Education Working Paper Archive

The Effect of Black Peers on Black Test Scores

September 12, 2007

David J. Armor
School of Public Policy
George Mason University

Stephanie Duck
School of Public Policy
George Mason University

Abstract

Recent studies have used increasingly complex methodologies to estimate the effect of peer characteristics—race, poverty, and ability—on student achievement. A paper by Hanushek, Kain, and Rivkin using Texas state testing data has received particularly wide attention because it found a large negative effect of school percent black on black math achievement. This paper replicates the HKR models using state testing data from North and South Carolina and national testing data from the Early Childhood Longitudinal Study. The replications fail to support the Texas results. In most models tested, black peer effects are small and not statistically significant, and in the few instances where effects are significant, they are much weaker than those found in Texas. Moreover, it appears that computational problems in the HKR study led to incorrect estimates for black peer effects.

THE EFFECT OF BLACK PEERS ON BLACK TEST SCORES [1,2]

There has been a long-standing debate among social scientists about the effects of school desegregation on the academic achievement of minority students. While desegregation plans have many components involving students, staff, and programs, all of which have potential impacts on learning, the most contentious issue has been whether school or classroom racial composition (or racial diversity) improves achievement.

Indeed, the origin of the controversy dates to the original *Brown* decision, which cited social science evidence showing that psychological harms from segregation might cause black students to do poorly in school. Although most legal scholars reject the notion that *Brown* relied on the psychological harm thesis, the citations—listed in the famous Footnote 11—sparked extensive research and writing on this question over the past five decades. In fact, the issue is still very much alive in school policy as well as the federal courts. During the 2006-07 term the Supreme Court reviewed two school desegregation cases from Seattle and Louisville where defendant school boards justified their continued reliance on race in student assignment because of educational benefits for minority students (Jefferson County Board of Education, 2006; Seattle School District No.1, 2006).

Expert studies and testimony in the Seattle and Louisville cases cited a study of black peer effects by Hanushek, Kain, and Rivkin (HKR).[3] Using a massive set of longitudinal test scores from the Texas state testing program, HKR developed special models for panel data and found a sizeable negative effect of school percent black on black achievement. This was surprising, given many prior studies over many years that had found weak relationships at best (St. John, 1975; Cook, 1984; Schofield, 1995; Armor, 2002). As the authors state, however, the Texas state achievement data allowed development and application of sophisticated fixed effect models that cannot be tested on smaller data sets.

Accordingly, the impetus for present study is to replicate the HKR model using other longitudinal achievement data to determine whether the large black peer effect found by HKR generalizes beyond the state of Texas. Because of the federal No Child Left Behind Act, the Texas data are no longer unique; state accountability systems in education have generated large-scale data bases of test scores in many states, and some states have allowed researchers access to that data. In addition, the Department of Education has developed several longitudinal achievement studies, one of which is the Early Childhood Longitudinal Study (ECLS). These single-cohort longitudinal studies also allow estimation of HKR-type panel models.

---

[1] Portions of this paper were presented at the Fall APPAM Research Conference in Madison, Wisconsin, November 4, 2006, and also appeared in the Education Working Paper Archives at the University of Arkansas.
[2] Special thanks to Brian Bucks for comments on earlier versions of this paper.
[3] Several versions of this paper have been circulated. See Hanushek, Eric A., John F. Kain, and Steven G. Rivkin, "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement," National Bureau of Economic Research, Working Paper 8741, revised in February 2004. The first version of this paper was published in 2002, and a more recent revision is dated March 2006. We will rely on the 2004 and 2006 versions in this paper

This paper replicates the HKR models using data from two state testing programs, North and South Carolina, both of which have data quite similar to the Texas data. It also tests the HKR model using the ECLS data.

Theories of Black Peer Effects

Although the main thrust of this paper is empirical, it might be useful to review the major theoretical perspectives behind the study of black peer effects. Three major theories attempt to explain why school segregation, or schools that are predominantly black, have adverse effects on black achievement and, conversely why desegregated schools should improve black academic achievement. The first is self-esteem theory, a second is educational inputs theory, and a third is peer group theory. A fourth theory attempts to explain why school composition should have limited effects on achievement.

Self-esteem theory postulates that school segregation creates a "stigma" for black children which affects their self-image and motivation to succeed. This is the theory behind the "psychological harm" thesis in the original *Brown v Board of Education* decision.[4] This theory was supported by some early social science research, in particular the famous "doll" studies of Kenneth and Mamie Clark (Clark, 1939). However, most research conducted after 1960 found little evidence for the theory, and indeed numerous self-esteem studies found that black children had higher self-esteem than white children and also that black children in segregated schools had higher self-esteem than black children in desegregated schools (Armor, 1995; Schofield, 1995).

The second theory is based on educational input-output models drawn from economics. In this case, input-output theory suggests that greater school inputs, such as more funds, higher teacher quality, and reduced class sizes will produce higher outputs in the form of student achievement. During the 1950's and 60's, many social scientists and educators believed that school resources were deficient in segregated black schools, thereby explaining the lower achievement of black students. Since desegregation would put black and white children in the same schools, school resources would be more equitable and therefore the achievement gap would diminish. The controversial Coleman Report challenged this assumption, finding that by the mid-1960s school resources were not that different between predominantly black and white schools, and, moreover, that school resources were only weakly related to student achievement (Coleman, et al, 1966). Although the validity of Coleman's conclusions were upheld in subsequent analyses of his data (Jencks, 1972; Mosteller & Moynihan, 1972), the issue continues to be debated among educational researchers (Hanushek, 1996; Hedges, Laine, & Greenwald 1994; Grissmer & Williamson, 1998; Rothstein, 1995). The current debate is not so much on whether school

[4] "To separate [black children] from others of similar age and qualifications solely because of their race generates a feeling of inferiority as to their status in the community that may affect their hears and minds in a way unlikely to ever be undone…" *Brown v. Board of Education*, 347 U.S. 483 (1954).

resources have an effect on academic achievement, but rather how large the effect is and whether equitable resources alone can solve the problem of low black achievement.

Perhaps the most relevant theory for this study is peer group theory itself. Classical peer group theory postulates that schools with a higher proportion of higher achieving middle-class students will determine the teaching standards of schools and classrooms, thereby improving the performance of lower achieving students (Armor, 2006). Although classical peer group theory does not require racial desegregation per se, just the presence of a sufficient number of middle class students, as a practical matter the high correlation between race and socioeconomic status implies that SES integration cannot be accomplished without racial integration.

A different type of peer group theory is black oppositional culture (Fordham and Ogbu, 1986). Their argument suggests that black students face significant pressure from their black peers to reject academic success because such success equates to giving up African-American culture and 'acting white.' Therefore, a desegregated school can reduce this negative effect by replacing black peers with white peers. While this theory has been embraced by some educators (e.g., Rothstein 2004), oppositional culture theory has not been supported by several empirical studies (Cook & Ludwig 1998; Ainsworth-Darnell & Downy 1998).

The fourth theory explains why desegregated schools or reduction in black peers may not improve black achievement. Family background theory posits that academic achievement is determined primarily by a series of parent characteristics, and therefore school characteristics—whether racial composition or school resources—has minimal impact on black achievement. These family factors, including parent IQ, parent education and income, family structure and size, birth weight and nutrition, and parenting behavior such as cognitive stimulation operate during infancy and explains the school readiness gap, but they continue to operate throughout the early school years and can explain a great deal of the achievement gap between black and white children (Brooks-Gunn, Klebanov, and Duncan, 1995; Armor, 2003; Brooks-Gunn and Markman, 2005). When comparing black students in segregated and desegregated schools, this theory underscores the need to control for as many family background characteristics as possible, since blacks in segregated schools may come from more disadvantaged families than blacks in desegregated schools.

Models for Assessing Peer Effects

Since the early versions of the HKR paper were circulated, a number of studies have appeared with a variety of different models for analyzing peer effects. Most of these newer models estimate peer effects at the classroom level, and they also place substantial emphasis on estimating the effect of peer achievement levels along with peer race. Indeed, it may well be that the reason for a negative black peer effect is not because of race but because black students on average have lower achievement than white students. Thus most studies of peer effects, including HKR, include average peer achievement along with peer racial composition in their models.

To date, other studies of peer effects have used either the North Carolina or the Florida state testing data. Using North Carolina data, Vigdor and Nechyba (2004) estimated classroom- and school-level peer effects for race and achievement levels; they find a classroom-level effect for peer achievement, but they do not find consistent effects for racial composition. Another study of achievement in Wake County, North Carolina, developed nonlinear models for peer effects at the classroom level (Hoxby and Weingarth, 2005); this study also finds weak effects for racial composition at the classroom level when peer achievement is properly accounted for. Finally, another study using North Carolina data finds a small effect of desegregation on closing the black-white gap, although only by .06 standard deviations (Cooley, 2006). A study of Florida state testing data also estimated peer effects at the classroom level, and it found “no effects of percent black students after controlling for student, school, and teacher effects” (Burke and Sass, 2004).

Thus other studies using North Carolina and Florida state testing data have failed to confirm a large negative effect arising from concentrations of black students, thereby rendering the HKR study even more unique. However, none of these studies are true replications because they all impose substantially different models and assumptions than those adopted by HKR. In particular, they all estimate peer effects at the classroom level rather than the school and grade level used by HKR. Because the Texas state test data do not include classroom identification, HKR have not examined peer effects at the classroom level. Since this study is primarily a replication of the HKR study of Texas, we confine our estimation of peer effects at the school and grade level rather than at the classroom level.

Moreover, we believe that the HKR approach is more relevant to the issue of school desegregation policy. Both courts and legislative bodies have generally distinguished between racial composition at school and classroom levels, and most school desegregation policies and court rulings focus on school rather than classroom composition. While absence of a black peer effect in desegregated schools might be caused by resegregation at the classroom level, the HKR study found a large black peer effect at the school and grade level, thereby rendering classroom racial composition moot for Texas. Of course, if the HKR result is not replicated in other states, then it would be appropriate to investigate classroom resegregation as a possible explanation for lack of a black peer effect at the school level.

The original HKR paper formulated a longitudinal model of achievement which postulated current achievement as a function of current student, school, and peer characteristics as well as the prior year’s achievement, as represented by equations (1) and (2) below (HKR, 2002, 2004). For child i in school s and grade g, A_{isg} is achievement, X is a vector of individual student characteristics (e.g., family socioeconomic status), S is school and teacher characteristics, P is average peer characteristics, and e is an error term.

$$A_{is(g-1)} = X_{is(g-1)}B + S_{s(g-1)}C + P_{s(g-1)}D + A_{is(g-2)} + u_{is(g-1)} \qquad (1)$$

$$A_{isg} = X_{isg}B + S_{sg}C + P_{sg}D + A_{is(g-1)} + u_{isg} \qquad (2)$$

By taking the first differences of the achievement equations for two consecutive years, HKR arrive at a gain-score (or value-added) model described in equation (3) below.

$$A_{isg} - A_{is(g-1)} = \Delta A_{isg} = X_{isg}B + S_{sg}C + P_{sg}D + u_{isg} \qquad (3) \quad \text{Gain model}$$

The residual term u is further decomposed according to several fixed effect terms which we will discuss later. In their most recent revision, HKR propose an additional lagged achievement model which becomes their preferred approach (HKR, 2006). In their new formulation, the coefficient for the prior year achievement $A_{is(g-1)}$ is not assumed to be 1, so that we have instead

$$A_{isg} = X_{isg}B^* + S_{sg}C^* + P_{sg}D^* + \theta A_{is(g-1)} + u_{isg} \qquad (4) \quad \text{Lagged model}$$

Coefficients for student (B and B*), school (C and C*), and peer effects (D and D*) are assumed to be constant from one grade or year to the next (e.g., parent SES exerts the same effect on annual achievement gains in grade 3 as in grade 4, etc.) in this model. However, student, school, and peer characteristics themselves can (and do) change from year to year, which would give rise to changes in achievement according to the model. It should be noted that the Texas analysis includes three cohorts of students, and our North and South Carolina analysis uses four or five cohorts.[5] Their equations show a superscript c for cohort which is omitted here for ease of notation.

The extent to which coefficients B, C, and D differ between (3) and (4) will depend on the size of θ. If θ is close to 1, which is often the case with achievement test scores measured one year apart, then the coefficients may not differ appreciably. Peer effects are distinguished for average racial composition (% black, % Hispanic), average poverty level, and average achievement level of a given grade within a school.

In addition to this basic functional form, HKR attempt to reduce error variance by controlling for several types of fixed effects. The Texas data has student test scores over several grades (4 to 7), several years (1993 to 1997), thousands of schools, and hundreds of thousands of students. By "stacking" gain or lagged test scores, where an observation is a single gain score or a single level score and a lagged score within in a given grade,

[5] By cohort, we mean a group of students who start a grade in a given year and then progress through later grades; e.g., one cohort would be students who start 3rd grade in 2001; a second cohort would be students who start 3rd grade in 2002.

school, and cohort, they can estimate fixed effects for various combinations of students, grades, schools, attendance zones, and years or cohorts. The analytic advantage of fixed effect models is removing unmeasured differences among students, schools, and attendance zones that are constant over the period of study. Specifically, four fixed effect combinations are estimated in most of their 2004 and 2006 models: student, year by grade, school by grade, and attendance zone by year.

In a more recent but quite different paper on peer effects, Hanushek and Rivkin have changed to aggregate models whereby scores on all individual student measures (test scores, socioeconomic status, etc.) are averaged for each school by grade by year group, separately for each race, and then the coefficients B, C, D and θ are estimated using the aggregated form of the data (Hanushek and Rivkin, 2007). The reason for shifting to an aggregate analysis has to do with computational problems when trying to estimate both student and school by grade fixed effects simultaneously, an issue discussed in the next section.

In switching to an aggregate model, Hanushek and Rivkin have dropped estimation of student fixed effects. However, since all of the other fixed effect factors (year by grade, school by grade, attendance zone by year) can be estimated with the individual student data, the advantages offered by aggregation is unclear. Aggregation has several analytic drawbacks, not the least of which is a problem caused by students who repeat one or more grades over time.[6] Moreover, retaining the individual student data allows estimation of student fixed effects, albeit not simultaneously with school by grade fixed effects.

Review of the Texas Results

A summary of key results for the original models estimated by HKR is provided in Table 1. The figures are adapted from Table 1 of their 2004 and 2006 versions of the paper, as indicated. The coefficients in their papers are scaled to reflect the proportion of black peers. For ease of interpretation and also comparison with our replication results, we have changed the scale of the coefficients in Table 1 to reflect a 10% change in percent black rather than a unit change in the proportion black (in the HKR papers all test scores are standardized to mean 0 and variance 1 for each grade and year, and black peers are measured as a proportion). Thus a coefficients should be read as the change in achievement, in sd's, for a 10 percentage point increase in black peers .[7]

[6] Students who repeat grades—about 5% of black students and 2% of white students—have to be eliminated when aggregating by year-grade-school groups, because they create extra year-school-grade groups with very low test scores. But they can be included and identified in the individual student analysis.

[7] It is unlikely that a black student would experience a 100-point change in percentage of black peers in one year, even in the most comprehensive desegregation plan. In Southern school districts during the early 1970s, a change of more than 60 or 70 percentage points was rare.

Table 1 Effect of School Percent Black on Black Math Achievement in Texas (HKR papers)

Dep. Variable	Model	Effect[a]	Sig.	Fixed effect and other controls[b]
2004 Paper				
Gain scores	(1)	-.007	**	None
Gain scores	(2)	+.029	***	Student fixed effects
Gain scores	(3)	-.030	**	Student & school x grade fixed effects
Gain scores	(4)	-.031	**	Student, school x grade, & att. zones x year fixed effects
2006 Paper				
Lagged scores	(5)	-.009	**	None
Lagged scores	(6)	-.021		Student & school x grade fixed effects (from Table 2)
Lagged scores	(7)	-.023	**	School x grade & att. zones fixed effects
Lagged scores	(8)	-.022	**	Fixed effects from model (3) plus school & teacher characteristics
Lagged scores	(9)	-.024	**	All controls from model (4) plus average peer achievement (twice lagged)

Sources: Hanushek, Kain, and Rivkin, 2004 & 2006, Table 1. * p≤.05 ** p≤.01 *** p≤.001

[a] Expected change in math or gain scores (in standard deviations), given a 10% increase in percentage black students in a given year.

[b] In addition to the controls listed, all models include a full set of grade-by-year dummies, indicators for school changes (other than to middle school), and indicators for free lunch eligibility.

There are several surprising features in these results. First, the gain score models show unusual and very large reversals in the signs of the coefficients as fixed effect controls are added. Comparing gain models (1) to (3), the coefficient changes from a modest negative effect (-.007) without controls to a much stronger positive effect (+.029) after controlling for student fixed effects, then it reverses to a strong negative effect (-.030) after adding school by grade fixed effects. Note that adding attendance zones does not have an appreciable effect in the gain score models.

The coefficients for the lagged achievement model do not show such large swings; in fact, considering the gain models, they are somewhat surprising for their lack of change. For example, there is little difference between removing student and school by grade only (-.021) and removing school by grade and attendance zone by grade only (-.023). There is also very little change after removing average peer achievement (-.024). This lack of a peer achievement effect is inconsistent with most of the research on this issue, as reported above, where addition of peer achievement usually reduces the black peer effect to a significant degree. Note, also, that addition of school and teacher controls has very little effect (-.022).

The size of the effect in these models is also surprising. An effect of -.030 implies that a 50 point reduction in school percent black—not uncommon in a comprehensive desegregation plan—would raise black math scores by .15 standard deviations in a single year. In their discussion of impact, HKR suggest this effect is cumulative over grades, in which case the impact would be multiplied by the number of grades or years that the desegregation lasted. If this reduction lasted five years, and assuming that white achievement does not change, then black achievement would rise by .75 standard deviations and the black-white achievement gap would be virtually eliminated.[8] It is not clear, however, that these annual effects are in fact cumulative; we shall return to a discussion of cumulative effects after presenting the replication results.

During the 1970s and 80s many school systems experienced substantial reductions in segregation lasting for many years, including many Texas school districts. For example, in Charlotte Mecklenburg, North Carolina, black students experienced an average reduction of more than 35 percent black classmates lasting for more than ten years, yet the achievement gap changed very little (Figure 1).[9] Appendix Figure A-1 shows that the initial change in black exposure to whites was more than 40 points .

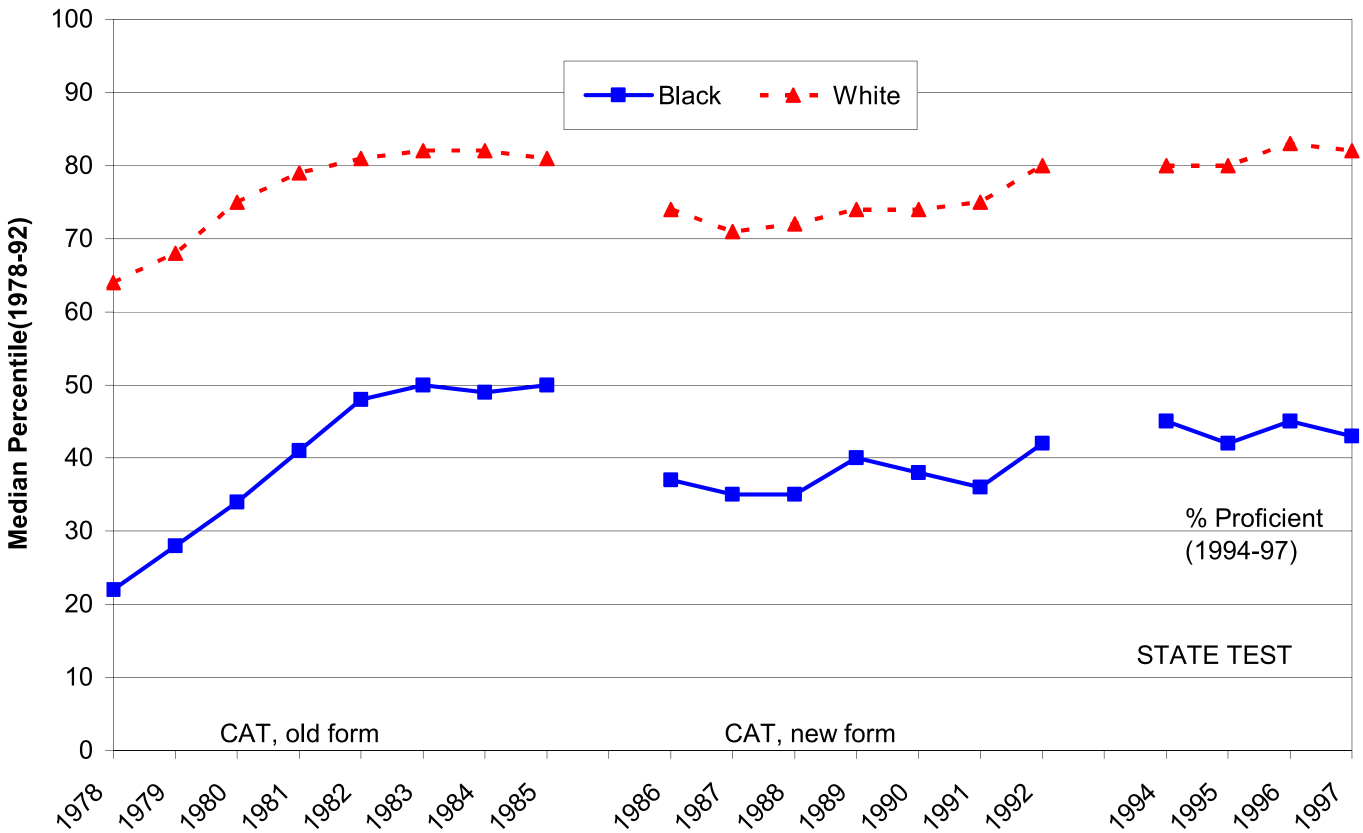

Figure 1 CHARLOTTE MECKLENBURG 6TH GRADE ACHIEVEMENT BY RACE

---

[8] HKR found no black peer effect on white achievement. HKR also report a black-white math gap in Texas of .7 standard deviations, and according to NAEP data, the black-white gap in Texas is about .8 standard deviations in 8th grade math.

[9] Adapted from Armor (2002) Chapter 6, figure 6.11. This chapter also describes case studies of Wilmington-New Castle County, Delaware; Dallas, Texas; Kansas City, Missouri; and Minneapolis, Minnesota, which failed to find significant black achievement gains following extensive, long-term desegregation (figures 6.6-6.13).

Finally, it should be noted that the all of coefficients in Table 1 that include fixed effect controls (that is, all but models (1) and (5)) are probably incorrect because of computational problems. Basically, the HKR procedures for estimating two or more large fixed effect variables simultaneously (e.g., student, school by grade, and attendance zone by grade) used inappropriate software commands.[10] The resulting coefficients are unstable and can be misestimated by substantial magnitudes, most likely causing the unusual sign-reversals observed in Table 1. HKR are now using aggregate models that disregard potential student fixed effects, although aggregate models can raise new data analysis problems as described above.

The Replication Data

Most states have implemented statewide achievement testing systems in compliance with No Child Left Behind requirements. In 1996 North Carolina was one of the first states to adopt an accountability system utilizing extensive achievement testing (called "End-of-Grade" tests), and consequently their state testing program is one of the most extensive and longest-running in the nation. The state also created a comprehensive testing data base, residing at the Duke University Education Research Data Center, that is available to researchers. We have drawn from that data base for this study.

South Carolina was not far behind its neighbor state; in 1998 it implemented a comprehensive statewide testing program called the Palmetto Achievement Challenge Tests (PACT). Data appropriate for the analyses in this study was made available by the South Carolina Department of Education. Finally, we have used similar longitudinal testing data from the federal ECLS project.

For the North Carolina replication, data for students includes reading and math test scores, grade, student race/ethnicity, free lunch status (a poverty indicator), and parent education. Data for each school consists of eight different measures of school resources and teacher characteristics. For the South Carolina replication, the student data includes the same characteristics as North Carolina except parent education; thus free lunch status is the only measure of student socioeconomic status. The school data is much richer, however, so we have data on 15 school and teacher characteristics. For both states, the main independent variable is the percentage of black students in each grade level at each school.

[10] We are grateful to Steve Rivkin for explaining how the fixed effect regressions were computed using Stata software. During these discussions he acknowledged that the procedure used for removing multiple fixed effect variables was incorrect. Most desk top computers do not have sufficient memory to hold the individual student data while simultaneously inverting a matrix of several thousand variables for the school by grade indicators. While a single large fixed effect variable such as students can be removed using the Stata **areg** command, **areg** cannot be used to remove two large fixed effect variables (students and school by grade). One can remove two large fixed effect factors using **xtreg** with double demeaning, but this procedure yields the fully interacted fixed effects for student by school by grade, which produces as many indicators as there are observations. Using a desk top computer, therefore, one can remove either student fixed effects or school by grade fixed effects but not both simultaneously.

For both North and South Carolina, the data include identifiers for district, school, grade, and student, making it possible to construct longitudinal files. Like the Texas data, longitudinal testing data is available for all students who remain in these states, even if they change schools or school districts. The school identifier allows creation of variables to indicate whether a student changes schools and whether that change reflects mobility or the normal transition from elementary to middle school. Most North and South Carolina school districts have K-5/ 6-8 grade structure, so most students make the transition from elementary to middle school from grade 5 to grade 6. In the South Carolina data we have also identified students who repeat a grade.

For North Carolina, we constructed four cohorts of students who attended grades 3 to 8 between 1997 and 2005. The first cohort attended grade 3 in 1997 and grade 8 in 2002; the last cohort attended grade 3 in 2000 and grade 8 in 2005. Each cohort consists of about 100,000 students, about 80,000 of which are the same students, so the total North Carolina analysis sample consists of nearly 500,000 different students. The South Carolina data has five cohorts who attended grades 3 to 8 between 2001 and 2006. Given the organization of the available data, each cohort did not necessarily attend all grades. One cohort attended grades 3 to 7 from 2001 to 2005, another cohort attended grades 4 to 8 from 2002 to 2006, two other cohorts attended four grades (5th to 8th and 3rd to 7th), and one cohort attended grades 3 to 5. Each cohort consists of about 45,000 students, and the total South Carolina sample is nearly 230,000 different students. [11]

The data from both states is organized into longitudinal and stacked test score formats. In stacked data, an observation is a single score from a given grade and year (t) paired with its lagged score (t-1); gain scores can also be computed for each observation. Thus a single student who has taken the state tests every year contributes up to five or six observations. Each observation also has the associated student and school characteristics for that grade, school, and year.

The structure of the ECLS data differs from the state data primarily in the lack of multiple cohorts. The ECLS data consist of a single cohort of students who started Kindergarten in 1998-99 and were tested in both the Fall and Spring of that year. That cohort was then followed and tested in the Spring of the 1st , 3rd, and 5th grades in 2000, 2002, and 2004. One advantage of ECLS data is that a large number of family background measures are available from parent interviews, usually a limitation in state data bases. Although the original cohort started with about 15,000 Kindergarten students, only about 10,000 of the original cohort have testing data from the 5th grade administration. The ECLS data is also stacked, and the lagged test score is t-1 for Spring of Kindergarten and first grades, and t-2 for Spring of third and fifth grades. Only those student and teacher variables that are available in the Spring of each wave are used in the analysis, in order to have complete longitudinal data for each student.

---

[11] For both states, these student counts are those that have test scores in year t and also a lag score in t-1; since some students (e.g., 3rd graders) do not have a prior test score, the total number of students is about 15% higher.

The North Carolina test scores are standardized to have a mean of 250 and a standard deviation of 10, which is similar to the scale scores they normally use. The South Carolina scores are standardized to have a mean of 100 and a standard deviation of 10. The standard deviation of 10 is chosen in order to match the metric we use for the ECLS data, t-scores, which have means of 50 and standard deviations of 10. Thus the coefficients for % black peers must be divided by 10 to obtain a standardized effect in standard deviation units.

Descriptive statistics for the final variables used in the analyses of all three sets of data are shown in Appendix Table A-1. More descriptive information about the specific variables used is in the results section.

Replication Methods

Our replication analyses follow the HKR linear fixed regression techniques for most models. In the case of individual student lag models with student fixed effects, which creates an endogenuity problem, the analysis uses two-stage least squares with twice-lagged test scores as the instrument (following the procedure adopted by HKR in their 2006 paper). Similarly, our measure for average peer achievement is based on twice-lagged achievement, which is averaged over each school and grade combination.

The state replications also use various combinations of fixed effect indicators used by HKR, at least those that can be estimated within computational constraints. We include year by grade and school by grade in all of our fixed effect models, which are clearly the most important fixed effect variables in the HKR papers. Year by grade indicators are needed to remove idiosyncratic variations in testing procedures from year to year as well as systematic variations in test content as students progress through the grades. School by grade indicators are needed to remove any number of unmeasured school characteristics, which is especially important since racial composition is also measured at the school and grade level. In an earlier version of this paper, we also removed attendance zone by year fixed effects and student fixed effects (separately) for the North Carolina data; these results will be mentioned in the North Carolina analysis.

The ECLS analysis is identical to the state analyses, with the exception that, since only a single cohort is available, school by grade fixed effects cannot be estimated. Moreover, since year is totally confounded with grade level, the ECLS fixed effect analysis consists of indicators for each of the five waves (which corresponds to both year and grade).

Inclusion of certain variables leads to a reduction of observations, most notably for average peer achievement which loses cases because it is computed for twice-lagged achievement (e.g., we cannot compute peer achievement for either 3rd or 4th graders because there are no 1st or 2nd grade test scores). Models that incorporate a test for average peer achievement have about 15% fewer cases. Given the nature of the ECLS data, we do not compute average peer achievement for these students.

Finally, since the state data includes all students in a given school, and since we cannot assume that errors are uncorrelated for students within schools, we also use Huber/White robust standard errors using schools as the clusters.

Results of the Replication Analyses

We first present a series of regression models without fixed effects for each of the three sets of data. These initial analyses are instructive for showing how peer race effects are modified by successive controls for student background, school and teacher characteristics, and peer achievement. We then present a summary table showing how peer race effects are further modified by removing fixed effects for year by grade and school by grade indicators.

North Carolina. Table 2 shows the results of five models separately by race and test content for the North Carolina data. The model starts with a simple bivariate regression between achievement at time t and % black peers. We then add, successively, student socioeconomic controls, lagged achievement (t-1), school and teacher characteristics, and finally average peer achievement.

Starting with black math achievement in model (1), there is a moderate relationship of nearly -.03 between black peers and black math achievement in any given year. This means that black students in 75% black schools typically score about 1.5 points lower than blacks in 25% black school, or .15 sd's. Note that this uncontrolled relationship is the same magnitude as the HKR gain score effect in Table 1, model (3), <u>after</u> they have removed most of the control variables.

After controlling for socioeconomic status and school stability (model 2), which have large and significant coefficients, the black peer effect drops sharply to -.02. This demonstrates that black students in predominantly black schools are more disadvantaged economically than their counterparts in desegregated schools, thereby explaining some of the original negative correlation. Model (3) shows that a student's prior year test score (lagged achievement) is a powerful predictor of the current test score, explaining more than fifty percent of the total variation in black math scores. The black peer coefficient drops to -.009. Note, also, that adding lagged achievement weakens student SES effects, demonstrating that prior year achievement embodies many aspects of family background.

Model (4) adds school and teacher variables that, in separate analyses, had the strongest impacts on test scores, but only % certified teachers retains a statistically significant impact on achievement growth after controlling for student background and lagged achievement.[12] Addition of school characteristics reduces the black peer effect to only -.005, which, while significant, means that a 50-point increase in black composition (ceteris paribus) is associated with a decrease in math scores of only one-fourth of a point. The magnitude of the

[12] In addition to the two variables shown, % high quality teachers also had a significant relationship with achievement, but it was not available for all years and caused a reduction in the number of observations.

Texas black peer coefficients in the HKR lagged models (models 6 to 9) are more than four times the magnitude of this North Carolina coefficient, even before controlling for various fixed effect factors.

Table 2 LAGGED ACHIEVEMENT MODELS FOR NORTH CAROLINA STUDENTS WITHOUT FIXED EFFECTS

Student, School, &	Black Students					White Students				
Teacher Characteristics	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
MATH PROFICIENCY										
% Black Peers	**-0.029**	**-0.019**	**-0.009**	**-0.005**	-0.002	**-0.012**	**-0.015**	**-0.007**	**-0.005**	0.000
Lagged Math Score (t-1)			**0.759**	**0.758**	**0.764**			**0.815**	**0.815**	**0.819**
Average Peer Math (t-2)					**0.043**					**0.054**
% Free Lunch		**-0.022**	**-0.005**	**-0.006**	**-0.005**		**-0.032**	**-0.007**	**-0.007**	**-0.007**
Parent Education (yrs)		**1.007**	**0.255**	**0.259**	**0.234**		**1.525**	**0.336**	**0.335**	**0.295**
Same school, t & t-1		**1.190**	**0.324**	**0.319**	**0.303**		**1.109**	**0.230**	**0.228**	**0.202**
% Certified Teachers				**0.015**	**0.017**				**0.012**	**0.012**
Pupils Ter Teacher				0.014	0.008				**0.020**	0.019
Constant	246.2	230.5	55.5	54.2	42.2	253.1	228.8	41.5	40.3	26.3
R2 / Adjusted R2	0.006	0.112	0.641	0.642	0.651	0.001	0.211	0.723	0.723	0.740
N Schools	1817	1793	1783	1783	1695	1882	1878	1872	1872	1776
N Observations	585660	549493	507317	507317	420959	1217413	1194155	1126811	1126811	895673
READING PROFICIENCY										
% Black Peers	**-0.031**	**-0.019**	**-0.007**	**-0.005**	**0.003**	-0.002	**-0.006**	**-0.002**	0.000	**0.007**
Lagged Reading Score (t-1)			**0.752**	**0.751**	**0.751**			**0.764**	**0.764**	**0.758**
Average Peer Reading(t-1)					**0.085**					**0.090**
% Free Lunch		**-0.028**	**-0.007**	**-0.007**	**-0.007**		**-0.032**	**-0.008**	**-0.008**	**-0.008**
Parent Education (yrs)		**1.129**	**0.300**	**0.302**	**0.274**		**1.421**	**0.348**	**0.347**	**0.300**
Same school, t & t-1		**1.168**	**0.327**	**0.325**	**0.333**		**0.966**	**0.270**	**0.267**	**0.276**
% Certified Teachers				**0.008**	**0.008**			**54.052**	**0.009**	**0.007**
Pupils Ter Teacher				**0.025**	0.009				0.002	0.004
Constant	246.6	229.7	56.8	55.9	34.9	252.8	230.5	54.1	53.4	32.8
R^2 / Adjusted R^2	.006	.126	.615	.616	.612	<.001	.200	.664	.664	.664
N Schools	1816	1792	1782	1722	1695	1880	1877	1870	1803	1775
N Observations	582234	546430	503381	503381	417942	1213864	1190812	1121758	1107484	891716

Coefficients in bold are significant at p≤.05 (robust standard errors using N schools as clusters)

Finally, after controlling for average peer math achievement in model (5), the black peer effect drops to -.002 and becomes statistically insignificant. Again, this result is consistent with other analyses of the North Carolina data (e.g., Hoxby, 2005), but it is not consistent with the Texas study, which found no significant impact of peer achievement.

Turning to black reading scores, the results closely follow those for math scores. The bivariate coefficient for black peers is nearly identical (-.031 vs. -,029), and after controlling for student, school, and teacher characteristics (model 4) the black peer coefficient for reading is identical to that for math (-.005). Like math scores, the black peer coefficient also becomes non-significant after controlling for peer reading achievement.

Generally speaking, the black peer effect is smaller for both white math and white reading scores; this is consistent with the Texas study which found no significant black or Hispanic peer effects for white math scores. While the black peer effect for white math is the same as black math in model (4), the coefficient is 0 for white reading. After controlling for peer achievement in model (5), the coefficient for white math is 0 and actually positive (.007) for white reading.

In an earlier version of this paper, fixed effects were removed for attendance zones and also for students (separately). Attendance zones did not change the black peer coefficients, and controlling for student fixed effects (but not school by grade) did not result in statistically significant black peer coefficients.

South Carolina. The South Carolina basic regressions are shown in Table 3. Interestingly, the bivariate black peer effect for black math is identical to the corresponding North Carolina coefficient (-.029).

Table 3 LAGGED ACHIEVEMENT MODELS FOR SOUTH CAROLINA STUDENTS WITHOUT FIXED EFFECTS

Student, School, & Teacher Characteristics	Black Students (1)	(2)	(3)	(4)	(5)	White Students (1)	(2)	(3)	(4)	(5)
MATH PROFICIENCY										
% Black Peers	**-0.029**	**-0.020**	**-0.009**	**-0.009**	**-0.007**	**-0.036**	**-0.029**	-0.003	-0.001	0.004
Lagged Math Score (t-1)			**0.697**	**0.696**	**0.702**			**0.778**	**0.777**	**0.786**
Average Peer Math (t-2)					**0.048**					**0.016**
% Free Lunch		**-0.036**	**-0.010**	**-0.010**	**-0.008**		**-0.058**	**-0.015**	**-0.014**	**-0.014**
Same school, t & t-1		**0.516**	0.251	0.224	0.208		**0.374**	0.117	0.085	0.117
Repeated Grade		**-4.147**	0.143	0.180	**-0.317**		**-6.233**	-0.054	-0.014	**-0.398**
Per Pupil Expend ($000)				**0.113**	**0.129**				**0.140**	**0.143**
% Prime Instruct. Time				**0.074**	**0.074**				**0.106**	**0.111**
Constant	97.1	99.4	30.3	23.2	17.5	104.3	105.5	23.5	13.3	10.0
R2 / Adjusted R2	0.006	0.046	0.534	0.534	0.551	0.007	0.095	0.633	0.634	0.647
N Schools	911	904	899	859	823	888	883	877	836	807
N Observations	258281	262494	252417	246130	194092	364024	361045	353503	345587	274205
READING PROFICIENCY										
% Black Peers	**-0.026**	**-0.012**	**-0.006**	**-0.006**	**-0.005**	**-0.037**	**-0.051**	**-0.009**	**-0.008**	-0.003
Lagged Reading Score (t-1)			**0.737**	**0.736**	**0.739**			**0.753**	**0.752**	**0.756**
Average Peer Reading (t-2)					**0.031**					**0.044**
% Free Lunch		**-0.044**	**-0.011**	**-0.011**	**-0.010**		**0.061**	**-0.016**	**-0.016**	**-0.016**
Same school, t & t-1		**0.801**	**0.661**	**0.669**	**0.772**		0.341	0.092	0.066	0.106
Repeated Grade		**-4.389**	**0.267**	**0.278**	-0.134		**-6.455**	**-0.377**	**-0.345**	**-0.709**
Per Pupil Expend ($000)				0.071	0.057				**0.151**	**0.161**
% Prime Instruct. Time				**0.056**	**0.058**				**0.074**	**0.083**
Constant	97.1	99.6	26.3	20.9	17.3	104.2	105.5	26.1	18.7	12.8
R^2 / Adjusted R^2	0.006	0.055	0.567	0.568	0.579	0.005	0.105	0.620	0.621	0.683
N Schools	912	905	900	859	823	888	884	878	835	805
N Observations	256056	260428	249324	243151	191593	364024	358746	350032	342196	271624

Coefficients in bold are significant at p≤.05 (robust standard errors using N schools as clusters)

More important, the reduction of the coefficient after adding student background factors and lagged achievement also parallels North Carolina results.

However, the black peer coefficient remains at -.009 after controlling for school and teacher characteristics (model 5). In preliminary analyses, the school characteristics with the strongest achievement correlations, prime instructional time and per pupil expenditures, were different than the strongest North Carolina. Prime instructional time is somewhat unique; it is the percentage of time when both regular teachers and students are in school. In any case, despite having significant effects, introducing these school measures does not alter the black peer coefficient. Finally, adding peer math achievement (model 5) does reduce the black peer effect to -.007, slightly larger than in North Carolina effect.

The results for black reading also resemble North Carolina, with the black peer effect decreasing to -.006 (vs. -.005) after removing student background, lagged achievement, and school factors. The black peer coefficient is also slightly larger (-.005 vs. -.003) after removing peer reading achievement. While the black peer effects are very small and not significant for white math in models (3) to (5), the black peer coefficients are actually larger for white reading than white math in model (4)(-.008 vs. -.006). After controlling for peer achievement, however, the coefficient falls to -.003 and is not significant.

ECLS. The final replication uses data from the Early Childhood Longitudinal Study, and the basic regression results are summarized in Table 3. Although the ECLS sample spans earlier grades than the state testing programs (K to 5 vs. 3 to 8), the bivariate black peer effect for black math scores, -.021, is the same order of magnitude as the North and South Carolina coefficients. It is possible that black peer effects are larger in later grades, particularly middle schools when students begin studying higher level math including algebra.

In any event, one advantage of the ECLS is a very rich array of socioeconomic and family characteristics measured in a parent survey, and model (2) reveals that the black peer coefficient decreases to -.016 and becomes non-significant for black math after controlling for these student background factors. Further, after adding lagged achievement, the coefficient drops close to 0. Lagged achievement also reduces the effects of all student background variables, and only parent education level remains significant. Finally, despite a full complement of teacher characteristics, none of them have significant effects on black math achievement after controlling for student background and lagged achievement.

Black reading scores present the same basic picture. Although the black peer coefficients are somewhat larger for black reading scores, they are not statistically significant for models (3) to (5). And, while the bivariate black peer relationship is quite large for both white math and white reading (-.05), the coefficients are small and not significant for models (2) to (5).

Fixed Effect Models. Up to this point, none of the regressions have included fixed effects. Table 5 shows fixed effect regressions for all three data sources. In North and South Carolina, the fixed effects are year by grade and school by grade; in ECLS, which is only a single cohort, only year (or grade) fixed effect indicators can be estimated. All of the other control variables used for model (4) in Tables 2 to 4 are included; the only control variable omitted is peer achievement.

Table 4 LAGGED ACHIEVEMENT MODELS FOR ECLS STUDENTS WITHOUT FIXED EFFECTS

Student, School, & Teacher Characteristics	Black Students (1)	Black Students (2)	Black Students (3)	Black Students (4)	White Students (1)	White Students (2)	White Students (3)	White Students (4)
MATH PROFICIENCY								
% Black Peers	**-0.021**	-0.016	-0.001	-0.001	-0.050	0.005	-0.006	-0.001
Lagged Math Score (t-1)			**0.753**	**0.781**			**0.778**	**0.764**
Parent Education Level		**0.998**	**0.466**	**0.474**		**1.158**	**0.287**	**0.267**
Income Level		**0.307**	0.066	0.071		**0.403**	**0.081**	0.052
Number of Siblings		**-0.403**	-0.042	0.030		-0.253	-0.009	-0.002
Age of Mother First Birth		**0.258**	0.019	-0.011		**0.243**	**0.050**	**0.051**
Teacher Experience (years)				-0.037				-0.008
Teacher Education Level				0.892				0.084
Teacher Certification				0.553				-0.014
Constant	**46.771**	**36.242**	**8.228**	**4.955**	**53.018**	**37.801**	**8.337**	**9.427**
R2 / Adjusted R2	0.006	0.136	0.623	0.651	0.004	0.162	0.668	0.649
N Schools	1064	936	926	918	5682	5482	5468	5429
N Observations	5303	4236	3300	2772	28333	26295	20824	18662
READING PROFICIENCY								
% Black Peers	**-0.023**	-0.011	-0.004	-0.005	**-0.052**	0.001	-0.011	-0.006
Lagged Reading Score (t-1)			**0.716**	**0.717**			**0.757**	**0.747**
Parent Education Level		**1.396**	**0.541**	**0.463**		**1.253**	**0.323**	**0.261**
Income Level		**0.276**	0.064	0.082		**0.423**	**0.093**	**0.120**
Number of Siblings		**-0.812**	-0.268	-0.065		**-0.638**	-0.114	-0.057
Age of Mother First Birth		**0.369**	**0.021**	0.032		**0.194**	0.034	0.028
Teacher Experience (years)				0.024				0.010
Teacher Education Level				0.555				**-0.474**
Teacher Certification				0.692				-0.121
Constant	**48.168**	**35.145**	**10.659**	**7.494**	**52.520**	**38.373**	**9.821**	11.573
R^2 / Adjusted R^2	0.006	0.183	0.596	0.613	0.005	0.163	0.644	0.628
N Schools	1064	936	926	916	5682	5482	5468	5428
N Observations	5285	4225	3277	2751	28309	26276	20796	18641

Coefficients in bold are significant at p≤.05 (robust standard errors using N schools as clusters

Considering black students first, of the six estimates for black peer effects, five range from 0 to -.006 and are not statistically significant. Only North Carolina reading shows a statistically significant black peer effect of

-.008. If we add peer achievement as a further control, the black peer effect for black reading drops to -.001 which is not significant.

Of the six estimates of black peer effects for white students, only North Carolina math shows a significant negative effect (also -.008). There is also a significant effect for South Carolina reading, but the effect is positive.

Table 5 FIXED EFFECT REGRESSIONS FOR NORTH & SOUTH CAROLINA AND ECLS*

Student & school	Black Students			White Students		
Characteristics	North Carolina	South Carolina	ECLS	North Carolina	South Carolina	ECLS
Math Achievement						
% Black Peers	-0.006	0.003	0.003	**-0.008**	0.002	-0.001
Lagged Math Score	**0.756**	**0.696**	**0.773**	**0.813**	**0.774**	**0.765**
Constant	56.6	28.6	3.6	42.0	21.7	9.4
R2 / Adjusted R2	0.664	0.563	0.654	0.740	0.654	0.649
N Schools	1723	859	918	1803	836	5429
N Observations	497706	246130	2772	1112486	345587	18662
Reading Achievement						
% Black Peers	**-0.008**	0.006	0.000	-0.004	**0.008**	-0.006
Lagged Reading Score	**0.747**	**0.738**	**0.711**	**0.760**	**0.742**	**0.746**
Constant	58.3	25.4	5.4	55.4	26.8	11.0
R2 / Adjusted R2	0.626	0.589	0.626	0.670	0.636	0.629
N Schools	1722	859	915	1803	835	5425
N Observations	493840	243151	2733	1107484	342196	18622

Coefficients in bold are significant at p≤.05 (robust standard errors using N schools as clusters

*Fixed effects for North and South Carolina are year by grade and school by grade; fixed effects for ECLS are year (or grade). In addition, regressions include all of the student, school, and teacher characteristics in shown in Tables 2-4, except peer achievement.

In summary, none of the analyses presented here find the magnitude of black peer effects that HKR found in their Texas study. Even for those models where statistically significant effects are found, the magnitude of these effects is substantially smaller than the Texas black peer effects, usually by three orders of magnitude. Before discussing some of the theoretical and policy implications of these findings, we address a question raised earlier about the extent to which the black peer effects in these lagged models are cumulative over time.

Are Black Peer Effects Cumulative?

Although most of the black peer effects found in our analyses are small and not statistically significant, there are several models (without some of the control variables) that show significant effects ranging from -.005 to -.008. A major question becomes whether these annual effects are cumulative. Even a small effect of -.005 for a one point change in percentage black could add up to a sizeable effect if it is cumulative over 5 to 10 years. For example, a 30 percent reduction in percent black peers, sustained over 10 years, could lead to an increase in black achievement by nearly .2 sd's if the -.005 effect is cumulative. To test this possibility, two different types of analyses are presented. The first is a longitudinal analysis of changes in test scores over time, and the second is a direct calculation of cumulative effects from the lagged regression models.

Longitudinal Analyses. Although the stacked testing data offers several analytic advantages, particularly if one wants to remove student or school fixed effects, they do not easily reveal how segregated and desegregated black students perform over time. Regression coefficients from the stacked test data simply reflect changes between any two consecutive years in the time period analyzed, and therefore the results do not reveal long-term trends or how black achievement actually changes over time in schools with differing racial compositions.

In order to obtain a clearer picture of changes over time for students in segregated vs. desegregated schools, we conducted longitudinal analyses that track average test scores from grade 3 to grade 8 for black students in schools of differing racial compositions (or Kindergarten to grade 5 for ECLS). First, because racial composition is related to black student socioeconomic status, we adjusted test scores for whatever SES measures were available in that data base.[13] Second, we computed the average racial composition of the schools that students attended over this six year period, and we collapsed the average into two categories: majority white (0-50% black) and majority black (51-100% black).[14] Finally, we plotted SES-adjusted test scores of the same black students as they move from grade 3 to 8 (or K to 5) according to average school composition.

This analysis allows us to compare black students who attend majority black schools from grade 3 to 8 with black students who attended majority white schools over the same period. Of course, some students could obtain an average of 40% black by attending 10% black schools for three years and 70% black schools for three years, which gives an average of 40%. But most students attend schools with similar compositions over grades 3 to 8, and this can be demonstrated by correlations.[15]

Figure 2 shows the longitudinal results for North Carolina black math scores. The results are consistent with the regression analyses, in that there is only a small difference in the trajectory of math scores for black students who attended majority white schools versus majority black schools. Those in majority white schools have SES-adjusted 3rd grade scores just above 248, they drop just below 248 during grades 4, 5, and 6, and then they increase during 7th and 8th grades to end up just about where they began in grade 3. Black students in

[13] We did this by regressing math or reading scores on SES characteristics at the student level, computing residuals, and then adding the population mean so the metric of the adjusted scores would be the same as the standardized actual scores.
[14] We initially categorized schools into 4 compositions but present two categories here for simplicity.
[15] In North Carolina, racial composition of schools in adjacent grades is correlated over .9 except for grades 5 and 6 when most students change from elementary to middle school; even here the correlation is .83. The racial composition correlation is .88 between grades 3 and 5 and .9 between grades 6 and 8. The correlation between racial composition in grade 3 and grade 8 is over .75.

majority black schools start at 248, also drop slightly in grades 4 and 5, then they drop a bit more in 6th grade so that they are about ½ point below their counterparts in majority white schools. They also show an increase in 7th and 8th grade, but they do not make up the ground they lost in the transition from grade 5 to 6, so by the 8th grade they score .7 points below black students in majority white schools. Since they started .2 below, this is a net loss of about ½ point (.05 sd's) over a five year period. It should be noted that since this analysis does not control for any school or teacher differences, some of this gap could arise from school differences that were shown to be significant in Table 2.

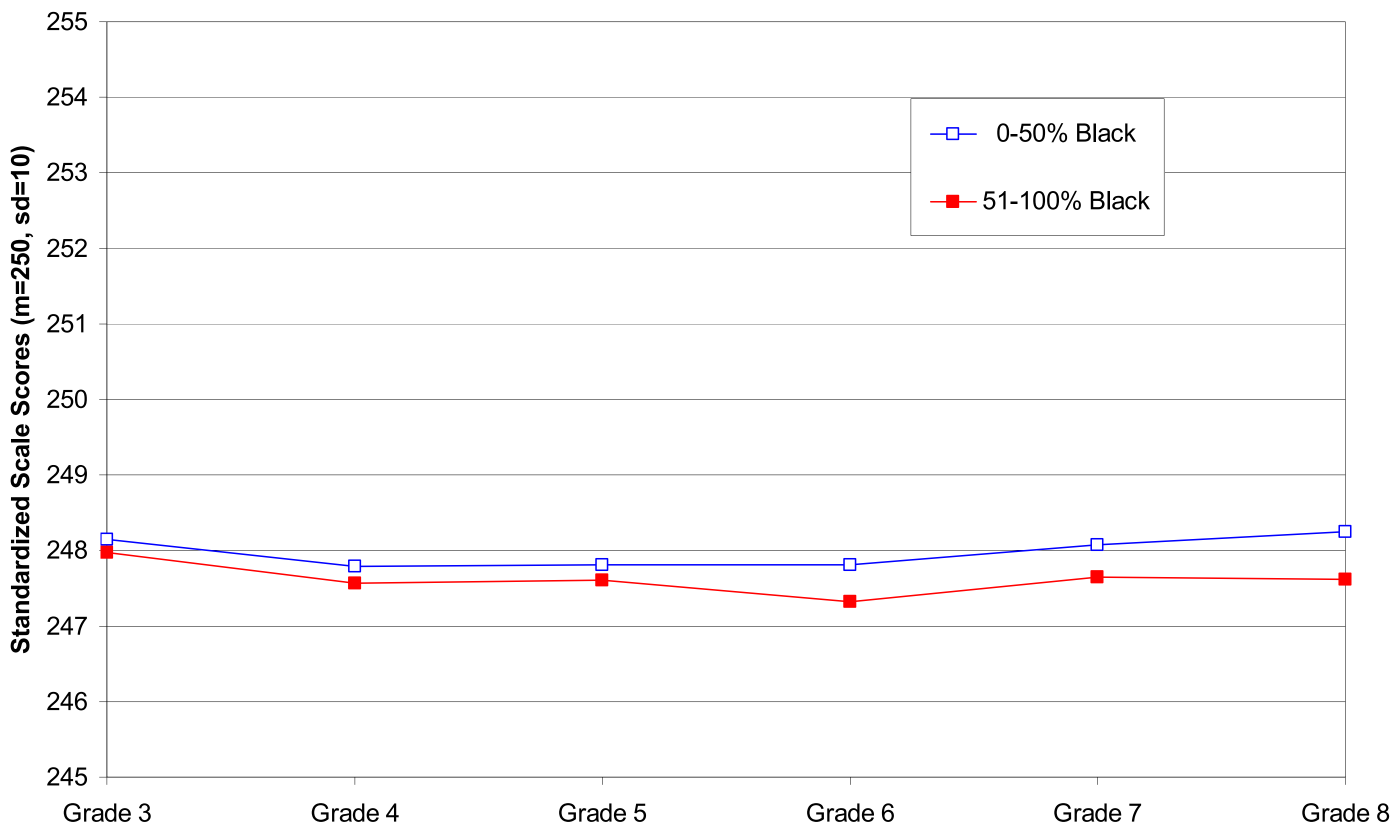

Figure 2 CHANGES IN BLACK MATH SCORES BY SCHOOL COMPOSITION, NORTH CAROLINA
(Adjusted for student SES)

Figure 3 shows a similar longitudinal analysis for black math scores in South Carolina. In this case black students in majority white schools start out slightly lower than those in majority black schools (both are slightly below 99) and they drop by nearly a point in the 4th and 5th grades. Their scores rise in the 6th grade, level off in the 7th grade, and drop again in the 8th grade to end just below 98. Black students in majority black schools follow approximately the same trajectory, except they drop a little more in the 4th and 5th grade, do not rise as much in the 6th grade, then fall sharply in the 7th grade but remain level in the 8th grade. In the 8th grade the blacks in majority black schools are ½ point below those in majority white schools, but since they started out .2 points above, the net change is -.7 points (or -.07 sd's) over the five year period. As for North Carolina, this longitudinal analysis does not control for school and teacher differences which might explain part of this small increase in the gap for students in majority black schools.

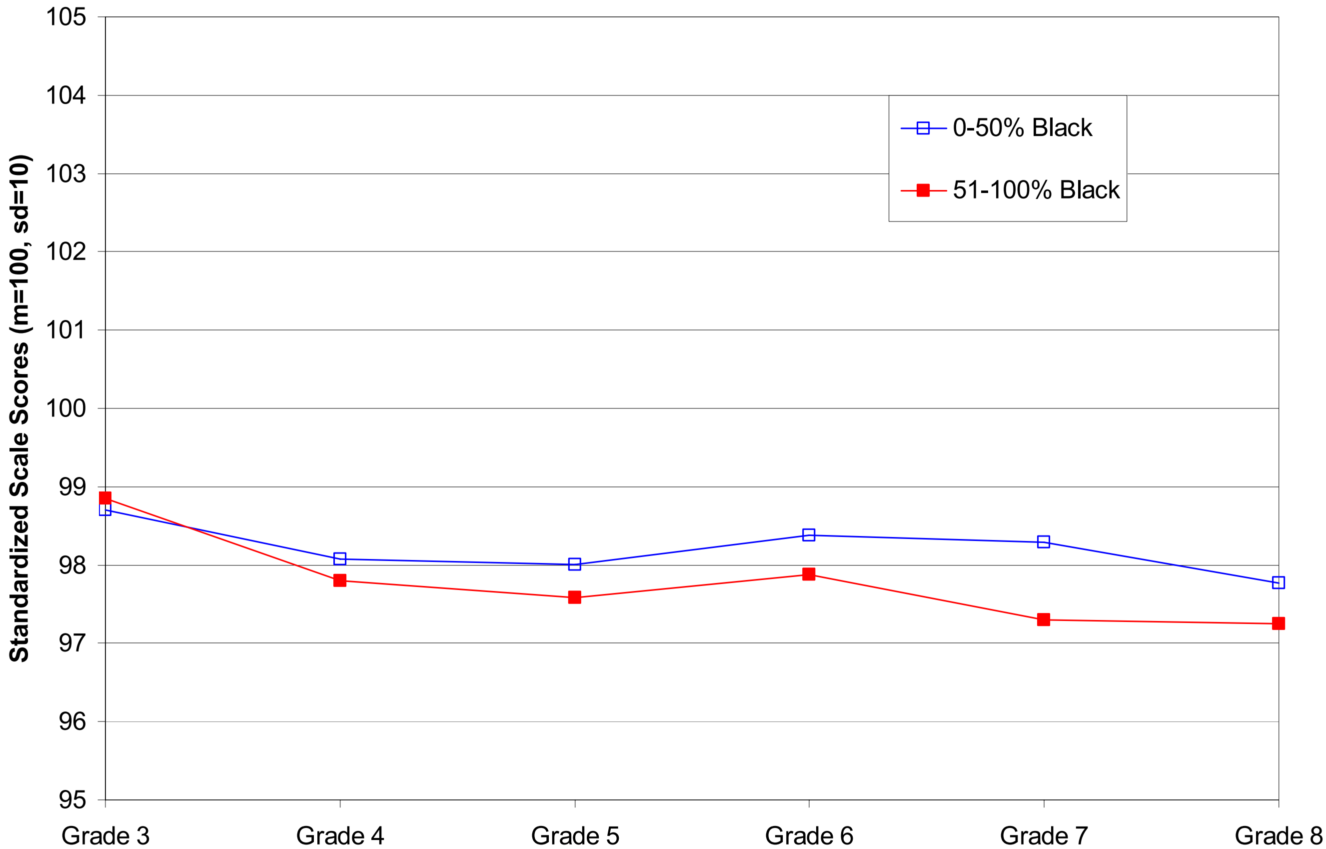

Figure 3 CHANGES IN BLACK MATH SCORES BY SCHOOL COMPOSITION, SOUTH CAROLINA (Adjusted for student SES)

Clearly, these longitudinal results do not support the notion that the small black peer effects found in the lagged regressions can be multiplied by the number of years students are in schools with a particular compositions. That is, if the coefficient is -.005 and a student experiences a 50 percentage point increase in school % black for five years, we would not multiply -.005 times 50 times 5 to get a 1.25 point decrease in scores. The average difference between majority black and majority white schools is about 50%, and yet Figures 2 and 3 show a net effect of only a fraction of a point over a five year period.

Before showing how to compute the expected change from the lagged regression models, we turn to a longitudinal analysis for the ECLS data.

Figure 4 shows the longitudinal changes for math scores (adjusted for SES) as black children advance from Kindergarten to 5th grade. Black children in majority black schools start out in Kindergarten 2 points behind those in majority white schools (for reasons that are not clear at this point), but by the first grade they catch up and from the 1st grade to the 5th grade there is virtually no difference between majority white and majority black schools. This graph is consistent with the regression results in Table 4 which shows black peer effects close to 0 for math scores.

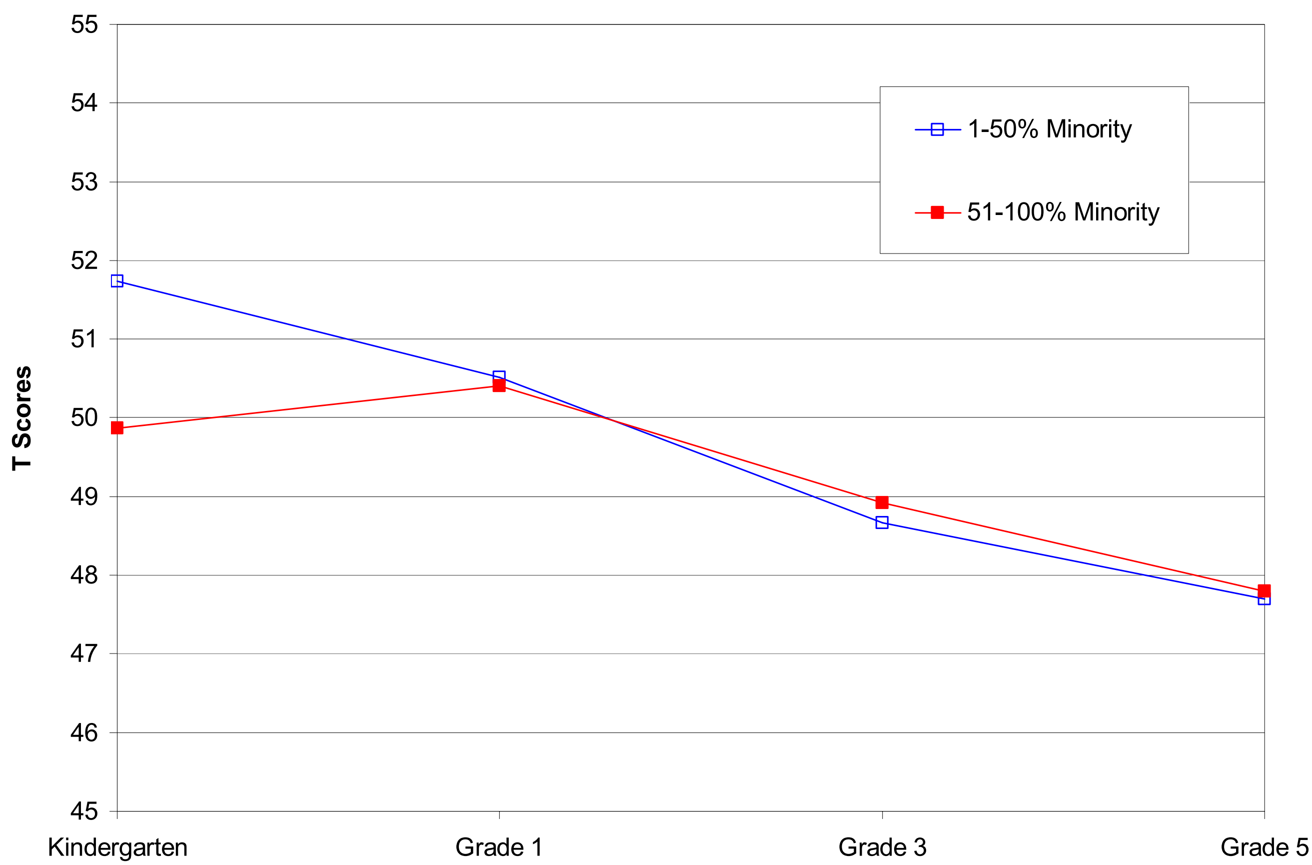

Figure 4 CHANGES IN BLACK MATH SCORES BY SCHOOL COMPOSITION, ECLS NATIONAL SAMPLE (Adjusted for student SES)

Although not shown in a chart, the results for reading are basically the same for the ECLS data, and for the North and South Carolina data, they show even less difference than the math changes. In North Carolina, blacks in majority black schools start out .2 points less in reading scores than those in majority white schools, and by the 8^{th} grade they are only .4 points less, for a net difference of .2 points. In South Carolina, the two groups start out at with the same reading scores, and by 8^{th} grade blacks in majority black schools fall behind only .1 point in reading.

Computational Method. In most desegregation plans, there is a large initial change in racial composition for black students, and then the racial composition is relatively stable except for demographic changes which are more gradual. The initial change rarely exceeds 60 or 70 percentage points, and for large urban districts that are already partially desegregated the change might be 50 percentage points or less. For example, in 1971 Charlotte-Mecklenberg implemented one of the most extensive desegregation plans in the nation, and its first-year change in average percent white for black students was just over 40 percent. The average percent white declined every year thereafter, some of it due to white flight (see Appendix Figure A-1).

We can estimate the change in achievement over time using the lagged regression equation (4). Assume we want to compare the achievement of black student i in predominantly black school 1 and black student j in desegregated school 2 as they progress from grade h to grade k. Assume, further, that the two students are identical on all family background characteristics, and that their two schools are also identical on all factors <u>except</u> percent black in their respective schools and grades, P_{1g} and P_{2g}. Then the difference between their predicted achievements at any grade g is simply

$$\hat{A}_{i1g} - \hat{A}_{j2g} = (P_{1g}D + \theta\hat{A}_{i1(g-1)}) - (P_{2g}D + \theta\hat{A}_{j2(g-1)}) \qquad (5)$$

Applying equation (5) successively from the starting grade h to the ending grade k and gathering terms, we arrive at the difference between predicted achievement at grade k in terms of the initial achievement difference at grade h, the percentage black at each school and grade, the black peer effect D, and the lagged achievement coefficient θ, as follows:

$$\hat{A}_{i1k} - \hat{A}_{j2k} = (\theta^{k-h})(\hat{A}_{i1h} - \hat{A}_{j2h}) + D\sum_{m=0}^{m=k-h-1}\theta^{m}(P_{1(k-m)} - P_{2(k-m)}) \qquad (6)$$

To simplify calculations for an illustration, assume that both students stay in their respective schools from grades 3 to 8 and that the percentage black remains constant in each school and grade, designated by P_1 and P_2.

In most cases, his should produce the maximum achievement difference by 8th grade since the percent black usually increases in a desegregated school following the initial reduction in percent black (see Appendix Figure A-1 for an example). After these simplifications, the equation for the predicted difference in grade 8 becomes

$$\hat{A}_{i18} - \hat{A}_{j28} = (\theta^5)(\hat{A}_{i13} - \hat{A}_{j23}) + D(P_1\text{-}P_2)\sum_{m=0}^{m=4} \theta^m \qquad (7)$$

.

To illustrate, assume that the two students start out with the same achievement in grade 3, that the predominantly black school is 75% black, and that the desegregated school is 25% black. Using the estimated coefficients for North Carolina math (in Figure 5 D = –.006 and θ =.75), the difference in math achievement by grade 8 would be

$$-.006\ (75 - 25)\ (1 + .75 + .75^2 + .75^3 + .75^4) \approx -.9$$

That is, the student in the predominantly black school would experience a net change of –.9 point by the 8th grade. So the one-year effect of -.3 is multiplied by 3 instead of by 5 years, reflecting the attenuation of the lagged relationship over time. This value is slightly larger than the net difference of –.5 points observed in the longitudinal analysis of North Carolina math scores in Figure 2.

Discussion and Conclusions

In our attempt to replicate the HKR Texas study, we have applied the same panel regression models to state data bases that are quite similar to the Texas data. These two state analyses come closest to the Texas study because they permit estimation of student and school fixed effects which were critical to the HKR analyses. We also apply a similar panel regression for the national ECLS sample. While the ECLS data consists of a single cohort and does not allow estimation of fixed effects for schools and students, it has the advantage of being a national sample and thereby important for the generalizability of our results.

Despite the fact that these three data sources involve different populations of students, different achievement tests, and different measures of student backgrounds and school characteristics, the results for black peer effects are remarkably consistent with one another. They all start with moderate relationships between racial composition and black reading and math achievement, but after controlling for student background, lagged achievement, selected school and teacher characteristics, and (for the state analyses) fixed effects for year by grade and school by grade, only one of the six black peer coefficients is statistically significant. Further, this coefficient is reduced to 0 if, in addition, we add average peer achievement as a control variable, which is included in most of the HKR models.

A more descriptive longitudinal trend analysis is also carried out for the three sets of data to describe actual changes in achievement for students in majority white or majority black schools. The results of these analyses are also consistent with the regression results. For the two state analyses, there is a very small effect of racial composition, such that black students in majority black schools experience a slight net decline in math on the order on one-fourth of a point, but the differences for reading are even smaller. In the ECLS analysis, black students in the two groups score nearly identically in grades 1 through 5 after adjusting for student background characteristics.

While these results are highly consistent with each other, they are not consistent with the Texas study, and therefore they fail to replicate those results. Can we identify reasons for the discrepancies? During the course of this study, it was discovered that the computational procedures for estimating multiple fixed effect variables in the Texas study were not properly specified, and therefore the coefficients shown in Table 1 are most likely incorrect. At this point, the Texas analysis has not been re-done using the individual student models from the original HKR papers, estimating school by grade and year by grade fixed effects but excluding student fixed effects. It is possible that, when this is done, the black peer effects in Texas will be consistent with the North and South Carolina results. There is a new analysis of the Texas data using aggregate models, but the models are not the same as those in the original HKR papers (Hanushek and Rivkin, 2006). Aggregate data also give rise to other data processing problems, and we believe that the original HKR models based on individual student data (sans student fixed effects) give more reliable results.

Given these findings, it does not appear that black students who spend most of their elementary and middle school years in predominantly black schools are adversely affected by high concentrations of black peers, once we take into account their family background. To the extent that black students in predominantly black schools score lower than their counterparts in integrated schools, the reasons lie in the fact that these students come from more disadvantaged families, and it is this family disadvantage, not school composition, that explains their academic performance.

From a theoretical standpoint, the data presented here fails to support the three theories outlined above that postulate a positive academic benefit from desegregated schools. Rather, the data are more consistent with the fourth theory, that family background is the primary cause of lower black scores in schools with higher concentrations of black students. From a policy point of view, and especially the school desegregation cases in Seattle and Jefferson County, there may be other reasons for assigning students to schools based on race, but higher academic achievement does not appear to be one of them. To the extent that school boards, courts, and other agencies rely on academic benefits to justify school desegregation policies, that rationale does not appear warranted at this time.

Of course, to this point black peer effects have only been studied in a handful of states and two national data bases, and the existing ECLS data stops in grade 5 (the Rumberger and Palardy NELS data covered grade 8 to 12). So it is possible that larger black peer effects will be found in other states or in national studies covering a broader span of grades. But that is only conjecture at this point; at the present time, there is a lack of evidence supporting the notion that predominantly black schools harm black achievement to a significant degree.

This conclusion does not mean we oppose school desegregation policies. There are other reasons for having desegregated schools, including the simple fact that some students and families prefer desegregated environments. Supporters of school desegregation policies cite social benefits of integrated schools, but it is beyond the scope of this paper to address these other rationales. Our primary purpose has here is to present evidence on the academic benefit thesis, which has figured prominently in the arguments favoring maintaining school desegregation plans.

References

Ainsworth-Danrell, James and Douglas B. Downey. (1998) Assessing the Oppositional Culture Explanation for Racial/Ethnic Differences in School Performance. *American Sociological Review*. 63,4, 536-553.

Armor, David (1995). *Forced Justice: School Desegregation and the Law*. Oxford: New York

Armor, David (2002). Desegregation and Academic Achievement. In C. Rossell, D. Armor, and H. Walberg (Eds.), School Desegregation in the 21st Century. Westport, CN: Praeger.

Armor, David. (2003). *Maximizing Intelligence*. New Brunswick: Transaction Publishers

Armor, David (2006). Lessons Learned from School Desegregation. In P. Peterson (Ed.), *Generational Change: Closing the Test Score Gap*. New York: Rowman, Littlefield Publishers, Inc.

Brooks-Gunn, Jeanne & Pamela Klebanov & Greg Duncan. (1996) Ethnic Differences in Children's Intelligence Test Scores: Role of Economic Deprivation, Home Environment, and Maternal Characteristics *Child Development* 67(2): 396-408.

Brooks-Gunn & Lisa Markman.(2005). The Contribution of Parenting to Ethnic and Racial Gaps in Schooling *The Future of Children*. 15(1):139-168.

Brown v Board of Eduation, 347 U.S. 83 (1954).

Burke, Mary A. and Tim R. Sass, Classroom Peer Effects and Student Achievement, unpublished manuscript, Department of Economics, Florida State University, December 23, 2004.

Clark, Kenneth and Mamie Clark (1939). The Development of Consciousness of Self and the Emergence of Racial Identity in Negro Children. Journal of Social Psychology, 10,591-599.

Coleman, James S, et al. (1966). Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office.

Cook, Thomas D. (1984). What Have Children Learned Academically from School Integration? An Examination of the Meta-Analytic Evidence. In Cook, et al, School Desegregation and Black Achievement. Washington, D.C.: National Institute of Education,

Cook, Phillip & Jens Ludwig (1998). The Burden of 'Acting White': Do Black Adolescents Disparage Academic Achievement in Christopher Jenks and Meredith Phillips eds. *The Black-White Test Score Gap* (Washington D.C:. Brookings Institution Press) 375-400.

Cooley, Jane (2006). Desegregation and the Achievement Gap: Do Diverse Peers Help? available on the world wide web at http://ncpims.mspnet.org/index.cfm/12920

Fordham, Signithia, & John Ogbu. (1986). Black Students' School Success: Coping With the 'Burden of Acting White.' *Urban Reivew*, 18, 176-206.

Grissmer, David Ann Flanagan a& Stephanie Williamson. (1998). Why Did the Black-White Test Score Gap Narrow in the 1970s and 1980s. In Christopher Jenks and Meredith Phillips (Eds.), The Black-White Test Score Gap. Washington, D.C. Brookings Institution Press

Hanushek, Eric. (1996). School Resources and Student Performance. In Gary Burtless (Ed.), Does Money Matter? The Effect of school resources on Student Achievement and Adult Success. Washington D.C.: Brookings Institution Press.

Hanushek, Eric A. and Steven G. Rivkin, School Quality And The Black-White Achievement Gap, Working Paper 12651, National Bureau Of Economic Research, October 2006

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin (2002). New Evidence about *Brown v. Board of Education*: The Complex Effects of School Racial Composition on Achievement,. Working Paper #8799, National Bureau of Economic Research, February.

Hanushek, E.A., John Kain & Steven Rivkin.(2004, 2006) New Evidence about *Brown v. Board of Education*: The Complex Effects of School Racial Composition on Achievement Institute for Research on Poverty Discussion Paper no 1284-04; Revised March 2006

Hedges, l.V., R.D. Laine, and R. Grenwald, R. (1994). An Exchange: Does Money Matter? An Exchange: Meta-Anaylysis of Studies of the Effect of Differential School Inputs on Student Outcomes (An Exchange: Part 1). Education Researcher 23 (3): 4-15.

Hoxby, Caroline (2005) & Gretchen Wiengrath. Taking Race Out of the Equation: School Reassigmnet and the Structure of Peer Effects available on the world wide web at www.economics.harvard.edu/faculty/hoxby/papers/hoxby_ weingarth_taking_race.pdf

Jenks, Christopher. (1972). *Inequality*. New York: Basic Books

Jefferson County Board of Education (2006). Brief for respondents submitted in *Meredith v Jefferson County Board of Education* 05-915, October 10.

Mosteller, Federick & Daniel P.Moynihan (1972). On the Equality of Educational Opportunity. New York: Random House.

Seattle School District No. 1 (2006). Brief for Respondents submitted in *Parents Involved in Community Schools v Seattle School District No.1* 05-908, October 10.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. Teachers, schools, and academic achievement. *Econometrica* 73,no.2 (March):417-458

Rothstien, Richard (1995). *Where's the Money Gone? Changes in the Level and Composition of Education Spending*. Washington D.C.: Economic Policy Institute.

Rothstein, Richard. (2004). *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap*. Washington D.C.: Economic Policy Institute.

Schofield, Janet Ward (1995).Review of Research on School Desegregation's Impact on Elementary and Secondary Students. In Banks and Banks (Eds.), *Handbook of Research on Multicultural Education.* New York: McMillan.

St. John, Nancy (1975). School Desegregation: Outcomes for Children. New York: Wiley & Sons.

Thernstrom, Abigail & Stephen Thernstrom. (2003). *No Excuses: Closing the Racial Gap in Learning*. New York: Simon & Schuster.

Vigdor, Jacob and Thomas Nechyba (2004). Peer Effects in Elementary School: Learning from 'Apparent' Random Assignment, unpublished manuscript, Duke University and NBER

APPENDIX

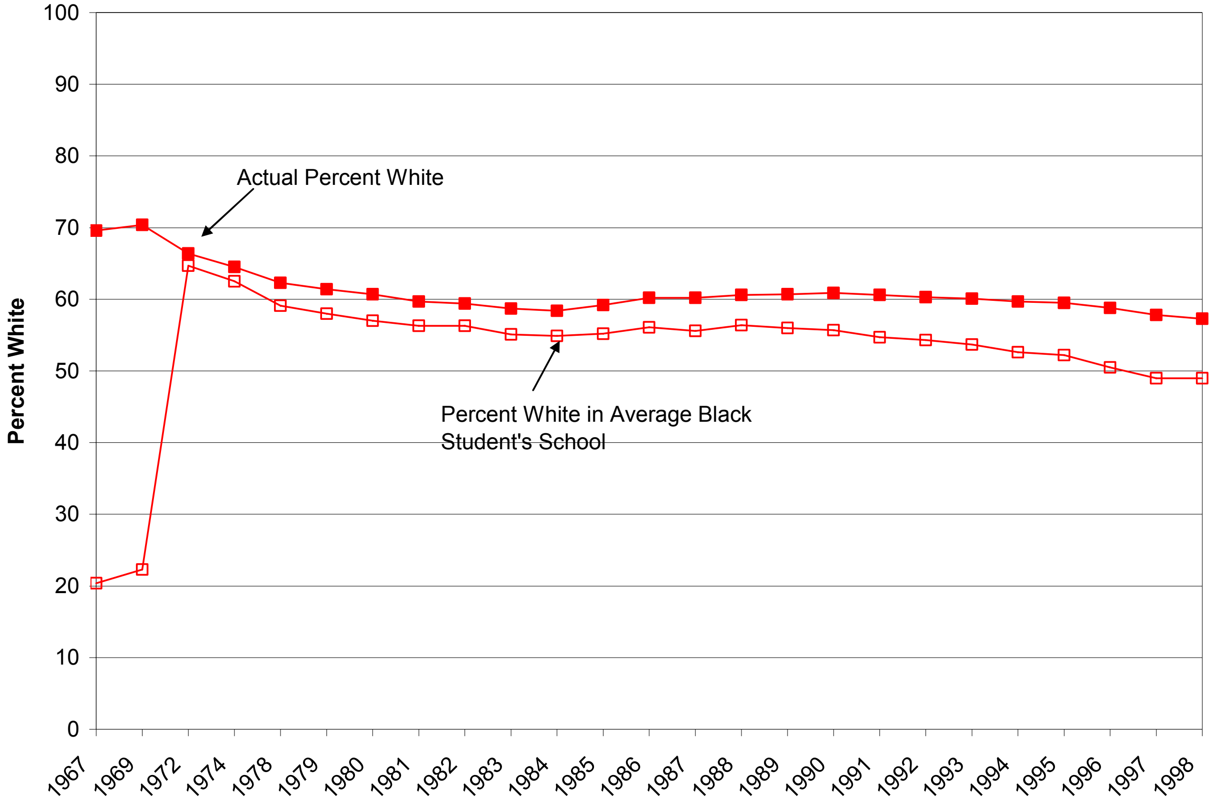

Figure A-1 CHANGE IN BLACK EXPOSURE TO WHITE STUDENTS IN CHARLOTTE-MECKLENBURG ELEMENTARY SCHOOLS

TABLE A-1 STATISTICS FOR NORTH CAROLINA VARIABLES, BLACK AND WHITE STUDENTS[a]

Characteristic	Mean	SD	N Obs.	Minimum	Maximum
Math (t)	250.2	10.0	1803073	208	286
Reading (t)	250.3	9.9	1796098	209	281
Lagged math (t-1)	250.6	9.8	1688025	211	286
Lagged reading (t-1)	250.7	9.7	1678012	209	281
Avg. peer reading(t-2)	251.6	3.0	984532	217	270
Avg. peer math(t-2)	251.5	2.8	984527	222	269
% Black peers	30.0	23.9	1830743	0	100
% Free lunch	34.5	45.3	1806482	0	100
Parent education (yrs)	13.5	2.3	1817535	10	19
Same school (yrs)	3.3	0.9	1830743	0	5
% Certified tchrs	85.4	9.9	1215347	0	100
Pupils per teacher	14.7	2.1	1215064	2	25

[a] All test scores standardized to mean=250 and sd=10 for all students in each year and grade

TABLE A-2 STATISTICS FOR SOUTH CAROLINA VARIABLES, BLACK AND WHITE STUDENTS[a]

Characteristic	Mean	SD	N Obs.	Minimum	Maximum
Math (t)	100.0	10.0	843061	61	139
Reading (t)	100.1	9.9	836148	56	141
Lagged math (t-1)	100.1	9.9	609006	61	139
Lagged reading (t-1)	100.2	9.9	603258	56	141
Avg. peer reading(t-2)	98.8	5.7	529107	63	124
Avg. peer math(t-2)	98.7	5.6	528579	60	123
% Black peers	41.0	27.3	880399	0	100
% Free lunch	48.6	47.8	879679	0	100
Same school, t & t-1	0.7	0.5	648585	0	1
Repeated grade	0.0	0.2	836148	0	1
Per pupil expend.	5968.0	1187.1	853507	0	36416
% Prime instruct. time	89.6	2.4	855914	76	99

[a] All test scores standardized to mean=100 and sd=10 for all students in each year and grade

TABLE A-3 STATISTICS FOR ECLS VARIABLES, BLACK AND WHITE STUDENTS[a]

Characteristic	Mean	SD	N Obs.	Minimum	Maximum
Math (t)	51.1	9.8	33636	7	89
Reading (t)	51.0	9.8	33594	7	95
Lagged math (t-1)	51.2	9.7	26919	7	89
Lagged reading (t-1)	51.0	9.7	26882	7	95
% Black peers	19.2	31.9	33755	0	100
Parent education	5.0	1.8	31789	1	9
Income category	8.0	3.2	32355	1	13
Number of siblings	1.4	1.1	32046	0	12
Age of mother, 1st Birth	23.7	5.5	32360	12	45
Teacher experience, yrs	13.9	10.0	32013	0	49
Teacher MA+ (prop.)	0.8	0.4	30346	0	1

[a] All test scores standardized to mean=50 and sd=10 for all students in each year and grade