. In practice, the effects on racial integration depend on the disparity between the incomes of whites and blacks in the area and the details of the income-based integration plan. Further, given the observed patterns of residential segregation by race and income within U.S. urban districts, for income balancing to lead to racial balancing of schools most districts would have to make transportation readily available to all students (so that they can attend schools outside their residentially segregated neighborhoods) and also to invest resources in particular

[10] The percentages for other large districts are 30.9 for Guilford, 9.4 for Cumberland and 23.9 for Winston/Salem /Forsyth (Clotfelter, Ladd and Vigdor, 2008, Table 2).

schools to counter the preferences of some parents to enroll their children in schools near their homes(Readon et al, 2006, p. 68).

Though the authors correctly emphasize that balancing schools by income might be desirable for reasons other than their effects on racial integration, their main conclusion is that integrating schools by income is at best a poor substitute for integrating schools by race. From the perspective of the black-white achievement gap, that is disappointing because, as documented by Corcoran and Evans in this volume, sorting decisions of teachers appear to be influenced more by the race of a schools' students than by their SES. Nonetheless, given the limits imposed by the courts on the power of districts to use the race of individual students in making school assignments, some districts may find that a carefully designed strategy for balancing schools by SES is the best tool available to them for promoting racial balance, and thereby indirectly leveling the distribution of teachers across students of different races.[11]

At the same time, given that schools are likely to remain racially segregated and may well become even more so in the future, other more direct strategies will also be needed to counter the disadvantage black students face relative to white students in the quality of their teachers. I now turn to some of those strategies.

III. Teacher Quality – Direct Policy Interventions

Among the direct policy interventions for reducing the black-white disparities in teacher quality are financial incentives intended to make schools serving minority students more attractive to teachers and the development of new pathways into teaching

[11] See section on parental choice for a discussion of the argument that parental choice might conceivably serve the same goal.

designed to provide more teachers for hard-to-staff schools. In addition, attention to professional development is potentially important for the black-white gap because of the weak credentials of many of the teachers in high minority schools

Financial incentives to alter the distribution of teachers

The fact that on average the teachers of black students have weaker credentials than those of white students reflects the way teachers are distributed both across and within school districts. At both levels, policy interventions related to teacher salaries could potentially be part of a productive policy strategy.

Across districts. The distribution of teachers across districts largely reflects considerations of supply and demand, including the preferences of teachers. Various authors have investigated the effects of various factors on the ability of districts to attract and retain teachers and have found that teacher retention tends to be higher in districts with better salaries, higher pupil test scores, smaller classes, and lower proportions of low-income and minority students (see, for example, Murnane and Olsen 1989, Mont and Rees 1996, Hanushek, Kain, and Rivkin 1999, and Scafidi, Sjoquist, and Stinebrickner 2002).

If money were no object, the offer of high teacher salaries would be a logical component of any policy strategy for attracting and retaining more highly qualified teachers to districts serving large proportions of minority students. Such salaries would have to be sufficiently high, however, to compensate teachers who would otherwise prefer to teach in districts with more congenial working conditions. Some studies suggest that such salary differentials would need to be quite high. Hanushek, Kain and Rivkin (2004) estimate that reducing the rate of attrition in a large urban district to that in

suburban schools would require a 43 percent salary difference for female non-minority teachers with three-to-five years of experience. Emerging from similar research for New York State is that salary differentials of $10,000 to $16,000 would be required to attract equally qualified teachers to low-performing public schools away from suburban schools (Boyd, Lankford, Loeb and Wyckoff, 2006).[12] The rub is that districts with large proportions of minority students are often unable to raise sufficient local tax revenue or they receive insufficient aid from their states to offer the higher salaries needed to attract high quality teachers.

Within districts. The distribution of teachers within districts introduces some additional considerations in part because district and school officials play a major role not only in assigning teachers to schools and to classrooms but also, as discussed in the previous section, in assigning students. Another key difference is that while salary schedules differ across districts they are uniform within districts. Under the current system, a teacher at any step in the salary schedule would receive the same salary regardless of the school at which she teaches within the district. As a result, the easiest way a district administrator can improve the real income, or job satisfaction, of an experienced teacher who remains within a district is to permit her to move to a school offering a more satisfying teaching experience. Such transfers generally work to the disadvantage of black students, particularly those from low income families, when they are concentrated in specific schools.

[12] Such estimates would overstate the required salary increases if the estimated wage elasticities of supply are too low. A study based on an $1800 bonus program for eligible teachers in low-performing middle and high schools in North Carolina finds retention elasticities that are significantly larger than other estimates in the literature.(Clotfelter et al, forthcoming). These larger estimates may emerge from the bonus program because the researchers are better able to separate the effects of the salary differential from the working conditions in the school given that the bonus applies to only a subset of the teachers within each school.

Once again financial incentives may be required to change the incentives for teachers to move among schools in this way. An example of this approach is North Carolina's $1800 annual bonus program for certified teachers in the shortage areas of math, science, and special education teaching in eligible middle and high schools. School eligibility was determined based on the percentages of low-income students and of student performing poorly in math and biology. Importantly the program was designed so that eligible teachers would continue to receive the bonus even if the school became ineligible as a result of its improved performance. Despite flaws in the way the program was implemented, the evaluators found that the program reduced turnover in the eligible schools by 17 percent (Clotfelter, Glennie, Ladd, and Vigdor, forthcoming). The program was not in place long enough for it to have any measurable impact on teacher recruitment.

Policy implications. Financial incentives are only one of several policy options for attracting quality teachers to high-minority schools. Another is for urban districts to hire teachers earlier in the year to avoid much of the late hiring that historically has put them at a disadvantage relative to suburban districts in competing for quality teachers (Jacob, 2007). Yet another is to improve the working conditions for teachers in high-minority schools, possibly by improving the leadership of those schools. At this point, the research is not sufficient to determine which strategy or combination of strategies is likely to be most effective. Ideally, states and districts would experiment with various forms of financial and other incentives, and researchers would be actively engaged in their evaluation.

Alternative pathways into the profession

Consistent with the theme of previous sections, schools serving low-performing minority students are likely to have the greatest difficulties attracting teachers of any quality to teach their students. As a result, their teachers are more likely to be uncertified, to be on some form of temporary license or to be a novice teacher than those of other schools. Some cities, including most notably New York City, have set up programs, including the NYC Teaching Fellows Program, to address this challenge by providing new pathways for potential teachers to enter the profession. Though these new pathways require far less initial training than the traditional pathway of standard teacher training and certification, the goal is to attract teachers who are sufficiently able to offset their initial lack of training. The best known national program of this form, Teach for America, recruits corps members from the top universities, provides intensive training during the summer and additional support as they pursue their teaching assignments in hard-to-teach schools in communities throughout the country. A potential difference between the TFA program and the new New York City-specific pathways is that there is no expectation that teachers will remain after their second year of required teaching.

A major policy question is how the teachers entering through these alternative pathways fare in the classroom. Studies of TFA teachers in Houston generated somewhat mixed results, with the conclusions differing in part on whether TFA teachers were compared only to certified teachers or to all teachers, regardless of their certification status (see summary in Goldhaber, 2008, p. 151). A recent randomized national field experiment of the TFA program presents a clearer picture. Regardless of the comparison group, the students with TFA teachers outperformed the students of other teachers in the relevant schools in math by as much as 0.15 standard deviations, but performed no better

or worse in reading (Glazerman, Mayer and Decker, 2006) . Positive findings for the TFA teachers also emerge from a careful study of the alternative entry paths in New York City. At the same time, the students of teachers who entered through one of the New York City specific programs exhibited slightly smaller gains than the students of other teachers (Boyd et al. 2005). Consistent with that finding, a more general study of the effects of teacher credentials in North Carolina also shows lower achievement gains for students of teachers who have licenses under that state's alternative entry program (Clotfelter, Ladd and Vigdor, 2007a and b).

The positive findings for TFA teachers notwithstanding, the jury is still out of the power of alternative entry programs to raise the quality of teaching in low-performing schools serving large minority populations. One characteristic of the TFA program that stands out, and is worthy of further attention, is the greater support it provides for its teachers once they are in the classroom than is the case for most other alternative entry programs.

Professional development

Despite the particular importance of professional development for many of the teachers in low-performing schools, the evidence on how best to proceed is scanty.[13] At the same time, the evidence is increasingly clear about what districts should avoid. In that category are financial incentives for teachers to complete master's degrees that are not tightly linked to their teaching responsibilities and investments in short term, generic professional development activities. Instead, professional development should be longer and deeper, and should be linked to the relevant standards, curriculum, and assessment system of the district or state. Even professional development programs that meet those

[13] This section draws heavily on Hill, 2007.

general criteria, however, may not be effective. Hence, the challenge for educational policy makers is to find programs that are demonstrably effective and are tightly aligned with the needs of the teachers and the goals of the district.

II. Policies directed toward classrooms and schools

The strategies in this section shift attention away from the quality of teachers to the size of classrooms and to school-based comprehensive reform efforts. As will become clear, though, concerns about teacher quality cannot be avoided, especially with respect to the class size discussion.

Smaller class sizes

Reducing the size of classes has long been on the policy agenda of state policy makers. As noted by Corcoran and Evans (this volume, Figure 1), average class sizes, as approximated by pupil-teacher ratios, declined from about 19.3 in the mid 1980s to 17 in 2004. For smaller class size to serve as a strategy to reduce the black white test score gap smaller class sizes would have to generate higher achievement for minority students than for white students.

Evidence on how class size affects student achievement emerges from two main sources: empirical studies of observational data and a well-known randomized field trial, called the Tennessee STAR (Student/Teacher Achievement Ratio) project. Though his periodic reviews of the various observational studies have led the economist Eric Hanushek (e.g. Hanushek, 1997) to conclude that smaller class sizes have no systematic positive effect on student achievement, his methodology and conclusions are subject to

significant criticism.[14] The STAR study, in contrast, provides compelling evidence not only that smaller class sizes generate higher achievement in the early grades but also that the effects are larger for minority students.

The STAR project, which was financed by the Tennessee Legislature and ran for four years in the mid 1980s, is highly touted because it was based on an experiment in which students were randomly assigned to classrooms of different sizes. Kindergarten, first, second and third classrooms of 13-17 students were compared to classrooms of 21-25 students. The curriculum and the tests were standardized to compare about 6,500 pupils in about 330 classrooms, at approximately 80 schools in math, reading and basic study skills. The initial study concluded that smaller classes generated gains in achievement scores, especially in kindergarten and grade one and for minority children (Finn and Achilles, 1990; summary by the Harvard Statistician Frederick Mosteller, 1995). Moreover the effects on minority students in the inner city were larger for minority students than for white students. These findings emerged not only from the original study, but also from careful follow-up studies by Alan Krueger (1999) and Nye, Hedges and Konstantopoulos (2000) in which the authors explicitly addressed some of the flaws in the implementation of the STAR experiment. The largest increase in test scores emerges for students the first year they attend a small class. After that year,

[14] Hanushek's approach and conclusions have been criticized on methodological ground by Hedges, Laine and Greenwald (1994) and Alan Krueger (2002) Particularly compelling is Alan Krueger's criticism that Hanushek's method of aggregating results across studies gives far too much weight to multiple estimates from studies that find no effects. With a more appropriate weighting based on the same set of studies reviewed by Hanushek, Krueger (2002) concludes that, after other factors that affect student achievement are appropriately controlled for, student achievement is higher in the smaller classes.

additional time spent in a small class has a positive, but weaker association with test scores.[15]

Subsequent studies using follow-up data indicate that the positive achievement effects of small class sizes in the early grades appear to persist through eighth grade (Nye, Hedges and Konstantopoulos 2004). Moreover, consistent with the initial studies of short term benefits, the benefits as of eighth grade were larger for minority students than for white students, although the differential between white and minority students was statistically significant only for reading (Nye, Hedges and Konstandopoulos, 2004, p. 99). Across the five years of the follow-up, minorities benefited from the early class size reductions on average in reading by an amount that was about 67 percent larger than the benefit to white students.

Most researchers now agree that small classes can be beneficial in the early grades, and particularly for minority students.[16] For policy purposes, however, three caveats are worth noting. The first is that smaller class sizes do not guarantee higher student achievement. This point clearly emerges from Murnane and Levy's study of 15 schools in Austin, Texas (1996). As a result of a desegregation court order, all fifteen schools were given additional funding to reduce class sizes. Though all the schools hired more teachers and reduced class sizes, achievement rose only in the two schools that made other changes as well, such as adopting new curriculum, bringing in health services and involving parents. The second caveat is that reducing class size is expensive since it

[15] Angrist and Lavy (1999) also find that smaller class sizes increased achievement in Israel using the natural, but quite random, variation in class sizes associated with that country's explicit policy of capping class sizes.

[16] One possible exception is Eric Hanushek (1999) who emphasizes that the benefits of smaller classes are limited to kindergarten and first grade and that the huge variation in student achievement suggests that teacher quality is much more important that class size.

requires additional teachers and classrooms. Hence, positive effects on student achievement along do not make it a cost-effective strategy.

The third caveat is that policy makers must be careful in extrapolating the results from an experiment such as Project STAR to a district or statewide policy to reduce class size. The reason is that large scale changes set in motion a variety of other adjustments that are not incorporated into the small-scale experiment. Most obvious in the case of a class size reduction is that it creates a need for many additional teachers and classrooms and is likely to induce some teachers to move from one school or district to another. As a result, when California enacted legislation in 1996 to reduce K-3 classes by about 10 students per class, the hoped for differential benefits for minority students did not materialize. The problem was that the larger teaching force required to staff the smaller classrooms led to a deterioration in the average quality of teachers in schools serving a predominantly black student body. This outcome occurred because such schools found it increasingly difficult to attract and retain quality teachers (Jepsen and Rivkin ,2002; Bohrnstedt and Stecher (eds), 2002).

Whole school reform

In contrast to piecemeal reforms that address specific inputs to the educational process such as the quality of teachers or the size of their classes, whole school reforms are designed to improve achievement by changing multiple factors within a school simultaneously and in a coherent manner. A variation of this reform effort is the promotion of small high schools, an effort supported in recent years by significant funding from the Bill and Melinda Gates Foundation. Whole school reform models are typically designed for schools serving low- performing students. Because many of these

schools disproportionately serve minority students, a successful reform effort of this type could potentially reduce the black-white achievement gap by raising the performance of low-scoring black students.

Because no one has a monopoly on ideas of how to reform schools, there exist a large number of whole school reform models, the best known of which is Success for All, developed by Robert Slavin at Johns Hopkins. Many other whole school reform models are connected with the New American Schools initiative that began in 1991. That non-profit set up a competition for the best whole school reform model and ultimately chose to support 11 models from 600 proposed designs. In 2002, Rand reported the results of its comprehensive study of the models by seven of the design teams that were in 550 schools.

The results of the New American Schools (NAS) project have been disappointing largely because many of the schools were unable to implement the model fully (Berends et al, 2002). In practice they needed a lot of assistance and often faced barriers in the form of district bureaucracies, state and district polices, and resistance from unions. Emerging from this experience are two lessons. One is that individual schools are part of a larger system, which makes it hard to change them without changing the system of which they are a part. Another is that within schools, the quality of leadership is key.

In contrast, most of the more than 45 studies of the Success for All model generate positive achievement effects. Success for All is an early intervention model designed to ensure that every student reaches third grade ready to read. Critics of the program argue that the positive findings may be biased upward either because most studies include the model designer, Robert Slavin, as a member of the evaluation team or

because of the way students or schools select into the program. The selection problem is at least partially addressed in a recent study in which schools interested in participating in the program were randomly selected to receive the Success for All program in grades K-2 or grades 3-5 (Borman et al, 2007). As is the case for most of the previous studies, this national randomized field trial high poverty schools across 11 states generates statistically significant positive achievement effects in reading for the Success for All program. [17]

One thorny issue that arises in the context of any school-level reform is the extent to which a model that works well at the initial site under the close supervision of the team that designed the model can be replicated elsewhere. Another is whether the results can be generalized to schools beyond those that choose to participate in the program. Even in the randomized study of Success for All just referred to, the results are at best generalizable to schools interested in implementing the program.

These issues are further illustrated by the First Things First (FTF) Program, a small-school reform model designed to improve the achievement of economically disadvantaged middle and high school students .Though FTF generated impressive achievement gains in math and reading and improvements in other outcomes for students in its home site of Kansas City, Kansas, the effects were far less positive and far less consistent in the expansion sites. The evaluators speculate that among the reasons for the less impressive results were the weaker support from the districts in the new sites, the relatively long time needed for program development, and the inability of the designer to

[17] At the same time, no such effects emerged from a completely independent observational study of three whole school reform models, including Success for All, in New York City (Bifulco, Duncombe and Yinger, 2005). Though carefully done, however, that study is subject to all the caveats of studies based on quasi-experimental data.

provide adequate technical support for the project over an extended period of time. (Quint et al, 2005). The evaluators conclude that implementing such a program is hard work and requires significant commitment of educators not only at the school level but also at the district level.

The bottom line is that some school-based reform models appear to have the potential to raise student achievement of some low performing students, including minority students. Taking such programs to scale, however is a difficult undertaking with no guarantee of widespread success. Further, the success of any school-based strategy will inevitably require the involvement and commitment of district, as well as school, level, officials.

IV. School-based accountability programs

By school based accountability programs, I am referring to systems that use measures of student outcomes – primarily student achievement as measured by test scores – to hold schools accountable for improving the performance of their students. The federal No Child Left Behind Act (NCLB) of 2001 is the most prominent example. That legislation requires every state to test all students in reading and math annually in grades 3-8 and once in high school. It uses those test scores, reported separately by racial and income subgroups within schools, to hold individual schools accountable for making adequate yearly progress toward the ultimate goal of 100 percent proficiency. Many states, particularly southern states such as Texas and North Carolina, had their own quite well-developed accountability systems well before the federal law spread school based accountability to all states.

This type of top-down administrative system differs from other forms of accountability, such as political accountability that would hold policy makers accountable through the political process or to accountability through market processes. In the next section on choice and competition, I return to market-based accountability. In this section, I restrict the discussion to test-based administrative accountability.

There are at least three rationales for this type of accountability, not all of which have direct links to the racial achievement gap. For the proponents of standards based reform, for example, test based accountability is simply one part of a more coherent reform strategy designed to promote the ambitious educational outcomes required in this increasingly global society. The goal of this reform strategy is to align all components of the education system, including teacher training and capacity building, toward the overall goal of high student performance, progress toward which is measured by student test scores. Though such proponents emphasize the importance of high standards for all students, the standards-based reform strategy is not directly targeted on achievement gaps. Instead, the goal is to increase overall achievement.

A second rationale for test-based accountability it that it serves as a stand-alone policy designed to address the perceived problem that educators are shirking their responsibilities and simply are not working hard enough or" smart" enough to generate the desired outcomes. Economists often use the language of the principal agent model to describe this situation. In the context of such a model, the challenge is to set up an appropriate incentive system to induce the agents – in this case the educators – to operate in ways compatible with the interests of the principal – in this case state policy makers and the public. By measuring, reporting and attaching positive consequences to strong

performance and negative consequences to weak school performance, policy makers provide incentives for schools and school districts to focus attention of what is being measured and ultimately to alter the way they operate. Concerns about the capacity of schools to respond or about inadequate resources clearly take a back seat to confidence in the power of incentives and sanctions to change behavior.

To the extent that shirking of teachers is the policy problem, rather than, for example, lack of resources, knowledge or professional skills, a test-based accountability system could potentially help narrow the black-white test score gap. For accountability to narrow the gap, however, two things would have to be true. The first is that, in the absence of an accountability system, the parents of black students would have to be less vigilant than the parents of white students in monitoring the quality of the children's schools and classrooms so that the introduction of accountability would have a differentially positive effect on black students. Though information on this point is lacking, the lower levels of education or income of black parents relative to white parents could well render them less able or willing to exert an influence in the schools that white parents. In addition, any differential monitoring would have to occur in schools that were not racially balanced. Otherwise any monitoring by white parents would benefit black students along with white students and hence the introduction of an accountability system would have little or no effect on the black-white achievement gap. If shirking is indeed the problem, accountability could well be a useful tool for closing gaps. To the extent that the low achievement of minorities reflects larger social forces or of their exposure to teachers with weak credentials, however, simply putting pressure on teachers to work harder or smarter will do little to reduce the gap.

Finally, as has been emphasized by groups such as the Citizens Commission on Civil Rights and the Education Trust, school accountability – especially as implemented under NCLB with its attention to subgroups – can be viewed as a tool for directly addressing the problem of educational inequities, and in particular racial achievement gaps. By setting high standards for all students and by focusing attention on the students whom the education system has been leaving behind, namely minorities, students from low income families and those who are disabled, school accountability programs could serve to raise the achievement of those historically low performing groups. .

Design matters

Regardless of the rationale, the potential for test-based accountability systems to contribute to reducing the black-white test score gap will depend on how the system is designed. Among the many important design issues, perhaps the most important is whether to use a status model or a model based on individual student growth to judge the effectiveness of individual schools. A status model essentially looks at levels of achievement – typically defined as the percent of students who reach a designated level of proficiency -- while a growth model – often called a value-added model -- focuses on the average gains in learning of individual students from one year to the next. NCLB is currently based on the status approach. If the important values are providing realistic incentives for school improvement, especially for schools at the low end of the performance distribution, the growth approach, though itself somewhat flawed, is clearly preferred to the status model.

The status model is appealing to some observers because it sends a clear signal that the goal is high achievement for all students. The problem, though, is that simply

sending a signal does not assure that the outcome will be achieved, and may well lead to unintended and undesired side effects such as narrow teaching to the test or possibly even cheating. As has been documented in many studies including Clotfelter and Ladd (1996), status models are not well designed to promote an equity agenda because they inevitably favor the schools with the most advantaged students. This pattern emerges because of the high positive correlation across schools between the socio-economic status of the students and their achievement. As a result, the more advantaged schools have greater incentives to improve than to do the schools serving low-performing students who perceive little chance of positive recognition.

If the important values are providing realistic incentives for school improvement, especially for schools at the low end of the performance distribution, the growth approach, though itself somewhat flawed, is clearly preferred to the status model.

Achievement effects of accountability

At this point, less is known about how accountability programs affect student achievement than one might expect given its centrality to the current education policy debate.[18] Though a recent study finds that student achievement in reading and math has risen in most states with three or more years of comparable test score data since NCLB was enacted, the authors emphasize the difficulty of determining whether NCLB caused the increase (Center for Education Policy, 2007). In addition, national trends of test scores based on the National Assessment of Educational Progress (NAEP) provide no support for a large and demonstrable effect of NCLB on student achievement. Though the reading scores of eighth graders were rising prior to NCLB, they have remained generally constant in the post NCLB period. At the same time, the upward trend in

[18] Much of the following discussion is based on Figlio and Ladd, 2008.

eighth grade math scores both before and after NCLB are consistent with the view that state-level accountability programs that preceded the federal legislation have raised student achievement in math.

In general, the state-level experiences provide a better source of information of the achievement effects of accountability, both because many states have been using test-based accountability systems for longer periods of time than the federal government and because of the possibility of comparing trends in a particular state to those in other states or to the nation. Even, here, however, the effects are not fully clear.

Of particular interest for the black-white achievement gap is the Texas experience, because that state provided the model for the U.S approach of focusing attention on subgroups within schools. Other evidence about the achievement effects of accountability emerge from cross-state studies that make use of the variation across states in the strength of their state accountability systems or in the timing of their introduction to tease out their causal impacts on student achievement. A central issue in all the studies is how best to measure achievement. The main choices are the high stakes state test on which a particularly accountability system is based or the NAEP which to which no specific stakes are attached but which has the advantage of being comparable across states. As illustrated below for Texas, positive results on the high stakes test may not translate into positive results on the low stakes test. In general the low-stakes NAEP test is probably the better indicator of student learning. The exception is in those cases in which the state or district's curriculum differs significantly from the material tested on NEAP.

Texas. Perhaps of most interest are the results for Texas, since that state's accountability system served as the model for NCLB. After a series of education reforms starting in the early 1980s, Texas introduced in 1990 a criterion-referenced testing program called the Texas Assessment of Academic Skills (TAAS) that was designed to shift the focus from minimum skills to higher-order thinking skills (see description in Haney, 2000.) Schools were held accountable not only for the overall pass rates on TAAS in the school but also for the pass rates of four student subgroups: African Americans, Hispanics, whites, and economically disadvantaged students

Between 1994 and 1998, TAAS test scores in both math and reading increased quite dramatically, suggesting that the state's accountability program had a large and positive impact on student achievement. Analysis by Klein et al (2000), however, showed that the large gains on TAAS did not translate into comparable large gains in the lower-stakes Texas NAEP scores.[19] Moreover, only for the white fourth graders did the reading gains of Texas students on NAEP exceed the gains of their counterparts nationwide. A somewhat more positive story emerges for Texas fourth graders in math. Once again, the TAAS gains exceeded the Texas NAEP gains but in this case, the latter gains exceeded the national gains for all three racial groups.

Most relevant for this discussion is that the TAAS and NAEP results generate conflicting stories about how accountability affected racial achievement gaps in Texas. In particular, the gaps between blacks and whites in fourth grade reading and math and in

[19] Using "effect sizes" which are measured in standard deviations and hence can be compared across tests, Klein et al. (2000) reported the following effect sizes for achievement in reading for Texas fourth graders between 1994 and 1998: TAAS scores increased by 0.39 for white fourth graders, 0.49 for black fourth graders, and 0.39 for Hispanic fourth graders. In contrast the gains on the Texas NAEP were far smaller at 0.13, 0.14, and 0.14, respectively. (Klein et al. 2000).

eighth grade math based on the TAAS scores decreased significantly between 1994 and 1998, while the comparable gaps based on the NAEP increased slightly (Klein et al, 2000, pp. 10-11). Similar patterns also emerge for Hispanics. Klein et al. speculate that the reasons for the differing patterns for TAAS and NAEP results is that Texas teachers may be teaching very narrowly to the TAAS and that the schools serving minority students may be doing this even more than other schools. Thus, even in Texas, the evidence is at best mixed about the power of an accountability system to reduce racial gaps.

Cross state studies. Other studies generate similarly mixed results with respect to effects by racial group. At least one careful study (Carnoy and Loeb, 2002) for the late 1990s find larger effect sizes on passing rates at the basic level on NAEP for black and Hispanic students than for white students. Other studies with different outcome measures find different patterns. In particular, Hanushek and Raymond (2005) find essentially no effects of accountability on the gains in achievement between fourth and eighth grade of black students, but positive effects for Hispanic students, effects that are consistent with early findings by racial group for 7th graders in Dallas (Ladd, 1999). Effects of accountability on racial achievement gaps are similarly mixed. The Hanushek and Raymond study finds that state accountability systems may have reduced the gap for Hispanics but expanded it for blacks.

Conclusions about accountability

Though accountability programs that focus on specific groups may have some potential for reducing black-white achievement gaps, the overall evidence suggests their effects are likely to be small and are most likely to emerge in the lower grades. There is

little evidence to date of their ability to reduce the gap at higher grades and in terms of higher order skills. Though somewhat discouraging, this conclusion should not be too surprising. These findings are consistent with the view that the underperformance of black children relative to white children in many cases has far less to do with the teacher shirking that motivates stand-alone accountability programs and far more to do with a host of other factors both inside and outside the schools. The challenge for education policy makers at this point is to develop accountability systems that take greater account of the different skills and capacities that children bring to school, that shift the focus away from test score results to the strengthening of stronger instructional practices within schools, and that cast the bright light of accountability on participants other than just teachers in the education process, including district and state policy makers who determine the terms under which individual school operates.

V. School choice programs

As is the case for test-based accountability programs, expanded parental choice of schools has been promoted for many reasons, not all of which are related to the challenge of reducing the black-white test score gap. The following discussion briefly evaluates three mechanisms through which expanded parental choice of schools might conceivably reduce the black-white test score gap. The fact that the evidence for success in each case is at best mixed raises doubts about the potential for parental choice programs to reduce the gap. Indeed, the more general concern is that expansion of choice could well widen achievement gaps. Because additional choice serves multiple goals, this conclusion need not mean that additional choice is undesirable. It does mean, however, that as policy

makers respond to rising demands for more parental choice, they need to take care in designing those programs in ways that are least likely to widen racial achievement gaps.

Effects on achievement through competition

Some proponents favor more parental choice because they believe that when schools are faced with the possibility of losing students and the funding that accompanies them, the schools will be forced to become less complacent and more productive. The U.S. evidence to date, however, suggests that competitive pressures of this type are not likely to have much impact on the black-white test score gap. One reason is that even at best competitive pressure appears to have very small positive impacts on student achievement (Gill and Booker, 2008). That conclusion is based on existing studies of various types of choice programs, many of which are still quite small. Although larger positive results could conceivably emerge as the U.S. introduces more choice and competition, the evidence from other countries with more extensive choice, such as Chile, tend to confirm this conclusion that any achievement effects arising through the mechanism of competition are likely to be small.

Another is that parental choice and competition may exacerbate the challenges faced by the low-performing schools serving many black students. To the extent that the students who exercise their power to leave are the more motivated students or the ones from families who are more actively involved in the school, the outcome could well be greater concentration of low-performing students, and hence more challenging-to-educate students, in those schools (Fiske and Ladd, 2000). The result is that the performance of students in those schools, including many black students, may well fall in the face of competitive pressures.

Achievement effects on the choosers

Potentially more important from the perspective of the black-white test score gap is that giving parents more choice may improve the schooling options, and hence outcomes, for black students more than for white students. Two considerations appear to support this possibility. One is that black students, especially those from low-income families , typically have fewer schooling options than white students. That occurs because the combination of their lower income and various features of the housing market, including zoning restrictions and discrimination, tend to limit the neighborhoods – and hence schools -- available to them under a geographic school assignment system and because the lower average incomes of their families restricts their ability to enroll in private schools. The other consideration is the perception -- and to some extent the evidence -- that private schools, particularly Catholic schools, generate higher achievement than public schools. In fact, the positive achievement effects of Catholic schools are far smaller than once believed. At the same time, the evidence is consistent with positive differential effects for African Americans in urban areas.

These considerations, along with others, open the possibility for charter schools or voucher programs to reduce gaps to the extent they give black students access to better schools. Charter schools, which are now enabled by state legislation in 40 states plus the District of Columbia, are public schools that are publicly funded, but are operated by non-governmental organizations under charter from a public agency and are schools of choice in that no students are assigned to such schools. Though the state enabling laws differ from state to state, one of the goals of such laws is typically to provide additional options for disadvantaged students. Voucher programs, in contrast, expand the options

for students not explicitly by the introduction of new schools but rather by providing public funding for students to attend private schools. Such demand-side funding may simply increase the demand for slots in existing private schools or, depending on the scale of the program, may expand the supply of private schools. Charter school programs are currently far more common than voucher programs in the United States. The most well known publicly funded voucher programs are in Milwaukee, Cleveland and Washington, D.C. A number of privately funded voucher programs, typically called scholarship programs, have operated in other cities, including New York City, Washington, D.C. and Dayton, Ohio.

The key policy question of interest here is whether black students who exercise their option to choose a different school under either of these programs achieve at higher levels than they would have had they remained in the traditional public schools. Answering that question with confidence is challenging because the students who take advantage of the new schooling options are likely to differ in systematic ways from those who remain in the traditional public schools. Hence, researchers have had to develop empirical methods that keep to a minimum the biases that arise from self-selection.

Charter schools. One method used quite extensively in the literature on charter school effects involves the estimation of longitudinal models based on the test scores of individual students who are observed in both traditional public schools and in charter schools, with indicator variables to control for the characteristics of students, such as their motivation, that do not change over time. Such models solve the selection problem by measuring the gains in achievement of students while they are in charter schools relative to the gains of those very same students while they are in traditional public

schools. The disadvantage of this approach is that the sample is restricted to the students who are switching from one type of school to the other and may not be representative of all charter school students. A second method is to make use of the multiple natural experiments that arises when charter schools are oversubscribed and have to select students through a random lottery process. With this approach, the students who lose the lottery can serve as a control, or comparison group, for the students who are selected into the school. Though the random assignment component of this approach solves the selection program, the downside is that the results are generalizable only to the types of charter schools that are oversubscribed.

Careful studies of charter schools based on longitudinal data typically show little or no positive overall achievement effects. In fact, research of this type finds large negative achievement effects in North Carolina, negative overall effects in Texas for newly established charter schools and no differential effects for more mature charter school, and similar patterns in Florida (Bifulco and Ladd, 2006; Sass, 2006, Hanushek, Kain, Rivkin and Brand, 2005). The more negative effects in North Carolina may well reflect the failure of that state to remove the charters of schools that are underperforming.

The results from studies based on oversubscribed charter schools are generally more positive (Hoxby and Rockoff, 2005). These latter studies are important as an existence proof. That is, they document the potential for certain types of charter schools – the oversubscribed ones included in the studies -- to raise achievement for the types of students likely to apply to them. To the extent that such models can be successfully expanded to other sites, and that such models serve black students, their success indicate the potential for some black students to benefit from charter schools. At the same time,

however, other black students could well be harmed by the availability of charter schools. That conclusion emerges clearly from the North Carolina experience where the black students who end up in racially segregated charter schools fare far less well than other charter school students. (As documented by Bifulco and Ladd (2007), the net effect of charter schools in that state has been to expand the black-white test score gap. Though that outcome may not emerge in other states, it highlights the need for policy makers to be alert to the effects of their policy decisions on the black-white test score gap.

Voucher programs. With respect to voucher programs, evaluation of the heavily studied initial Milwaukee program shows that means-tested voucher programs can be designed to successfully expand the schooling opportunities for black families. Whether such programs increase the learning of the participants, however, is more controversial. The best of the three studies of the initial Milwaukee program finds small positive achievement gains in math but none in reading (Rouse, 1998) , but all the studies of that program are bedeviled by the challenge of finding an appropriate control group since students were not randomly assigned to receive a voucher.

A better approach for measuring achievement effects is to do field experiments in which applicants to the voucher program are randomly selected into the program or into a control group. Such field experiments have been used to evaluate privately funded voucher programs in New York City, Dayton, Ohio and Washington, D.C. Based on three years of the voucher programs in New York and Washington, D.C. and two years in Dayton, researchers William Howell and Paul Peterson (2002) find no evidence of a general achievement difference between the public and the private schools. In no year and in no individual city (other than the second year in Washington) was there evidence

that students who shifted to private schools achieved at higher average levels than students who remained in the public school system. Further, when the analysis was disaggregated by the race of the students, no differences emerged for either white or Hispanic students.

Positive differences in achievement did emerge, however, for African Americans. Based on their preferred estimates, which disproportionately weight the results from New York City on the ground they were the most stable over time, the authors conclude that African Americans who switched to private schools scored about 3.9, 6.3 and 6.5 percentile points higher than comparable students in the control group in the first three years of the program. These effects are about two-thirds the size of the differences that emerged for minority students exposed to smaller classes in the Tennessee class size experiment. The differences were consistent, however, neither across neither cities nor grades. In New York City, for example, the positive differential emerged clearly and consistently only for students in the fifth grade (Howell and Peterson, 2002, Table 6.2 and Table D.1). Further, a reanalysis of the New York data by Krueger and Zhu (2002) has generated questions about the robustness of the positive findings for African Americans in that city. Finally, it is not at all clear that any positive effects of private schools can be extrapolated to an expanded voucher program, even one targeted at African American students. There is no guarantee that any new private schools established in response to an expanded voucher program would be of the same quality as the more established schools involved in this small scale initiative.

Thus, the power of voucher programs to reduce the black white achievement gap by raising the achievement of black students who use vouchers to attend private schools remains to be documented. The evidence to date is not promising.

Promoting racial integration through parental choice. One final mechanism, albeit an unlikely route, through which greater choice could potentially reduce the black-white gap is worth exploring. To the extent that parental choice were to reduce racial segregation it could, for the reasons discussed earlier, potentially lead to a more even distribution of teacher quality across rates.

To be sure, the expansion of options for parents to choose the schools their children attend has historically generated greater, not less, racial segregation in the U.S. That outcome has resulted in part from white flight, which may have been motivated in part by "outgroup avoidance" (Saporito 2003), that is, the desire of the dominant group to minimize contact with the other group. Greater choice would also increase segregation if members of each racial group prefer to associate with others like themselves. Nonetheless two other mechanisms could potentially operate in the other direction. Given the high levels of residential segregation in U.S. metropolitan areas, greater choice could possibly reduce school segregation by providing families access to schools that are more integrated than the types of neighborhoods available to them. This mechanism is most applicable to the black families whose housing decisions are constrained by zoning restrictions or racial discrimination or in the housing market. In addition, explicit policy decisions about resources and the location of schools could promote greater integration. If parents prefer schools with more resources, the generous funding of schools in minority neighborhoods in the form of magnet schools, for example, may attract students

from other neighborhoods with different family backgrounds. Also, the creation of specialized schools with specific themes, such as district wide magnet schools in science or theater, may draw students from across the district or even from other districts. In this way, school specialization can widen a school's catchment area beyond racially isolated neighborhoods, thereby reducing racial segregation. Thus, whether choice programs are likely to increase or decrease racial segregation is an empirical question.

Bifulco, Ladd and Ross (2007) use data from Durham, NC to examine that issue in the context of choice programs consistent with the 2007 Supreme Court ruling. This urban school district represents a useful case study because it has avoided the use of racial criteria in its school assignment programs since the 1999 4th Circuit Supreme Court case that put a damper of race based policies in that circuit, it has long had a liberal school transfer program, and it offers a variety of schools of choice, including magnet schools, charter schools and year round schools. The availability of data on individual students makes it possible for the authors to track students to the schools they attend. The main question is whether the choice programs in that urban district increased or decreased the racial segregation of the schools.

Consistent with various predictions from the literature, the authors find evidence that substantial numbers of white families used the school choice options to avoid schools with concentrations of racial minorities, and that some black families used the options to select more racially isolated environments. The segregating effects of such choices, however, were largely offset, especially at the middle school level where the district has had some success in establishing magnet schools attractive to white families, by the students who made racially integrating choices. As a result, Durham's school choice

programs increased racial segregation but only by a small amount. Although the small size of the increase may be welcomed by some observers, the finding that racial segregation increased at all is not good news for those hoping that choice programs might serve as a mechanism for reducing racial segregation. Moreover, such programs resulted in far greater segregation by class and student achievement, an outcome that might also work to the disadvantage of black students who tend to be overrepresented among low SES and low-performing students. At the same time, studies such as this one are unable to explore the broader, general equilibrium effects of school choice policies. The availability of such policies might, for example, affect the residential choices that parents make and thereby indirectly influence a range of other policy outcomes.

VI. Conclusion

Emerging from this discussion is that none of the various school-related policies discussed here is likely to play a major role in reducing the black-white achievement gap. Some policies, however, undoubtedly have more potential than others. Most promising appear to be strategies to promote small class sizes in the early grades and to even out the quality of teachers across schools serving different racial groups. Instead, major reductions in the achievement gap will require policy attention to the larger social forces that lead to differences by race in what children bring to the classroom.

Despite this pessimistic conclusion about the power of school policies by themselves to reduce the gap, well designed school-related policies are still a crucial component of any gap-reduction strategy. The reason is that even vast improvements in various social policies relevant for education, such as improved health care or nutrition for infants and young children and expanded access to high quality pre-school

opportunities for children from disadvantaged families, will fail to reduce the gap if the education system itself distributes resources unequally across students of different races. As highlighted in the first section of this paper, the more unevenly that students of different races are distributed across schools, the more potential there is for resources, such as quality teachers, to be unevenly distributed by race. Hence, a major challenge for policy makers is to maintain whatever pressure they can to limit the resegregation of schools. Unfortunately, because the Supreme Court has limited the direct powers of districts to promote racial integration through explicitly race-based student assignment programs, districts must rely on less direct strategies such as balancing schools by socio-economic status or the judicious use of funding for magnet schools or the location of new charter schools or special programs to promote the goal of racial balance. As illustrated by the recent experience in Charlotte-Mecklenberg, North Carolina, failure to pay attention to racial balance can have serious consequences for the black-white distribution of teachers. But even in the absence of future changes of this type in the distributions of students and teachers, policy makers will need to be vigilant in pursuing strategies designed to counter the black-white differences in educational inputs that currently exist. That will require policies specifically designed to improve the skills of the teachers serving black children and to improve the schools they attend.

References

Angrist, J and V. Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics.. 533-574.*

Berends, M., Bodilly, S. and Kirby, S. 2002. A Decade of Whole School Reform: The New American Schools Experience. RAND Research Brief.

Bifulco, R, W. Duncombe, and J. Yinger. 2005. "Does Whole-School Reform Boost Student Performance? The Case of New York City." *Journal of Policy Analysis and Management. Winter. vol. 24(1),*

Bifulco, Robert and Helen F. Ladd. 2006. "The impact of charter Schools on student achievement: Evidence from North Carolina." *Education Finance and Policy, 1. pp. 50-90.*

Bifulco, Robert and Helen F. Ladd, 2007. "School choice, racial segregation and test score gaps: Evidence from North Carolina's charter school program. *Journal of Policy Analysis and Management.* 26.

Bifulco, R., H. Ladd, and S. Ross. 2007. "Public School Choice and Integration: Evidence from Durham, N.C." Available as a CALDER working paper. Caldercenter.org.

Boger, John Charles. 2000. "Willful Colorblindness: The New Racial Piety and the Resegregation of Public Schools," *North Carolina Law Review* 78 (September 2000), 1719-1796 .

Bohrnstedt, G. W. and B. M. Stecher (eds). 2002. *Class Size Reduction in California.* CSR Research Consortium. California Department of Education.

Borman, G., Slavin, R. Cheung, A. Chamberlain, A. Madden, N. and Chambers, B. 2007. "Final Reading Outcomes of the national Randomized Field Trial of Success for All. *American Educational Research Jounal. 44(3), 703-731.*

Boyd, D., P. Grossman, H. Lankford, S. Loeb, and J. Wyckoff. 2006. "How changes in entry requirements alter the teacher workforce and affect student achievement." *Education Finance and Policy. 1(2) 176-216.*

Carnoy, M. and S. Loeb. 2002. "Does external accountability affect student outcomes? A cross-state analysis." *Educational Evaluation and Policy Analysis* 24(4), pp. 305-331.

Center on Education Policy. 2007. *Answering the Question That Matters Most: Has Student Achievement Increased Since No Child Left Behind?"* By the author. June.

Century Foundation Task Force on the Common School (ed.) 2002. *Divided we fail: Coming together throughout public school choice.* New York: The Century Foundation Press.

Chaplin, D. 2002. "Estimating the impact of economic integration of schools on *racial integration." In The Century Foundation Task Force on the Common School (ed.*) *Divided we fail: Coming together throughout public school choice.* New York: The Century Foundation Press. Pp. 86-113.

Clotfelter, C. and H. Ladd. 1996. "Recognizing and rewarding success in public schools." In H. Ladd, editor, *Holding Schools Accountable: Performance-Based Reform in Education*. Washington, D. C. Brookings Institution Press, pp.23-64.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor, 2003. "Segregation and Resegregation in North Carolina's Public School Classrooms." *North Carolina Law Review*, 81, no. 4, May.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness *Journal of Human Resources.* vol. 41, number 4, Fall, pp. 778-820.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor. 2007a. "Teacher credentials and student achievement: Longitudinal analysis with student fixed effects" *Economics of Education Review*, Dec. .

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor. 2007b. Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. NBER and CALDER working paper, 2007.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor, 2008. "School Segregation under Color-Blind Jurisprudence: The Case of North Carolina." Available as a Sanford Institute Working Paper and on the CALDER web page (Caldercenter.org).

Clotfelter, Charles, Elizabeth Glennie, Helen F, Ladd, and Jacob L Vigdor. Forthcoming. "Would Higher Salaries Keep Teachers in High Poverty Schools? Evidence from a Policy Intervention in North Carolina." *Journal of Public Economics.*

Corcoran, Sean P. and William N. Evans. 2008. "Stalled Progress in Closing the Race Achievement Gap: The Role of Teacher Quality." In Katherine Magnuson and Jane Waldfogel, eds. *Steady Gains and Stalled Progress: Inequality and the Black-white Test Score Gap*. New York: Russell Sage.

Figlio. David and Helen F. Ladd. 2008 “School Accountability and Student Achievement.” In Helen F. Ladd and Edward B. Fiske (eds), *Handbook of Research in Education Finance and Policy* Routledge Press. PP. 166-182.

Finn, J. D and C.M. Achilles. 1990. “Answers and Questions about Class Size: A Statewide Experiment. *´American Educational Research Journal, 27, 557-577,*

Fiske, Edward B. and Helen F. Ladd. 2000. *When Schools Compete: A Cautionary Tale*. Washington: Brookings Institution Press.

Gill, Brian and Kevin Booker. 2008. “School Competition and Student Outcomes.” In Helen F. Ladd and Edward B. Fiske, *Handbook of Research in Education Finance and Policy.* Routledge, pp. 183-202.

Goldhaber, Dan. 2008. “Teachers Matter, but Effective Teacher Quality Policies Are Elusive”. In Helen F. Ladd and Edward B. Fiske (eds) *Handbook of Research in Education Finance and Policy.* Routledge. Pp. 146-165.

Glazerman, S. , D. P.Mayer, and P. T. Decker. 2006 . “Alternative Routes to Teaching: The impacts of Teach for America on Student Achievement and Other Outcomes.” Journal of Policy Analysis and Management, 25(1), 75-96.

Haney, W. 2000. “The Myth of the Texas Miracle in Education.” *Educational Policy Analysis Archives* 8(41).

Hanushek, Eric. 1999. Some Findings From an Independent Investigation of the Tennessee STAR Experiment and From Other Investigations of Class Size Effects.” *Educational Evaluation and Policy Analysis.* 21(2), 143-163.

Hanushek, E. J. Kain and S. Rivkin, 2004 “Why Public Schools Lose Teachers.” *Journal of Human Resources.* 39(20 326-354.

Hanushek, E., J. Kain, and S. Rivkin and G. Brand. 2005. “Charter School Quality and Parental Decision Making with School Choice. “ NBER Working Paper 11252. Cambridge, MA.

Hanushek, E and M. Raymond. 2005. “Does School Accountability Lead to Improved School Performance? “ *Journal of Policy Analysis and Management* , vol. 24, no. 2, pp. 297-329

Hedges, L., R. Laine and R. Greenwald. 1994. “An Exchange” Part I: Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes.” *Educational Researcher.* 23(3) 5-14.

Hill, Heather C. 2007. “Learning in the Teaching Workforce.” *Future of Children,* vol. 17, no. 1 (Spring), pp. 11-127.

Howell, William G. and Paul E. Peterson. 2002 *The Education Gap: Vouchers and Urban Schools* .Washington, D.C. Brookings Institution Press.

Hoxby, C.M. and J.E. Rockoff. 2005. "The Impact of Charter Schools on Student Achievement." Unpublished paper, Harvard University.

Jacob, Brian . 2007. "The Challenges of Staffing Urban Schools with Effective Teachers." *The Future of Children.*

Jencks, Christopher and Susan Mayer. 1990. "The Social Consequences of Growing Up in a Poor Neighborhood." In *Inner City Poverty in the United States,* eds. Laurence E. Lynn, Jr. and Michael G.H. McGeary. Washington, D.C.: National Academy Press. Pp. 111-186.

Jencks, Christopher and Meredith Phillips (eds). 1998. *The Black-White Test-Score Gap* Washington, DC: Brookings.

Jepson, C. and Rivkin, S. 2002. "What is the Tradeoff Between Smaller Classes and Teacher Quality?" Working Paper 9205. National Bureau of Economic Research.

Kahlenberg, R. 2001. *All Together Now: Creating Middle-Class Schools Through Public School Choice."* Washington, D.C.: Brookings Institution.

Klein, S., L. Hamilton, D. McCaffrey, and B. Stecher. 2000. "What Do Test Scores in Texas Tell Us?" *Education Policy Analysis Archives* 9(49).

Krueger. Alan. 1999. "Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics.* November.

Krueger, Alan. 2002. "Understanding the Magnitude and Effect of Class Size on Student Achievement." In L. Mischel and R. Rothstien, eds., *The Class Size Debate.* Chapter 1. Washington, D.C.: Economic Policy Institute.

Krueger. A. and P. Zhu. 2002. "Another Look at the New York City School Voucher Experiment." NBER Working Paper 9418. Cambridge MA: National Bureau of Economic Research.

Ladd, H. 1999. "The Dallas School Accountability and Incentive Program: An Evaluation of Its Impacts on Student Outcomes." *Economics of Education Review.*

Lankford, H. Loeb, S. and Wyckoff, J. 2002. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. " *Educational Evaluation and Policy Analysis,* 24(1), 38-62.

Mont, Daniel and Daniel I. Rees. 1996. "The Influence of Classroom Characteristics on High School Teacher Turnover," *Economic Inquiry* 34 (January), 152-167.

Mosteller, Frederick. 1995 "The Tennessee Study of Class Size in the Early School Grades. " *The Future of Children.* Vo. 5, no. 2. Critical Issues for Children and youths, pp. 113-127.

Murnane. R. and R. Olson. 1989. "Will There be Enough Teachers?" *American Economic Review, 79, ppl. 242-246.*

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. 2000. "Do Minorities and the Disadvantaged Benefit More from Small classes? Evidence from the Tennessee Class Size Experiment" *American Journal of Education,* 109, 1-26.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. 2004 "Do Minorities Experience Longer Lasting Benefits from Small Classes? " *The Journal of Educational Research.* 98(2), 94-100.

Murnane, Richard J. and Frank Levy. 1996. "Evidence from Fifteen Schools in Austin, Texas" in Gary Burtless,ed. *Does Money Matter?* Brookings.

Reardon, Sean F., John T. Yin and Michael Kurlaender. 2006. "Implications of Income-Based School Assignment Policies for Racial School Segregation. *Educational Evaluation and Policy Analysis.* vol. 28, no. 1 (Spring) , pp. 49-76.

Rouse, C. 1998. . "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics* 113, 553-602.

Quint, J. Bloom. H, Black, A. R. Stephens, L. and Akey, T. (2005). *The Challenge of Scaling up Educational Reform.* New York, NY: MDRC (http://eric.ed.gov/ERICDocs/data,ericdocs2sqI/content_storgage_01.0000019b/80.29.da/76.pdf) accessed January, 2008

Scafidi, Benjamin, David Sjoquist, and Todd R. Stinebrickner. 2002. "The Impact of Wages and School Characteristics on Teacher Mobility and Retention," Unpublished paper, September.

Saporito, Salvatore. 2003. "Private Choices, Public Consequences: Magnet School Choice and Segregation by Race and Poverty." *Social Problems* 50, 181-203.

Sass, Tim 2006. "Charter schools and student achievement in Florida. *Education Finance and Policy. 1, 91-122.*

Vigdor, Jacob and Jens Ludwig. 2008. "Segregation and the Black-White Test Score Gap." In Katherine Magnuson and Jane Waldfogel, eds. *Steady Gains and Stalled Progress: Inequality and the Black-white Test Score Gap*. New York: Russell Sage.

Table 1:Minority students in high and low percent minority schools, by school level for selected years, all schools in North Carolina (percentages, except where noted).

	Quartile 1 (high minority)	***Quartile 4 (low minority)***	***Difference (percentage points)***
Elementary			
1995	67.7	4.8	-62.9
1999	74.6	6.2	-68.3
2004	81.0	8.9	-72.2
Middle			
1995	66.5	7.8	-58.7
1999	70.8	8.7	-62.1
2004	77.7	10.4	-67.2
High school			
1995	66.3	5.9	-60.4
1999	69.6	7.2	-62.4
2004	74.0	9.3	-64.8

Notes. Quartile 1 and quartile 4 refer to quartiles of the distribution of schools by level and year based on the percentage of students in the school who are black, Hispanic, or Indian. The entries are the average percentages of minorities, weighted by the size of each school. Based on data from the North Carolina Department of Public Instruction, provided through the North Carolina Education Research Data Center,

Table 2 Teachers with less than 3 years experience in high and low percent minority schools, by school level for selected years (percentages, except as noted)

	Quartile 1 (high minority)	***Quartile 4 (low minority)***	***Difference (percentage points)***
Elementary schools			
1995	17.9	13.4	4.5
1999	21.9	14.5	7.4
2004	19.3	12.3	7.0
Middle schools			
1995	20.8	14.4	6.4
1999	25.1	17.2	7.9
2004	25.2	13.5	11.7
High schools			
1995	15.1	12.3	2.7
1999	18.1	13.4	4.7
2004	18.3	12.0	6.3

Notes. Quartile 1 and quartile 4 refer to quartiles of the distribution of schools by level and year based on the percentage of students in the school who are black, Hispanic, or Indian. The entries are the average percentages of minorities, weighted by the size of each school. Based on data from the North Carolina Department of Public Instruction, provided through the North Carolina Education Research Data Center,

Table 3. Teacher Quality by Race of Student the Charlotte/Mecklenburg School District, 2000/01 and 2005/06

Percentages of teachers with specified characteristics for typical student in each racial category

	3 + years experience		Top ¼ of test scores		Certified teacher*	
	2000/01	2005/06	2000/01	2005.06	2000/01	20005/06
Black	73.7	71.2	22.4	21.4	89.7	88.4
White	76.6	75.4	30.8	30.0	91.9	92.2
Difference	2.9	4.2	8.4	8.6	2.2	3.8

Note: Exposure rates of students by race to teachers in various categories are calculated as the average of teacher characteristics across schools weighted by the number of black and white students, respectively, in each school. * Teachers with initial or continuing certification in LicSal licensure data. Based on data from the North Carolina Department of Public Instruction, provided through the North Carolina Education Research Data Center,