Center for American Progress

Lost Learning, Forgotten Promises

A National Analysis of School Racial Segregation, Student Achievement, and "Controlled Choice" Plans

Douglas N. Harris

November 24, 2006

Progressive Ideas for a Strong, Just, and Free America

Center for American Progress

Table of Contents

Summary

The struggle to desegregate America's schools while ensuring equal educational opportunities for students of all races is one of the greatest social challenges the nation has faced over the last half century. While significant progress has been made since the Supreme Court's 1954 landmark decision in *Brown v. Board of Education*, thousands of schools around the country are still almost completely segregated. In the coming months, the Court will once again address the issue when it considers the constitutionality of "controlled choice" programs in Louisville and Seattle. These efforts, unlike the controversial busing of the 1960s and 1970s, are implemented without court intervention and allow parents a variety of school choices while still ensuring some degree of racial integration.

This report considers the educational consequences of the considerable racial segregation that remains in schools today and the potential of controlled choice to address them. It begins with an extensive review of research regarding the effects of school integration. Previous research provides relatively strong evidence that desegregation helps minority students reach higher academic achievement and better long-term outcomes such as college attendance and employment.

Previous studies on the subject, however, are either decades old or focus on relatively small groups of students. This report provides a new, exhaustive analysis of racial segregation across the country. Using test score information required by the federal *No Child Left Behind* (NCLB) Act, the study analyzes the effects of segregation in more than 22,000 schools across the country that enroll more than 18 million students. Most previous studies on the subject have included no more than a few thousand students, making this study arguably the largest ever conducted on the effects of segregation.

The new information is used to address two basic questions: First, do minority students learn more in integrated schools? Second, would racial integration improve the equity of learning outcomes in general and in the Louisville and Seattle districts that are the subjects of the Court case? The answers to these questions appear to be "yes." Specifically:

- African Americans and Hispanics learn more in integrated schools. Minorities attending integrated schools also perform better in college attendance and employment.

- Controlled choice and other forms of desegregation benefit minority students.

- Racial integration is a rare case where an educational policy appears to improve educational equity at little financial cost.

These results have significant implications for the Supreme Court's upcoming decision. In the original *Brown* decision, as well as a more recent case involving race and admissions to universities, a majority of the Court argued that considering race in school assignment constitutional partly because racial integration is an important part of the learning environment. By showing that less learning takes place in segregated schools,

the results in this study support the contention that racial diversity is important to the learning environment in schools. If the goal is to improve achievement, then opposing controlled choice is counterproductive.

While the Court's decision will have obvious implications for the future of desegregation programs, it may also complicate the implementation of NCLB. By evaluating schools based on test scores of racial subgroups, this federal accountability policy is, like controlled choice, *explicitly race conscious*. As a legal matter, the rejection of controlled choice by the Court could therefore put at risk the racial considerations in NCLB. Moreover, as a practical matter, if race is going to be a factor used to measure school success, then it stands to reason that schools should be able to consider race through such programs as controlled choice when addressing apparent school failures.

After a half-century of court cases and new policies, the nation still finds itself with highly segregated and inequitable schools. The main issue before the Supreme Court, and the nation's citizens, is whether we will continue to accept these inequities or move forward in fulfilling the promise of *Brown* and the moral and educational imperatives of racial integration.

The Promise of *Brown*

"Our decision, therefore, cannot turn on merely a comparison of these tangible factors in Negro and White schools . . . We must look instead to the effect of segregation itself on public education." Chief Justice Earl Warren, Brown v. Board of Education, 1954

May 17, 1954 is one of the most important days in the history of American education and, arguably, the history of the nation as a whole. On that day, Justice Earl Warren delivered the Supreme Court's majority opinion in *Brown v. Board of Education* that overturned the "separate but equal" doctrine and made it unconstitutional to establish laws that required racially segregated schools. Together with the civil rights movement they helped to spark, *Brown* and subsequent court decisions changed the way white Americans saw African American children and how African American children saw themselves. These momentous decisions changed expectations, created new hope, and made a bold promise of equal opportunity.

Part of that promise has no doubt been fulfilled. In the South in 1954, nearly 100 percent of African American children attended schools where nearly all of the other students were also minorities, and the situation was not much better in other parts of the country.[1] By 2000, this number had dropped to 37 percent nationwide.[2] There were also gains in educational resources and outcomes. In 1952, Mississippi spent more than three times as much per pupil on white students compared with African American students. Also, in 1946, there was a 22 percent gap nationwide in salaries paid to teachers and a 42 percent gap in pupil-teacher ratios in African American schools compared with all-white schools.[3] These resource gaps now have been almost completely erased and, partly as a result of integration and more equitable resources, the achievement gap between whites and African Americans declined by more than one-third between 1975 and 2000.[4]

While these are tremendous accomplishments, the task of desegregation—and the larger goal of educational equity—are far from finished. The remaining gaps in outcomes, while much smaller than in the past, still loom large. Likewise, there is still substantial—and now *increasing*—racial segregation in schools, reversing decades of improvement.[5] This re-segregation, as discussed in the next section, is partly the result of subsequent court decisions that have severely limited the desegregation options available to states and school districts.

The Supreme Court's upcoming decision on "controlled choice" could be yet another step backwards. The case involves programs in two school districts, Louisville and Seattle, which give parents a choice about where they send their children but also often consider the student's race in placement.[6] A large number of districts have adopted this strategy in past 15 years and as many as 1,000 districts nationwide consider race in some way in student assignment to schools.[7] The upcoming Court decision may therefore have far-reaching, national consequences.

Given that the programs were generally designed by districts to avoid court-ordered desegregation, and to comply with the Court's earlier decisions, it came as a surprise to many observers that the Court decided to hear the appeals.[8] Race is only one factor in the assignment of students to schools and, when used in this way, it appears—or least *appeared*—to meet current legal and constitutional requirements. The fact that the

Court has agreed to hear the case may signal an inclination to move even further from the goals and strategies of *Brown* and to send a broader message about the views of the new Supreme Court, with its two new justices, on racial issues.

This message has already been sent by the Bush administration, which has urged the Court to reject controlled choice.[9] This position might be expected given the administration's consistent opposition to explicit considerations of race. However, the administration's position on this question is, in many ways, in conflict with its other positions on educational reform. First, the stated goal of the president's signature educational program—The *No Child Left Behind* (NCLB) Act—has been to improve student achievement. The review of research and new analysis in this report suggest that desegregation can help raise achievement and meet the NCLB goals. Second, by requiring that students in various subgroups achieve academic proficiency on student achievement tests by 2014, NCLB was designed with the apparent intention of improving equity.[10] Again, the evidence in this report suggests that minority students learn more in integrated schools. Third, by reporting scores by racial subgroups, *NCLB, like controlled choice, is explicitly race-conscious*.[11] If race is going to continue to be a factor used to measure student and school success, then it stands to reason that schools should be able to consider race in addressing the apparent school failures. In all of these respects, the administration's opposition to controlled choice seems to contradict its positions on educational reform.

Below, I provide a brief overview of the legal history of desegregation, starting with the *Brown* decision. This includes an introduction to the two main legal standards that racial integration policies must meet in order to be constitutional. It will be necessary for the two districts, as defendants in the case, to establish that the government has a "compelling interest" in pursuing racial integration and controlled choice is "narrowly tailored" as a policy to pursue that interest.[12]

Both of these legal issues rest partly on social science evidence about the effects of racial integration. I therefore revisit evidence from the decades-old desegregation experiments and summarize newer evidence about how the racial composition of classrooms and schools affects student achievement. An important part of this review is that it goes beyond student achievement and considers evidence about effects of desegregation on high school graduation, college attendance, employment, and wages. Because education is ultimately intended for long-term life success, these outcomes are at least as important as short-term achievement.

There is also social science theory supporting this evidence. One of the most important ways that racial integration benefits minorities is that it places them in schools with more advantaged classmates—students whose parents have higher incomes and educational levels. It is well established that students' academic and long-term success are closely associated with these measures of their socio-economic status and social class.[13] It is therefore not surprising that having classmates with higher socio-economic status is also important. Schools with advantaged students have advantaged learning environments—fewer classroom disruptions, higher expectations, and greater parental support at school and especially at home.

A second important factor is that educational resources are equitable when schools are integrated. One reason is that school funding is based significantly on local property taxes. Neighborhoods around high-minority segregated schools have lower property values and therefore receive less funding. In addition, the most important school resource—teachers—are less likely to teach in high-minority schools for a variety of reasons. In addition to the previously mentioned issues of disruptions, parents, and expectations, teachers are often ill-prepared for the unique challenges in these educational environments. Thus high quality teachers are less likely to apply to teach in these schools and, when they do, are more likely to get burned out and leave for other schools. In short, even if all schools had the same level of *funding*, high-minority schools would still fewer school *resources*.

While there is already evidence that minority students learn less in segregated schools, past evidence is several decades old or focuses on relatively small groups of students. I therefore provide an extensive new analysis of student achievement data collected as a result of NCLB. States are now testing students more frequently and reporting the data by racial subgroups, making it possible for the first time to measure the progress of students as they proceed through school, in thousands of schools across dozens of states. This database has only become available this year and this is the first analysis that uses the data to study the effects of segregation on a national scale. Finally, I provide separate analyses of Louisville and Seattle because of their roles in the Court case and because they illustrate the larger implications for other school districts.

This new review and analysis suggest that, to promote strong academic learning and long-term employment success, school districts have a compelling interest in racial integration. The discussion of alternative policy options also suggests that racial integration—and controlled choice, in particular—are narrowly tailored to achieve these goals.

Controlled Choice: A Brief Legal History and Introduction

The fact that the Court's *Brown* decision has fallen short of producing desegregated schools is unsurprising given the series of subsequent Court decisions that have severely limited the available desegregation strategies. I discuss some of these cases below to provide the legal context for the upcoming controlled choice decision.

The *Brown* decision, at its most basic level, required that schools end *de jure* segregation—that is, it decreed unconstitutional laws that require racial segregation of students. In a subsequent decision the following year, sometimes called *Brown II*, the Court also went beyond this when it required schools to desegregate with "all deliberate speed" and ordered lower federal courts to oversee these efforts. The clarity of the Court's message appeared to have put to rest the "separate but equal doctrine" established by the Court in *Plessy v. Ferguson* in 1896.

As significant as the *Brown* decision may have been, however, it had only a small effect on actual, or *de facto*, segregation. It was not until the Court's 1968 decision in *Green v. County School Board of New Kent County* that the guidelines provided by the Court to lower federal courts were concrete enough to overcome the hostile public reaction to the idea of desegregated schools. This began the most intensive period of school desegregation, from 1968-1972, during which the percentage of African American students in intensively segregated schools dropped from 64.3 to 38.7 percent.[14]

While the Court surely anticipated the public hostility that came with *Brown*, it may not have fully grasped the constitutional issues that would arise in the process of addressing that hostility and actually achieving desegregation. Indeed, as we will see below, most of the viable options set out by Court justices for reaching the goals of *Brown* were later rejected by future justices.

The first decision that significantly limited options to pursue desegregation was *Milliken v. Bradley* in 1974, where the Court ruled that school districts in the Detroit metropolitan area could not be required to integrate across district boundaries. Given the high level of housing segregation in the Detroit-area, as well as many other metropolitan regions, this decision effectively took at least one-third—and arguably much more—segregation off the table.[15] As a result of the *Milliken* decision, those parents who wished to avoid desegregation could and did so by moving to the suburbs.

The options for lower courts to require desegregation within districts also became limited. Two Court decisions in the 1990s, *Board of Education of Oklahoma v. Dowell* (1991) and *Freeman v. Pitts* (1992), took away district responsibility to periodically adjust school attendance boundaries to maintain or increase racial integration. Attendance zones, common to nearly all school districts, refer to the neighborhoods that students reside in and the schools that these students are designated to attend by the district. Once again, this meant that parents wishing to avoid integration could move their residence, this time to a different attendance zone within the district. These two court decisions made clear that desegregation was allowable only as a temporary means to address historical discrimination, not as a permanent means to provide equal opportunity as housing conditions changed.[16]

Controlled Choice and Other Forms of Desegregation

While there have been many legal setbacks, Court decisions have not yet shut off all paths to desegregation. Desegregation within and across districts can no longer be *required* by lower courts, but it can be *allowed* through voluntary district efforts. Indeed, many districts, believing in the objective of integration and hoping to avoid lawsuits, responded to the earlier Court decisions by searching for other ways to integrate schools. Controlled choice became one popular option.

The concepts and mechanisms of controlled choice are perhaps best explained by Dr. Charles Willie of Harvard University, one of the main architects behind one of the earliest controlled choice programs in Boston. Willie describes controlled choice as a desegregation policy that allows parents wide latitude in choosing the schools their children attend and to ensure that these choices do not result in segregated schools, allows school districts to take race into consideration in assigning some students.[17]

Importantly, controlled choice is also built on the idea that desegregation is best done as part of a comprehensive solution to school improvement. The parental choice component of controlled choice is not only a means of obtaining political support for the idea, but a way of improving schools so that they are desirable to white and middle class parents. Following the economic theory of school choice, parental choice is intended to create market pressures on schools to improve. Over-subscribed schools are seen as successful and models to be emulated, while under-subscribed schools are pressured to improve. In this respect, the objective of controlled choice is to improve both the equity and quality of education.

It is important to understand how controlled choice differs from the traditional attendance zone approach, as well as various approaches that have been used to facilitate desegregation: busing, paired attendance zones, periodic shifts in attendance zones, and magnet schools. Each of these approaches, as we will see, provides a different balance of several often competing priorities: parental choice, proximity of school to home, and of course potential success in achieving integration.

The cases of *Board of Education of Oklahoma v. Dowell* (1991) and *Freeman v. Pitts* (1992), discussed above, highlighted the use of shifting attendance zones as a path to desegregation. This approach, by continually adjusting zones to account for housing patterns, limits parental choice and creates a degree of uncertainty from year to year about where children will go to school. However, it also ensures some degree of integration, so long as districts are willing to create attendance zones that do not align with nearby neighborhoods.

Perhaps the most well known and controversial approach to desegregation involves busing minority students to majority white schools. While busing obviously has the potential to desegregate schools, it has been extremely unpopular and therefore largely abandoned in practice. Alternatively, with paired attendance zones, two conventional (and usually contiguous) attendance zones are combined together and students spend roughly half of their schooling in neighborhood schools and the remainder in schools in the contiguous zone. Again, parents have few choices, but their children are now guaranteed to attend a school that, while not necessarily in their neighborhood, is still relatively close to home.

A final set of alternatives, magnet schools, take a very different approach and involve the creation of specialty schools, usually in low-income neighborhoods, that are designed to attract students from middle- and upper-class neighborhoods. This approach provides a combination of choice and proximity. Also, like controlled choice, it is based on the principle that different students have different needs and that specialty schools, combined with parental choice, can improve the overall quality of education throughout a district. Under both magnet schools and controlled choice, the success of the program in achieving both desegregation and quality depends on whether the schooling options are truly desirable to white and middle class families.

Weighing the advantages and disadvantages of these approaches, it is easy to see why controlled choice has been an attractive option. If it works as intended, it provides the best of all worlds, allowing parents to choose where their children go to school while simultaneously achieving desegregation and improving the overall quality of schooling.

Having provided this background about the various ways to achieve racial integration, I next consider the potential benefits of integration for minority students.

Revisiting the Effects of School Segregation

While the moral victory of eliminating de jure segregation was arguably the most significant contribution of *Brown*, it is also important to understand the measurable effects of desegregation. That is, what happens to minority students when they attend high-minority schools compared with majority white schools? Previous research on this question falls into three main categories: the effects of the decades-old desegregation experiments, newer evidence on the effects of classroom racial composition, and studies of the long-term effects of segregation on outcomes.

In discussing these studies, I refer to the estimated "effects of desegregation" and the "effects of racial composition." It is important to clarify in advance what these terms mean and do not mean. First, the effects are not due to race per se, but to the other characteristics of students' classmates, such as parental education and school resources, both of which are correlated with race. Therefore, these "racial" effects capture a complicated set of influences.

Second, the word "effect" often implies that the factor being discussed is the "cause" when, in fact, it is somewhat more complex. In technical terms, the effects discussed below are correlations between racial composition and student outcomes (e.g., achievement) and the degree to which they can be interpreted as causal effects varies from study to study. This is really a larger issue of research quality. While all of the studies discussed below estimate correlations between racial segregation and student outcomes, they vary in the degree to which they reflect causal relationships. The gold standard in research is often considered to be the experiment in which people are randomly assigned to receive some "treatment" and others are assigned to a "control" group that receives no treatment. Because participants are randomly assigned, it is generally reasonable to assume that differences in outcomes between these two groups are caused by the treatment. However, experiments are difficult to conduct and are thus relatively rare in education, leading to a frequent concern that the student groups being compared are not equivalent. This problem, often called "selection bias," means that it is difficult to separate the effect of the treatment from the other systematic differences between the groups that may also have influenced the same outcomes.

In this report, the challenge is to separate the effect of segregation from the factors that lead students to end up in segregated schools, both of which can influence academic achievement and long-term outcomes. For example, African American students from low-income families are on the average more likely to end up in high-minority schools than are wealthier African Americans. If students learn less in high-minority schools, as the present analysis suggests, is it because of segregation or because of students learn less when they come from less wealthy families? This is sometimes a difficult question to answer, but one the present study tries to address. In short, correlation with segregation does not necessarily imply causation.

Three categories of studies are considered below. The first group uses experimental designs. However, these studies are limited in three ways: they are now very old, they include very small samples of students, and they focus on a limited array of student outcomes. Therefore, I also consider two other sets of studies that address both these limitations while still accounting for selection bias.

Effects of Desegregation on African American Achievement

As noted earlier, the most aggressive efforts to desegregate public schools occurred during the late 1960s and early1970s as the Supreme Court and federal government gradually increased pressure on schools to integrate. In some cases, including some true experiments, the effects of desegregation were actively studied by researchers.

In 1984, the National Institute of Education (NIE) commissioned seven extensive reviews on the effects of desegregation.[18] All of the NIE-commissioned reviews suggest that African American students had higher test scores as a result of desegregation.[19] In one important review, Crain and Mahard (1983) found large benefits for African Americans in a large majority of the studies that used the best research methods.[20] Of the total reported effects, 85 percent were positive and statistically significant. Bused students appear to gain an average of eight percentile points over the control group. It is important to emphasize that the control and treatment groups in these cases were chosen randomly so that this effect is a reasonable estimate of the achievement difference that bused students would have attained had they remained in segregated schools.

There are also several patterns in the results that are important for the design and implementation of desegregation. First, the results suggest that voluntary desegregation efforts, such as most controlled choice programs, are more likely to have positive effects than court-imposed programs.[21] Putting this conclusion even more strongly, Bradley and Bradley (1977) found that all of the voluntary desegregation programs show positive effects on African American achievement.[22]

Minority students who are desegregated at a younger age, in elementary school, also seem to benefit more than those desegregated later in their school careers.[23] Three-fourths of the studies where desegregation occurred in kindergarten showed achievement gains and the effect sizes were larger than in desegregation efforts aimed at older students.[24] One possible explanation is that older students, who have become more accustomed to racial segregation, have greater difficulty making the adjustment and therefore gain less than more adaptable younger students. In addition, it is common for middle and high schools to place students in different academic tracks, leading to racial segregation within these schools.

Peer Effects and Achievement

Formal desegregation programs are not the only source of information regarding the achievement effects of racial composition. A more recent and growing research literature on "peer effects" tries to understand the issue using sophisticated analyses of recently available large-scale administrative databases, which track performance of large numbers of individual students over time. This means that it is possible to compare learning of the same student in different classroom settings and in different racial compositions. While random assignment is in some ways preferred, comparing students to themselves in this way also has advantages in accounting for selection bias.

One advantage of the peer effect studies is that they make it possible to separate the effects of peer race from the effects of peer achievement. As indicated earlier, minority students tend to have lower achievement and it is reasonable to expect that a student who has classmates with lower achievement will learn less herself. Lower-achieving classmates may directly affect each other by being less able to help each other with their academic work. Indirectly, it is also likely that lower-achieving students have more behavioral problems, less support at home, and lower academic expectations.

As with the desegregation experiments, peer effects evidence tends to support the idea that desegregation improves student achievement. In two extensive studies using data from Florida and North Carolina, researchers found fairly strong evidence that having peers with higher achievement raises individual achievement.[25] In the North Carolina study, Cooley (2005) finds that the achievement of African Americans was affected by the achievement of both white and African American classmates.[26] This study, as well as another in Texas[27], also found that the race effects were largest for African American students who initially had higher scores.

Two other studies of Texas present similar results for African Americans, but also include Hispanics (Hanushek et al., 2002; Hoxby, 2000).[28] In both studies, the authors found that having a lower proportion of African American classroom peers has a positive influence on achievement for students in all racial groups, but especially for other African American students.[29] Interestingly, Hispanics in Texas do not appear to be influenced by the percentage of their peers who are Hispanic.

Cooley (2005) also simulates different possible student assignment policies. She finds that randomly assigning students (in effect, distributing students evenly by race) would result in non-white achievement gains of 0.05 standard deviations per grade. Assuming similar gains could be achieved at other grade levels, this could imply large cumulative gains over the course of students' entire school careers. Over 13 grades (K-12), this could lead to gains as large as 0.65 standard deviations or 80 percent of the entire achievement gap. Even these large estimates may understate the effect of desegregation because the simulations control for many factors, such as teacher quality, that would be likely to improve for minorities as a result of racial integration as well. On the other hand, while random assignment simplifies the simulation, it is not especially realistic from a policy perspective. In this sense, the potential gains estimated by Cooley are probably overstated.

In short, African American students learn more when they have white peers and peers with higher test scores. These results are also consistent with the studies of older desegregation experiments discussed earlier and evidence on long term outcomes discussed below.

Long-Term Effects of Desegregation on African Americans

The above studies focus on student achievement and there are many reasons to do so. In addition to being a primary objective of schooling, students with higher test scores are more likely to go to college and to have greater success in the labor market. For example, there is evidence that a one standard deviation increase in student test scores is associated with an eight to 20 percent increase in wages when those students enter the labor market.[30] But there is much variation in labor market success that is not explained by test scores. Moreover, these economic gains from higher test scores are about the same as the gains from an additional year of schooling.[31] It is therefore important to go beyond achievement to understand the long-term effects of segregation.

Earlier, various explanations were given regarding why desegregation might help improve minority achievement—more advantaged peers and greater school resources. For longer-term outcomes, the story seems to be more about "social learning" than academic learning. Some researchers have found, for example, that desegregation may give African American students the confidence and ability to cope with situations involving white students who make up the majority of people in colleges and workplaces.[32] Desegregation also establishes networks that provide information and personal connections that help students navigate their educational and career paths. As we will see below, the evidence supports the theory that this social learning from desegregation has significant long-term influences on outcomes.

One of the busing desegregation experiments discussed above is the Boston METCO program, which involved randomly assigning high school minority students from Boston to majority white schools in nearby suburbs. Armor (1972) finds that 84 percent of the bused students went on to college after high school compared with 56 percent in the control group.[33] Two other national studies are worth noting. Grogger (1996) and Rivkin (2000) both find that having more white classmates is positively associated with higher wages for African Americans when they finish school.[34] Rivkin also finds that the effects on long-term student outcomes are smaller in urban school districts that were under involuntary desegregation, consistent with the earlier evidence on desegregation experiments.

As with achievement, the effects found in the most convincing studies of long-term racial composition are almost universally positive. Moreover, they arise not only because of the apparent achievement gains and their indirect effects, but also because of the social learning environment of minority students in majority-white schools.

Addressing the Counterarguments

There are alternative interpretations of the above research that deserve attention. Most importantly, some researchers, and particularly critics of desegregation, argue that the results are inconsistent across studies. On one level, this is a valid criticism. If a program such as desegregation has an effect, and it is implemented multiple times under similar school conditions, then the effect should be similar across applications. By analogy, if we mix the same combination of chemicals together in the same container, and we repeat the exercise, then the chemical reaction should be the same every time.

The problem, however, is that the conditions under which programs are implemented, and the implementation itself, can vary considerably. For example, in the busing studies, minority students in some schools may have been "tracked" into classrooms with other minority students and therefore received fewer benefits from white or higher achieving peers; in other schools, students may have been placed in more integrated classrooms, creating more positive effects. Therefore, while researchers are correct to be concerned about variation in effects, it would be wrong to dismiss the results for this reason alone. This is especially true in the case of desegregation where the effects are consistently positive.

Results also tend to vary based on research methods. As noted above, what are generally considered methodologically strong studies—those that account for selection bias—are more likely to find positive and significant desegregation benefits for minorities. If we ignored the research methods, then the effects on achievement are somewhat more ambiguous, but ignoring these methods also goes against standard research practice.

It is difficult to say how much inconsistency should be expected based on the differences in implementation and research methods. But it is striking that the most rigorous studies in all three categories—desegregation effects on achievement, peer effects on achievement, and racial composition effects on long-term outcomes—all point to the conclusion that desegregation benefits minorities. As we will see below, this conclusion is reinforced by a new analysis of the NCLB database.

New Data: Student Testing and NCLB

NCLB and its requirements for annual testing and reporting of scores by racial subgroups provide an important new source of evidence regarding the effects of desegregation. Regardless of one's views on the new law, there is no question that it has produced an unprecedented amount of new information about academic achievement in our nation's public schools.

In addition to requiring the new testing, the federal government has commissioned an outside organization, the American Institutes for Research, to collect the data from all states in a single database, called the National Longitudinal School-Level State Assessment Score Database (referred to as the "NCLB database" throughout this study). The NCLB database merges the test score data required by NCLB with the U.S. Department of Education's long-standing Common Core of Data (CCD), adding information about eligibility for free or reduced price lunches (a measure of family income) and other basic school characteristics.

Using the NCLB data, it is now possible for the first time to calculate students learning in schools across the country. For example, in Tennessee, which has arguably the most complete data of any state, scores are reported in the NCLB data for grades 3-12 in years 2004 and 2005. It is therefore possible, for example, to calculate learning gains for students who were in grade three in 2004 and who continued to grade four in the same school in 2005. The same is true for the cohort of students in grade four in 2004 who continue on the grade five in 2005 and so on through grade 12.

Because NCLB requires that scores be separated by racial and other subgroups, it is also possible to calculate changes in scores not only by grade and year cohort, but also by race. Extending the Tennessee example, this means that the students in grade four in 2004 can be divided into various subgroups—African Americans, Hispanics, and so on. From this information, it is possible to calculate how much each sub-group is learning and to compare learning gains with racial composition.

While most states are still phasing in their standardized testing to meet the new federal requirements, 22 already report enough data to calculate learning gains for students in elementary grades. A smaller number of states report enough information to make these calculations in middle school (16 states) and high schools (seven states). A total of 23 states are studied for at least one school level.[35]

More than 22,000 schools enrolling more than 18 million students are included in the analysis. The students in the NCLB database are 20 percent African American and 24 percent Hispanic compared with 16 and 15 percent, respectively, for the national as a whole. [36] The nine-point difference for Hispanics is not surprising given that California, Florida, and Texas are all included and have disproportionate shares of this group. Likewise, many states with large urban areas and African American populations are excluded. Finally, even within states that provide data, the most heavily white schools sometimes do not report data for minorities because there are too few of them to provide a reliable estimate of achievement. There is no reason to expect, however, that the role of race in this large sample of schools is different from the larger population.

While the two states that are involved in the upcoming Supreme Court decision—Kentucky (Louisville) and Washington (Seattle)—are among the states providing insufficient achievement data, it is possible to estimate the impact of segregation in these locations. This is accomplished using data from the other states and simulating the effect that eliminating controlled choice might have on students in these districts. Given the extremely large number of schools included from other states, there is good reason to believe that the conclusions of the present study can be applied to these states and districts as well.

Below I explain why the NCLB data are particularly useful for this type of analysis. I also discuss differences in state achievement tests as one limitation of the data.

The Importance of Learning Gains

Standardized testing of students has grown steadily over the past century in American public education. But until the past decade, tests were administered infrequently and designed only to obtain periodic snapshots of student performance. Within the past decade, however, many states have begun testing more frequently and redesigning the tests to measure changes in student achievement—*student learning gains*. This trend has been significantly accelerated with the passage of NCLB and its extensive new testing requirements.

Of course, there are many ways—positive and negative—that testing influences how schools work and considerable debate concerning how the tests should be used. Less debatable is that the increase in testing yields useful new information about school performance. In the past, the available data helped us answer the question, "how much do students *know*?" Most schools reported, for example, the percentage of students scoring above a particular cut-off or the percentile ranking of the average student at the school at a particular point in time.

Understanding how much students know is important, but it tells us little about what schools contribute. Instead, we must ask, "how much have students *learned*?" The difference between these two questions, and their answers, are often misunderstood in debates about education and in the design of accountability policies.[37] To understand how much students learn, we need to know first where they start. We know, for example, that minority and low-income students start off far behind their peers from the first day they enter kindergarten.[38] Clearly, these differences cannot be attributed to the kindergarten teacher or subsequent teachers. It is much more reasonable to attribute learning gains to teachers and schools because these take into account where students start.[39]

The importance of learning gains can also be understood as an issue of selection bias. As discussed earlier, experimental research designs are desirable because random assignment helps to reduce or eliminate relevant differences between the control and treatment groups. Without random assignment, the differences between any two groups might be caused by factors other than the treatment. In this case, if we were to compare the test score levels of the average high-minority school with the average low-minority school, we would see substantial differences for the reasons discussed earlier.[40] But are the lower score levels in high-minority schools caused mainly by the schools themselves? It depends. Again, there are substantial differences in the initial test score levels of minority and white students and a direct comparison on this basis results in selection bias. Learning gains, by considering where students start, substantially reduce this problem.

In short, the differences in learning gains between segregated and less segregated schools provide valuable evidence about the possible benefits of controlled choice and other forms of desegregation. The NCLB data makes it possible to study student learning in far more schools than was previously possible.

State Differences in Test Score Types

One of the common complaints about NCLB, even among its advocates, is that, because it relies on state-determined tests and state-determined proficiency definitions, it is difficult to make comparisons across states. Some states use "criterion-referenced" tests intended to measure the degree to which students have learned a certain set of academic standards.[41] Other states use "norm-referenced" tests which, in contrast, measure how well students perform relative to other students. The content of achievement tests also varies. Michigan's math test might, for example, focus more on geometry, while the one in Florida emphasizes algebra.

In addition to content, standardized tests vary regarding the test scale. To see how, consider what would happen if we were to count the number of correct answers on a test. The problem with this approach is that a correct answer about basic arithmetic would be given the same weight as a correct answer for a more difficult question about trigonometry. Further, one student may answer all of the easier questions correctly but miss all of the harder questions. Another student, in contrast, might get the same total number correct but do so by choosing the right answer for half of the easy questions and half of the harder questions. In this case, counting the number of correct answers would make it appear that the two students have the same achievement when, more realistically, it appears that the latter student has higher achievement.

To account for this measurement difficulty, test developers will usually "scale" the test to account for test item difficulty. In this analysis of racial composition, there are two reasons to believe that test scaling might be an important issue. First, as we will see, student achievement is lower in high-minority schools. Second, achievement gains may be easier to make when starting off with a low initial score. One of the main reasons is the so-called "ceiling effect" which arises when standardized tests only capture achievement up to a particular level. A student who starts a school year with a very high level of achievement may have difficulty showing measurable gains, even if they have learned a great deal. By analogy, a basketball player who makes 98 percent of his free throws will have great difficulty showing improvement in free throw shooting. This means that, if we ignore for the moment the possible effects of segregation, achievement gains may be *more likely* to occur in high-minority schools where initial achievement levels are low. In some states, it may also be easier to make gains when starting off around the average test score compared with some lower or higher level.

More information about the data is provided in Appendix A. In the next section, I use the NCLB data to provide an extensive new analysis of the effects of desegregation.

New Evidence on Segregation Effects

This section provides evidence regarding the effects of segregation on student learning using the NCLB data. This is followed by simulations of the student achievement effects of controlled choice in Louisville and Seattle. Specifically, I estimate the drop in learning that might be expected if the Court rejects controlled choice and the districts revert to their previous levels of segregation.

The effects of racial composition on student learning are estimated using regression analysis, which helps to isolate the effects of various factors on student learning. This type of analysis also makes it possible to account for some of the differences in achievement tests across states simultaneously. Previous evidence suggests that the benefits of desegregation for African Americans are largest in earlier grades; therefore, each figure is for a different school level (elementary, middle, and high). The full results, as well as explanations of how the analysis was carried out, are provided in Appendix B.

Achievement Gains and Segregation

Figures 1-2 below plot the regression-adjusted student achievement gains for African Americans and Hispanics according to the school percentage minority. In each figure, a downward sloping line suggests that students learn more in schools with more white students. Whether each line is above another line is unimportant.

Figure 1: Elem. School, Adjusted Ach. Gains, by % Minority

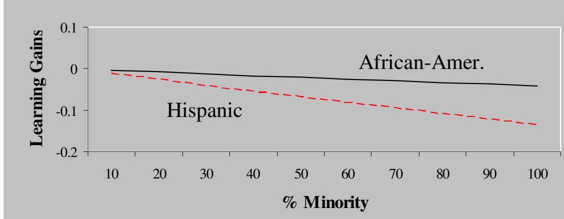
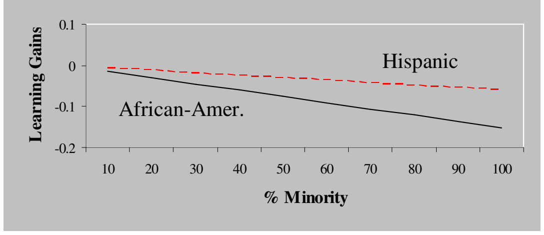


Figure 1 suggests that both African Americans and Hispanics learn less in high-minority schools. For example, African American students in schools with only 10 percent African American students learn about 0.04 standard deviations more per year than those in schools with 100 African American students. To put this in perspective, the achievement gap between whites and African Americans is about 0.8 standard deviations (see Appendix A). This means that the effect in Figure 1 is equivalent to about five percent of the achievement gap. It is important to emphasize, however, that this effect is only for a single grade and desegregating across grades would likely yield a large cumulative benefit—much larger than 0.04. On the other hand, the idea of moving a large number of students from extremely high-minority schools to extremely low minority ones is unrealistic. In the next section, I consider how large the cumulative benefits might be from a more realistic desegregation policy.

Also shown in Figure 1, the effect of a large change in percent minority for Hispanics is much greater than that for African Americans—0.14 standard deviations versus 0.04. Again, the main point here is that there is a negative relationship between percent minority and learning gains.

The lines in Figure 2 for middle school are somewhat different, but the same general pattern emerges. Both African Americans and Hispanics continue to learn more in schools with fewer minorities. We might have expected the lines to be somewhat flatter in middle school, given the evidence above from desegregation experiments, but this pattern emerges only for Hispanics.

Figure 2: Middle School, Adjusted Ach. Gains, by % Minority



At this point, it is important to discuss not only the patterns in segregation's effects, but how precise these patterns are in statistical terms, what statisticians call "statistical significance." In this respect, an estimated slope is only significant if we can say with a high degree of confidence that it is different from zero. Three of the four of estimates above (all but African Americans in elementary schools) are statistically significant by this definition. None of the high school results are statistically significant and these are therefore not shown. (See Appendix B.)

One important limitation, however, is that the regressions have so far ignored the roles of poverty and peer achievement that the previous discussion emphasized as being the underling sources of racial composition effects. Estimates of the separate effects of these factors are shown in Appendix B. They show, as expected, that a substantial portion of the "racial composition" effect is really due to poverty and peer achievement. Unlike the results from the above figures, there is apparently no direct effect of racial composition on either African Americans or Hispanics.

Instead, the role of racial composition really appears to arise through poverty and peer achievement. Poverty is negatively and significantly associated with learning in five of the six cases. In addition, the estimated poverty effects are considerably larger than the racial composition effects reported in Figures 1-2. For example, the -0.14 effect of racial composition on Hispanic elementary students, reported in Figure 1, is considerably smaller than the -0.20 effect of poverty reported in the Appendix (Table 2B). The general importance of poverty holds for both racial groups.

Finally, having peers with higher initial achievement is positively associated with learning gains for African Americans and Hispanics and these results are generally statistically significant. For example, increasing the peer achievement level by one standard deviation for African American elementary school students is associated with a roughly 0.1 standard deviation increase in learning gains.[42] The effects are larger and more consistently significant for Hispanics compared with African Americans.

Overall, these results suggest that there is relationship between racial composition and learning gains and that this reflects an indirect effect of other factors such as poverty and peer achievement. In other words, it is not race per se that affects learning, but the conditions under which minority students are raised and the characteristics of their classmates. For both reasons, integrated schools seem to benefit minority students.

Simulating the Effects of Controlled Choice in Louisville and Seattle

The above findings provide a basis for understanding the effects on achievement from specific desegregation policies. In order to illustrate the implications of the results, I use this evidence to estimate how much desegregation policies influence achievement for minorities in the two school districts that are the subjects of the upcoming Supreme Court decision—Louisville and Seattle.

In these two districts, it is useful to consider what would happen if the Supreme Court rejected controlled choice and the districts were no longer allowed to consider race in school assignment. This thought experiment involves two key questions: First, to what degree would schools resegregate in the absence of controlled choice? Second, how much would this resegregation influence student achievement?

To help answer the first question, the first two rows of Table 1 provide information about the racial composition of the respective districts and the minority "exposure rate." This last number, commonly used in studies of segregation, refers to the school percent minority for the average minority student. A district is perfectly integrated if the percent minority for the whole district is the same as the exposure rate. Alternatively, a completely segregated district yields an exposure rate of 100 percent.

The 2001 data show that the Louisville district was remarkably well integrated. In elementary schools, for example, there was only a four-point difference between the district percent minority and the exposure rate (40.7 versus 36.7). In Seattle, there was more segregation, due partly to the fact that 2001 was the first year that controlled choice was adopted.

The last row of Table 1 shows a hypothetical exposure rate that, in each district, is 15 points above the 2001 actual exposure rate. This change, while hypothetical, is intended to keep the exercise simple and to be fairly conservative about the effects of controlled choice on integration.

Table 1: Race in Louisville and Seattle

	Louisville, KY			*Seattle, WA*		
	Elem	*Middle*	*High*	*Elem*	*Middle*	*High*
District Percent Minority	36.7	33.9	29.3	23.5	31.7	29.7
Exposure Rate:						
w/ Controlled Choice	40.7	38.5	35.2	53.1	40.4	36.5
w/o Controlled Choice (hypoth.)	55.7	53.5	50.2	68.1	55.4	41.5

What would the above change in the exposure rate have on minority achievement? The regression estimates shown in the previous figures are based on the 23 states in the NCLB database. While Kentucky and Washington are not among the states in the database, it is reasonable to expect that data from other states provide estimates that can be applied to these excluded states. Indeed, considering other states is advantageous because they encompass a large number of other urban areas that are similar to Louisville and Seattle. Therefore, the simulations below assume that the effects of racial composition for the NCLB database also apply to these two districts.

The results of the simulation, shown in Table 2, represent the learning loss from the higher exposure rate for minority students. These learning losses are "cumulative" in that they assume such gains would occur in each of 13 grades, kindergarten through grade 12. This is calculated by multiplying the change in the exposure rate by the regression estimates and then adding the annual gains across grade levels.[43] As noted above, the regression results shown in Figures 1-2 were simplified and did not incorporate the roles of poverty and peer achievement. In the simulation results, I use the regressions that simultaneously consider these other important factors as well as race.

The results of this exercise, shown in the last row of Table 2 suggest that minority students in Louisville and Seattle would lose considerable ground in achievement if controlled choice were eliminated—as much as 1.343 standard deviations, a level that researchers would generally consider to be quite large.[44] To put this in perspective, these estimates suggest that the achievement gap could be completely eliminated through desegregation. These large effects remain regardless of the handling of statistical significance (see the difference between the Type #1 and Type #2 analyses[45]). The results are also in line with those found by Cooley (2006). Recall that her estimates imply a cumulative achievement effect of as much as 0.65 standard deviations, which is roughly the average of the estimates in Table 2.

Table 2: Simulated Effects of Eliminating Controlled Choice in Louisville and Seattle

Simulation Type	*Louisville, KY*	*Seattle, WA*
Type #1 Analysis		
African American	-0.449	-0.778
Hispanic	-0.218	-0.314
Type #2 Analysis		
African American	-0.937	-1.343
Hispanic	-0.260	-0.368

Nevertheless, there are reasons to think that these estimates overstate the true effects.[46] For example, these estimates probably still reflect some degree of selection bias, which would tend to inflate the estimates. Also, the earlier desegregation experiments found positive but smaller effects. Nevertheless, while there are still questions about the size of the effects, these results point toward the same conclusion as those of past studies. In short, minorities gain from desegregation. For the purpose of the Supreme Court decision, this further suggests that the government does have a compelling interest in integrating schools by race.

Implications for Controlled Choice and the Supreme Court Decision

In the Supreme Court's decision in *Brown*, Chief Justice Warren wrote that "we must look instead to the effect of segregation itself on public education." In this study, I have followed that advice, taking a fresh look at decades of the most rigorous research and providing important new evidence using the NCLB data. The number of students and schools included in these new data is, to my knowledge, the largest ever included in an analysis of segregation.

Each type of evidence points to the conclusion that minority students make greater learning gains in schools with more white peers—not because of race itself, but because of the economic and academic advantages of students in these schools and the important influence that classmates have on minority learning. While the sizes of these effects are still somewhat unknown, it is difficult to dispute that the effects exist and that the government has a compelling interest in addressing them.

The evidence that desegregation improves student outcomes is arguably stronger than evidence on other major systemic reforms recently considered. In the 1990s, school systems tried to decentralize and de-bureaucratize urban schools where a high percentage of minorities attend school. This was followed by a wave of test-based accountability and school choice programs, such as charter schools and vouchers, aimed at the same groups of students. While there is some evidence that these reforms have some small benefits for minority students, the effects appear much smaller and less consistent than those of desegregation.[47] It is worth continuing some of these new policy experiments in order to learn more about their long-term effects, but there is little evidence to date that even the broad application of accountability and school choice would have the same effects as desegregation.

The recent history of the Court's decision has made it much more difficult for lower courts and school districts to pursue desegregation, however. Lower courts cannot require school districts to desegregate across district boundaries and they are limited in the ways they can require desegregation within districts. There is only one main option left—controlled choice implemented by school districts without court intervention. The decision before the Court is to determine whether this last remaining option will be allowed to stand. In making their ruling, the Supreme Court justices should know that racial integration is as essential to providing equal educational opportunity today as it was when Justice Warren announced the Court's landmark *Brown* position in 1954.

Appendix A: Methodology

This appendix describes the NCLB data in more detail. The differences in tests and test scales across states, mentioned in the main text, represent the first topic. This is followed by a summary of choices I made regarding the grades and years of the test scores, the number of observations, and the interpretation of the magnitude of the differences in test scores. Tables related to these topics follow.

At the outset, it is important to emphasize that the focus of the entire analysis is on math scores because it is widely believed by researchers that schools have greater influence over math compared with reading scores, which are affected more by students' home environments. Specifically, students vary in the degree to which they read books at home and in their parents' use of verbal language, both of which influence student reading scores.

Test Score Differences Across States (Tables 1A)

The discussion in the text indicates that it is difficult to make comparisons across states because each state uses a different test. This section explains in more detail which test score data was used in each state and, where there were multiple options, why I focused on particular test scores.

NCLB requires that scores be reported in terms of "percent proficient" and nearly all states do this, but it is difficult to use these to calculate learning gains. To see the problem, consider a group in which 75 percent of students are proficient in grade three and suppose that this number increases to 80 at grade four. We could subtract these two and conclude that there was a five percentage point "gain." But this is somewhat misleading. It could be that the students who moved above the bar had been just below the bar previously, implying that little learning took place. In addition, some of the students who were above the bar, may have actually declined, but yet still stayed within the proficiency definition. In that case, the apparent five point gain might actually reflect a loss in learning. An even more fundamental problem is that it is difficult to define proficiency in comparable ways across grades. Therefore, using proficiency scores is avoided where possible.

Unfortunately, for 12 states, proficiency scores are the only ones available and learning gains are calculated as follows. To take the simplest example, consider a state, such as Ohio, where only one level of proficiency is reported. In this case, the number of proficient students is calculated for the entire state; this is translated into a percentile and, from there, into Normal Curve Equivalent (NCE)—the NCE of a student scoring at exactly the cutoff score.

A challenge arises at this point because there is no information about the distribution of students above and below the cutoff. There are at least two possible approaches to addressing this. The first is to calculate the midpoint NCE between each cutoff score. For example, suppose with only one cut-score the is 25 NCEs. Those above the band are assumed to average (25-0)/2=12.5 NCEs and those above are assumed to average (100-25)/2=62.5. Further, suppose that a school has 50 percent of its students above and below the cut-off. The school's average NCE would then be (0.5*12.5)+(0.5*62.5)=37.5. This approximation probably does not make a difference in the majority of proficiency-only states where there are three or more cutoffs, but it may in states, such as Ohio, with a smaller number of cutoffs. It is also important to recognize that while this simpler approach may bias the achievement level, this same bias should arise for other scores and therefore cancel out in the calculation of gains.

Fortunately, 10 states report scores in some other form. The preferred approach is to use national NCEs or percentiles. There are two reasons for this. First, these represent the average performance of students in each subgroup and therefore avoid the problem with proficiency scores that one cannot determine whether students are just above or below the cutoff. Note that percentile scores can be easily converted to NCEs and are therefore interchangeable. It is also straightforward with NCEs to calculate an "effect size" based on the student-level standard deviation (see below).

Scale scores are also preferred to proficiency scores. Like percentiles and NCEs, they capture the learning level of every student. These scores are also designed so that, at least in theory, a 10-point gain means the same thing no matter where the student starts off. The disadvantage is that it is difficult to estimate the student-level standard deviation. For this reason, NCEs and percentiles are still preferred. Further discussion on the calculation, usage, and importance of the standard deviation is provided below.

Finally, it is important to emphasize again that even when states report scores in similar ways, or where they can be converted to appear similar, the scores are still based on different underlying tests and therefore are still not completely comparable across states.

Choice of Grades and Years (Table 1A)

Table 1A also shows that achievement gains are calculated in most states by comparing one cohort in one year to the same cohort in the next grade in the subsequent year. An advantage of single year gain is that students move from school to school across years, changing the composition of the student cohorts. This is likely to produce some measurement error, though it probably does not significantly bias the results.[48] In any event, because mobility is more likely across multiple years, I use one-year changes where possible to minimize the potential problem. The calculation of annual gains is not possible in seven states with the NCLB data. In these cases, I compare, for example, the third grade cohort in 2003 with the same cohort in 2005 two years later when the same students have reached the fifth grade.

Among the states where annual gains can be calculated, it is also possible to calculate gains for multiple cohorts within the same school. In these cases, I chose the lowest grades (e.g., the third grade cohort over the fourth grade cohort). This choice is fairly arbitrary and there are only a few states where the issue arises.

Number of Observations (Table 2A)

Table 2A provides information about the number of observations and compares this information to NCES data regarding the characteristics of schools in the state as a whole. This comparison provides insight into how well the NCLB data represents the entire picture of learning in the respective states. While there is no reason to expect that schools are systematically included or omitted from the NCLB data, the potential significance of such non-random selection is certainly higher as the number of excluded schools increases.

The "total number of schools in the state" comes from the NCES Common Core of Data. The "total number of useable observations" is the number of schools in the NCLB data that provides all the necessary information to carry out the analysis for at least one subgroup. In all but one case (Tennessee), the number of schools reported to be in the state is greater than the number in the NCLB data. This is expected, given that the NCLB data include some missing observations. Most of the missing observations are due to an insufficient amount of testing information in certain grade levels. For example, the state with the greatest gap between reported and useable observations, Louisiana, is reported to have 1,510 schools, but only 116 report information sufficient for this analysis. This is mainly due to the fact that elementary and middle schools are excluded from the state's available data.

Understanding the Magnitude of the Differences in Learning

When considering any educational program, whether desegregation or class size reduction, it is important to understand the size of the effects in absolute terms and in comparison with alternative programs.

To allow for such comparisons, education researchers typically report the "effect sizes" of educational programs. This requires calculating the standard deviation of the outcome measure (in this case student test scores) and then dividing the change in scores by this standard deviation. A small effect size is considered to be between 0.10 and 0.19 standard deviations, while medium and large effects sizes are on the order of 0.20-0.29 standard deviation and greater than 0.30 standard deviations, respectively. In the text, I also compare the effects to the overall size of the achievement gap between minority and white students.

An important complication in calculating the effect sizes in this case is that the data is not reported for individual students, but rather for racial subgroups. This is important because standard deviation depends on the level of aggregation of the data. When we average test scores for students into subgroup averages, we reduce, or "average out," a substantial part of the variation and therefore reduce the standard deviation. The research standards discussed above are based on student-level data where standard deviations are higher. This implies, further, that using aggregated data—without adjustments—will make the effect sizes appear larger than they are.

To account for this, I estimated the student-level standard deviation using data mentioned earlier regarding the size of the achievement gap between minority groups—that is, the difference in achievement levels between different racial/ethnic groups. I first calculated the raw gap between African Americans and minorities using whatever data were available in the respective state. I then assumed that the "real" achievement gap in each state is 0.80 standard deviations, which is a conservative estimate based on other evidence from SAT, ACT, and other tests (Harris and Herrington, 2006). Finally, I solved the following equation for the standard deviation *sd* in each state *i*.

$$\frac{RawGap_i}{sd_i} = 0.80$$

An alternative possible approach would be to adjust the across-group standard deviations based on evidence that the student-level standard deviation is two to four times larger than across school standard deviation. Unfortunately, in some states, this approach yields some implausibly large achievement gaps across racial groups.

Table 1A: Information on State Tests and Learning Gains, by State

State	*Score Type*	*Score Set #1*		*Score Set #2*		*School Levels for Gains*
		Year	*Grade(s)*	*Year*	*Grade(s)*	
Alabama	Perc	2002	3	2003	4	E
Alaska [b]	---	---	---	---	---	---
Arizona	Scale	2003	3	2005	5	E
Arkansas	Prof (3)	2004	4, 6	2005	5, 7	E, M
California	Scale	2003	3,7,9	2004	4,8,10	E, M, H
Colorado	Prof (3)	2004	5,7,9	2005	6,8,10	E, M, H
Connecticut	Scale	2003	4,6	2005	6,8	E, M
Delaware	Scale	2003	3	2005	5	E
Florida	Scale	2004	3,7,9	2005	4,8,10	E, M, H
Georgia	Scale	2004	3,7	2005	4,8	E, M
Hawaii [b]	---	---	---	---	---	---
Idaho [b]	---	---	---	---	---	---
Illinois	Prof (3)	2002	3	2004	5	E
Indiana	Prof (2)	2004	3,6	2005	4,7	E, M
Iowa [a]	---	---	---	---	---	---
Kansas [a]	---	---	---	---	---	---
Kentucky [a]	---	---	---	---	---	---
Louisiana	Scale	2004	10	2005	11	H
Maine	---	---	---	---	---	---
Maryland	Prof (1)	2004	3,7	2005	4,8	E, M
Massachusetts [a]	---	---	---	---	---	---
Michigan [a]	---	---	---	---	---	---
Minnesota	Prof (4)	2003	3	2005	5	E
Mississippi	NCE	2001	4,7	2002	5,8	E, M
Missouri [a]	---	---	---	---	---	---
Montana [a]	---	---	---	---	---	---
Nebraska [a]	---	---	---	---	---	---
Nevada	---	---	---	---	---	---
New Hamp. [a]	---	---	---	---	---	---
New Jersey [a]	---	---	---	---	---	---
New Mexico	Prof (3)	2004	3,7	2005	4,8	E, M
New York [a]	---	---	---	---	---	---
North Carolina	Scale	2004	3,7	2005	4,8	E, M
North Dakota [b]	---	---	---	---	---	---
Ohio [c]	Prof (1)	2003	4,9	2004	10	E, M, H
		2004	6	2005	6,7	
Oklahoma [a]	---	---	---	---	---	---
Oregon	---	---	---	---	---	---
Pennsylvania [a]	---	---	---	---	---	---
Rhode Island [a]	---	---	---	---	---	---
South Carolina	Prof (3)	2004	3,7	2005	4,8	E, M
South Dakota [b]	---	---	---	---	---	---
Tennessee [c]	Prof (2)	2004	3,7,9	2005	4,8,10	E, M, H
Texas	Scale	2004	3,7,9	2005	4,8,10	E, M, H
Utah [a]	---	---	---	---	---	---
Vermont [a]	---	---	---	---	---	---
Virginia	Prof (2)	2003	3	2005	5	E
Washington [a]	---	---	---	---	---	---
West Virginia [b]	---	---	---	---	---	---
Wisconsin [a]	---	---	---	---	---	---
Wyoming [a]	---	---	---	---	---	---

Notes: "Test types" are defined as follows: "Prof (#)"= proficiency with the number of separate cut scores in parentheses; "Scale" = scale score; "NCE" = Normal Curve Equivalent"; "Perc"=percentile ranking converted to NCE. "Score set #1" and "Score set #2" refer to the two sets of test score data from which the cohort achievement gains are calculated (see earlier discussion). Each set comes from a specific year and uses specific grades. The grades reported include all those available in the NCLB data, though some these are not used in the analysis. The last column indicates the school levels for which it is possible to calculate learning gains: “E” = elementary; “M”=middle; and “H”=high school.

[a] It is not possible to estimate cohort achievement gains in this state.

[b] Even though it is possible to calculate gain scores, this state is dropped due to an insufficient number of African American or Hispanic students (see Table 1B).

[c] This state did not report the number of students tested in each grade. Weighting of subgroup observations was therefore based on values imputed from the number of students in the school, the number of grades in the school (yielding number of students per grade), and the percentage of students in each racial category.

Table 2A: Student Race Information, by State

State	*% AA*	*% Hisp.*	Total # Schools in State (CCD)	*# Elem. Schools w/ racial subscores*				*# Middle Schools w/ racial subscores*				*# High Schools w/ racial subscores*			
				Total	*AA*	*Hisp.*	*White*	*Total*	*AA*	*Hisp.*	*White*	*Total*	*AA*	*Hisp.*	*White*
Alabama [a]	39.4	1.4	1,516	708	453	24	526	371	232	17	299	0	0	0	0
Arizona	4.8	34.0	1,570	692	34	458	560	0	0	0	0	0	0	0	0
Arkansas	22.4	3.6	1,109	514	348	318	499	101	49	14	95	0	0	0	0
California	8.1	40.4	8,343	4,564	925	3,557	3,050	1,611	660	1,337	1,335	1,027	491	908	861
Colorado [a]	4.8	23.2	1,560	286	4	53	268	343	46	201	324	264	34	150	247
Connecticut	14.6	13.2	1,104	154	41	32	100	171	57	59	147	0	0	0	0
Delaware	31.3	6.5	201	50	33	3	46	0	0	0	0	0	0	0	0
Florida	28.0	17.2	3,111	1,718	1,153	919	1,397	679	536	446	609	474	389	327	435
Georgia	49.6	5.4	1,843	1,076	818	225	799	427	376	150	367	0	0	0	0
Illinois	18.8	11.5	4,302	1,822	668	465	1,339	0	0	0	0	0	0	0	0
Indiana	11.5	3.7	1,958	1,068	250	92	1,001	276	96	61	266	0	0	0	0
Louisiana	50.1	1.3	1,510	0	0	0	0	0	0	0	0	0	0	0	0
Maryland	37.8	4.9	1,357	820	637	197	612	273	255	130	225	0	0	0	0
Minnesota	8.0	4.0	2,348	694	134	55	635	0	0	0	0	0	0	0	0
Mississippi	54.5	0.1	1,015	330	253	1	202	239	174	1	185	0	0	0	0
New Mexico	1.9	51.3	746	71	2	64	44	26	4	22	20	0	0	0	0
North Carolina	33.5	5.2	2,106	1,192	897	476	1,087	566	458	306	533	0	0	0	0
Ohio	16.4	1.8	3,852	731	230	14	598	568	230	23	473	686	172	34	654
South Carolina	45.5	2.2	1,101	552	452	0	444	254	243	0	223	0	0	0	0
Tennessee	22.3	1.3	1,589	771	276	50	660	243	141	42	213	41	18	4	37
Texas	14.1	39.5	7,228	3,526	1,602	2,974	2,471	1,650	834	1,305	1,424	1,394	609	1,032	1,177
Virginia	27.8	5.0	1,918	1,043	592	143	935	0	0	0	0	0	0	0	0
Totals			88,984	22,382	9,802	10,120	17,273	7,798	4,391	4,114	6,738	3,886	1,713	2,455	3,411

Notes: As defined by NCES, the "total number of schools" refers to regular schools, excluding alternative schools and other special schools.

Appendix B: Regression and Simulation Analyses

The results regarding the effects of segregation are based on regression estimates of equation (1):

$$A_{ijst} = \beta_{1s} A_{ijst-1} + \beta_{2s} A^2_{ijst-1} + \beta_3 X_{jst} + \beta_4 M_{jst} + \varepsilon_{js} \qquad (1)$$

where A_{ijst} is the achievement of student subgroup i in school j in state s at time t. All versions of equation (1) are estimated separately for each racial subgroup.
Lagged and squared lagged values of achievement are independent variables. The coefficients β_{1s} and β_{2s} are indexed by s to indicate that they are allowed to vary only by state. The other coefficients are not indexed and are not allowed to vary by state.

An alternative specification, used as part of the sensitivity analysis, involves ignoring the scaling issue and using the simple change in scores, A_{ijst}-A_{ijst-1}, as the dependent variable. The advantage of this specification is that it avoids serial correlation in the error terms that is common when including a lagged value of the dependent variable on the right-hand side. Unfortunately, this specification does not account for the test scaling issues, which turn out to be important.

The set of variables X_{jst} indicate whether the school is a magnet or charter school and how many migrant students are in the school. In some specifications, as indicated below, percent poverty and peer achievement are also added.

Before discussing the simulation results, it is also important consider some differences between the present sample of schools and those of the national as whole. The NCLB database has a somewhat larger proportion of African Americans and Hispanics because of the 23 states for which data is available have higher than average percentages for these groups. This is especially true for Hispanics because California, Florida, and Texas are all included and have disproportionate shares of this group. The sample is also disproportionately from elementary schools, and to lesser extent, middle schools.

There is no reason to believe that either of the above differences between the sample and the national population is systematic in a way that the results might be biased. However, a small bias is possible as a result of the fact that schools with small percentages of any one group do not report data for those groups. This means that schools at either end of the racial composition scale—extremely integrated and extremely segregated—will tend not to report data for the small racial groups within those schools. In the present study, the difficulty is estimating the learning gains for schools with very high percentages of whites and very small percentages of minorities.

Results for Figures 1-2 (Table 1B below)

The results presented in Figures 1-2 are based on the results shown in Table 1B below. Additional estimates were made including district fixed effects. While doing so has some intuitive appeal, because the policies being considered here involve redistributing students within school districts, the point estimates are often implausibly large, perhaps because of more intensive sorting of students within districts, which compounds the selection bias problem. In any event, the estimates with district effects are not reported but are available from the author.

Adding Poverty and Peer Achievement (Table 2B)

The estimates in Table 2B include race, poverty, and peer achievement. It is important to note that peer achievement, because of the characteristics of the data, is defined as the achievement of peers of other races. For example, in the regression models focusing on African Americans, peer achievement refers to the achievement level of students who are white and/or Hispanic. Further, peer achievement is multiplied by the percentage of students in the respective minority groups so that the effect of other-race peer achievement depends on the number of other-race peers. (The peer achievement variable is therefore an interaction term and should be interpreted accordingly.)

Adding these variables reduces the sample size, mainly because peer achievement is often excluded. I have also estimated models that restrict the sample in the first set of estimates to the smaller sample of schools that have observations in the second set, though the results change little from those reported.

Because including all the variables reduces the number of observations (mainly because of missing observations on peer achievement), I reestimated Table 1B with this more limited sub-sample. These results, available from the author, show that the change in the racial composition effect is in fact due to the addition of poverty and peer achievement, not to differences in the samples. The results in Table 2B are used as the basis for the simulations.

Replicating the Peer Effects in Florida, North Carolina, and Texas (Tables 3B-5B)

While the NCLB data are unprecedented in the number of schools in which desegregation can be studied, it would be desirable to have this data for individual students, as opposed to subgroups. In Tables 3B-5B, I show the results from the peer effect studies discussed in the main text that use student-level data for North Carolina (Cooley, 2005) and Texas (Hanushek et al., 2002; Hoxby, 2000). I then attempted to "replicate" the results of these studies, applying similar specifications to the data for the respective states. If the results are the same, this provides some evidence that the results found with the NCLB are not due to the use of racial subgroups (as opposed to student-level data) and that the results with the NCLB data might be interpreted as causal.

The results with the NCLB data are relatively similar, though weaker, compared with Cooley's findings in North Carolina. Cooley included both own-race and other-race peer achievement. Only the achievement of other-race peers can be measured with the NCLB data and the effects of this variable are similar across the two studies (positive, but marginally significant or marginally significant). The positive and significant effects on other-race peer achievement for white students in the replication probably reflect a correlation between other-race peer achievement and own-race peer achievement, the latter of which is omitted. The effect of racial composition in North Carolina is also somewhat larger for white students in the Cooley study. It is unclear why this would be the case, but combined with the results on peer achievement, suggests that these results may understate peer effects.

Before comparing the results of this study with the previous Texas peer effect studies, it is important to note an inconsistency between the previous studies themselves.

Hoxby (2000) finds that racial composition affects both whites and African Americans (and Hispanics), while Hanushek finds that the racial composition effect is insignificant for whites. (The closest specification in Hanushek et al. is not shown in the tables, though the effects of racial composition are essentially the same as those reported in Table 5B when peer achievement is added.) The only apparent difference in specification between the two is that Hanushek et al. control for both percentages of African American and Hispanic, while Hoxby considers only the former. This could explain the difference, although there is no obvious reason to think that it would. The results with the NCLB data more closely approximate the Hoxby results in this respect.

Table 4B compares the Hoxby results with the NCLB subgroup replications. The two sets of results are nearly identical with regard to the effect of percent African American on African Americans, though the results for Hispanics differ; specifically, the effects of percent Hispanic on each racial subgroup is negative and significant in each replication, but never significant in the Hoxby estimates.

Table 5B presents the results from the first specification used by Hanushek et al. The results are nearly identical for African Americans; the sole exception is that the negative effect of percent Hispanic on African Americans is negative and significant in replication, but negative and insignificant in the Hanushek et al. results. The results are also similar for Hispanics where, in both studies, there is no evidence of peer effects for Hispanics. The results are somewhat more different for white students where, while nearly all of the estimates are insignificant, the point estimates are frequently of different signs, only one of which is the replication significant.

Overall, the results with the NCLB data are similar to those in the three peer effect studies for African Americans. If anything, the NCLB results may understate the effects of racial composition on this group. The opposite seems to be true of Hispanics where the NCLB data suggests that racial composition is important, but the peer effects studies find no such effect.

Results by Test Score Type

In the text and in Appendix A, I describe the differences in test scores across states, as well as the desirability of using percentiles, NCEs, and scale scores. However, in 12 states, the NCLB data only allowed analysis of test score proficiency. Therefore, I reestimated the models only for those states which use the three preferred types of test scores: Alabama, Arizona, California, Connecticut, Delaware, Florida, Georgia, Louisiana, Mississippi, North Carolina, and Texas.

The effects of percent minority on African Americans become more precise at all grade levels. However, the results for Hispanics and whites are essentially unaffected. This is not at all surprising in the case of Hispanics because a large percentage of these students are in the states that are in both sets of estimates—particularly, California, Florida, and Texas. These results are available from the author upon request.

Table 1B: Regression-Estimated Effects of Percent Minority

	African Americans			Hispanics		
	Elem	Middle	High	Elem	Middle	High
% Minority	-0.0403	-0.1510***	-0.0484	-0.1355***	-0.0342***	0.0184
	(0.0956)	(0.0439)	(0.0421)	(0.0132)	(0.0131)	(0.0230)
Observ.	9,559	4,270	1,076	9,891	4,011	1,397
R^2	0.5370	0.7206	0.7978	0.4513	0.7555	0.7440

Source: Author's analysis of NCLB database using equation (1). Dependent variable is second-year test score level for African Americans/Hispanics in elementary, middle, and high schools, respectively. Independent variables include student test scores for same cohort in previous year and grade (and its square), school percent migrant students, and school status as charter or magnet school. Robust standard errors in parentheses. Sample weights based on number of students in the school. Significance levels indicated by: * = 0.10, ** = 0.05, and *** = 0.01.

Table 2B: Regression Results With Poverty and Peer Achievement

	African Americans			Hispanics		
	Elem	Middle	High	Elem	Middle	High
% Minority	-0.0259	0.0346	0.0571	0.0468	0.0261	0.0552*
	(0.0423)	(0.0257)	(0.0397)	(0.0297)	(0.0268)	(0.0319)
% Poverty	0.0624	-0.1997***	-0.2693***	-0.1979***	-0.0857**	-0.0655*
	(0.0551)	(0.0354)	(0.0596)	(0.0240)	(0.0341)	(0.0353)
Peer Achieve.	0.0023***	0.0008***	0.0006	0.0005**	0.0007**	0.0003*
	(0.0004)	(0.0003)	(0.0006)	(0.0003)	(0.0003)	(0.0004)
Observ.	7,713	3,754	950	7,068	3,488	1,224
R^2	0.5632	0.7259	0.7836	0.4693	0.7468	0.7231

Notes: See notes to Tables 1B.

Table 3B: Replication of Cooley (2005)Peer Effect Study of North Carolina
(Racial grouping in brackets; elementary schools only)

Variable	*Cooley [Minor.]*	*Replicat. [AA]*	*Cooley [White]*	*Replicat. [White]*
% AA	---	Neg.	---	Neg.*
% Minority	Neg.***	---	Neg.***	---
Own-race peer ach.	Pos.*	---	Pos.***	---
Other-race peer ach.	Pos.	Pos.	Pos.	Pos.*

Table 4B: Replication of Hoxby (2000)Peer Effect Study of Texas
(Racial grouping in brackets; elementary schools only)

Variable	*Hoxby [AA]*	*Replicat. [AA]*	*Hoxby [Hisp.]*	*Replicat. [Hisp.]*	*Hoxby [White]*	*Replicat. [White]*
% AA	Neg. **	Neg.**	Neg.**	Neg.***	Neg.**	Neg.**
% Hisp.	Neg.	Neg.**	Pos.	Neg.***	Neg.	Neg.***

Table 5B: Replication of Hanushek et al. (2002)Peer Effect Study of Texas
(Racial grouping in brackets; elementary schools only)

Variable	*Hanush. [AA]*	*Replicat. [AA]*	*Hanush. [Hisp.]*	*Replicat. [Hisp.]*	*Hanush. [White]*	*Replicat. [White]*
% AA	Neg. **	Neg.**	Neg.	Neg.	Neg.	Neg.
% Hisp.	Neg.	Neg.**	Pos.	Neg.	Pos.	Neg.*
Avg. peer ach.	Neg.	---	Pos.	---	Pos.	---
Other-race peer ach.	---	Neg.	---	Pos.	---	Neg.

Notes: Each "replication" is based on NCLB data for the respective state.

About the Author

Douglas N. Harris has a Ph.D. in economics and is an Assistant Professor of Education and Economics at Florida State University. He will start in the same position at the University of Wisconsin-Madison in January. His research on educational equity and accountability has been published by the *American Journal of Education* and presented at recent meetings of the *Harvard Achievement Gap Initiative* and the *UNC Center for Civil Rights*. In addition to equity, his research includes topics such as teacher quality, accountability, school finance, and the relationship between education and economic competitiveness. Dr. Harris is currently leading a team of researchers in a USDOE-funded study to investigate approaches to training and selecting teachers and a separate project to evaluate the effects of NBPTS teacher certification using value-added modeling (funded by NBPTS). His research has been published in academic journals and books such as *American Journal of Education*, *Brookings Papers on Educational Policy*, *Economics of Education Review*, *Education Economics*, *Educational Policy* and the *Peabody Journal of Education*. His research has also been cited in national publications such as *Education Week*, *The New York Times*, *The New York Times Magazine*, and *The Washington Post*. He is also a parent and has been a school board member and active member of local community organizations. For more information, he can be reached at harris@coe.fsu.edu.

Acknowledgements

The author gratefully acknowledges funding support from the Center for American Progress, data assistance from the American Institutes for Research, and helpful comments from Cynthia Brown, Cassandra Butts, Linda Chard, Charles Clotfelter, Derek Douglas, Debbi Harris, Sara Goldrick-Rab, Richard Kahlenberg, Sean Reardon, Steve Rivkin, Neera Tanden, and participants in the AERA workshop on controlled choice held in Washington, D.C. during September 15-17, 2006.

Notes

[1] Schools were segregated by law in all Southern states (Alabama, Georgia, Florida, Mississippi, North Carolina, South Carolina, Tennessee, Virginia), six other border states (Delaware, Kentucky, Maryland, Missouri, Oklahoma, and West Virginia), three other states (Arizona, Kansas, and New Mexico), and the District of Columbia. Schools were also *de facto* segregated even in states that did not legally require it. See Clotfelter, C. (2004). *After Brown: The Rise and Retreat of School Desegregation*. Princeton, NJ: Princeton University Press.

[2] "Racially segregated" is defined here as having 90 percent or more students who are African American. The cited numbers are from Clotfelter (2004).

[3] Harris, D.N. & Herrington, C.D. (2006). Accountability, standards, and the growing achievement gap: Lessons from the past half-century. *American Journal of Education*, 112(2), 209-238.

[4] These achievement gap trends are based on the National Assessment of Educational Progress (NAEP) as discussed in: Harris and Herrington (2006) and Harris, Douglas N. (forthcoming a). "K-12 Education Policies and the Equity of Educational Outcomes: From Desegregation to Accountability," In Helen Ladd and Ted Fiske (Eds) *Handbook of Educational Policy*.

[5] Orfield, Gary (2001). *Schools More Separate: Consequences of a Decade of Resegregation*. Cambridge, MA: The Harvard Civil Rights Project.

[6] The Seattle program was adopted in 2001, while Louisville's was adopted in 1998. Also, the city of Louisville is located in Jefferson County, KY. The case is being brought against the county school district. The respective suits are: *Parents Involved v. Seattle School District* and *Meredith v. Jefferson County Board of Education*.

[7] Greenhouse, Linda (2006). "Court to Weight Race as a Factor in School Rolls," *The New York Times*, June 6, 2006.

[8] Greenhouse, Linda (2006).

[9] Savage, David G. (2006). "Bush administration opposes integration plans," *The Los Angeles Times*, August 25, 2006.

[10] The original proposal by the administration did not include test score breakdowns by racial sub-group, but this was added as part of the legislative process.

[11] The author thanks Dr. William Trent for a conversation that led to this point.

[12] These two legal standards are elements of strict scrutiny, which courts apply in cases where constitutionality of a government action is in question. There is some debate about whether strict scrutiny should apply, but it is likely that the Court will apply this standard and the legal discussion in the text is based on this assumption.

[13] One of the oldest and most well known studies on this topic is: Coleman, J. (1966). *Equality of Educational Opportunity*. Report OE-38000. Washington, D.C.: U.S. Department of Health Education and Welfare, Office of Education.

[14] See Clotfelter (2004). This same source indicates that the level of segregation in 2000 was roughly the same as it was in 1975, suggesting that a majority of all school desegregation over the past half-century occurred during the brief period, 1968-1972.

[15] In 1970, roughly one-third of desegregation was between districts and two-thirds was within-districts. Partly because of the Court's decision in *Milliken*, white families moved to the suburbs where there was no real possibility of racially diverse schools. As a result, by 2000, more than two-thirds of segregation was between districts. See Clotfleter (2004) Table 2.4.

[16] Kahlenberg, R. (2000). *All Together Now: Creating Middle Class Schools Through Public School Choice*. Washington, D.C.: Brookings Institution Press.

[17] Willie, C.V. and Alves, M.J. (1996). *Controlled Choice: A New Approach to School Desegregated Education and School Improvement*. New England Desegregation Assistance Center. Providence, RI: Brown University.

[18] The NIE studies are:

Armor, D. (1984). "The evidence on desegregation and Black achievement," In T. Cook, D. Armor, R. Crain, N. Miller, W. Stephan, H. Walberg, & P. Wortman (Eds.), *School Desegregation and Black Achievement* (pp.43-67). Washington, D.C.: National Institute of Education

Cook, T. (1984), "What have black children gained academically from school integration? Examination of the meta-analytic evidence." In T. Cook, D. Armor, R. Crain, N. Miller, W. Stephan, H. Walberg, & P. Wortman (Eds.), *School Desegregation and Black Achievement* (pp.6-42). Washington, DC: National Institute of Education.

Crain, R.L. (1984). Is nineteen really better than ninety three? Washington D.C.: National Institute of Education.

Miller, W. & Carlson, M. (1984). "School desegregation as a social reform: A meta-analysis of its effects on Black academic achievement." In T. Cook, D. Armor, R. Crain, N. Miller, W. Stephan, H. Walberg, & P. Wortman (Eds.), *School Desegregation and Black Achievement* (pp.89-130). Washington, D.C.: National Institute of Education.

Stephan, W.G. (1984). "Black and Brown: The effects of school desegregation on Black students. In T. Cook, D. Armor, R. Crain, N. Miller, W. Stephan, H. Walberg, & P. Wortman (Eds.), *School Desegregation and Black Achievement* (pp.131-159). Washington, D.C.: National Institute of Education.

Walberg, H.J. (1984). "Desegregation and education productivity." In T. Cook, D. Armor, R. Crain, N. Miller, W. Stephan, H. Walberg, & P. Wortman (Eds.), *School Desegregation and Black Achievement* (pp.160-193). Washington, D.C.: National Institute of Education.

Wortman, P.M. (1984). School desegregation and Black achievement: An integrative review. In T. Cook, D. Armor, R. Crain, N. Miller, W. Stephan, H. Walberg, & P. Wortman (Eds.), *School Desegregation and Black Achievement* (pp.194-224). Washington, D.C.: National Institute of Education.

[19] Schofield, J.W. (1995). "Review of research on school desegregation's impact on elementary and secondary school students." *In Handbook of Research on Multicultural Education*, ed. J.A. Banks and C.A.M. Banks. New York: Macmillan Publishing.

[20] Crain and Maynard (1983) focus on 21 longitudinal studies involving random assignment of students to control and treatment groups. Crain, Robert L. and Rita E. Maynard (1983). "The effect of research methodology on desegregation-achievement studies: A meta-analysis," *American Journal of Sociology*, 88(5), 839-854.

[21] This is based on several sources: Cook (1984), Shofeld (1995), and Stephan (1984). This assumes of course that desegregation actually takes place. If voluntary desegregation is simply a means of maintaining segregation then this form will obviously be ineffective.

[22] Bradley, L.A. and G.W. Bradley (1977). “The academic achievement of Black students in desegregated schools: A critical review,” *Review of Educational Research*, 47, 399-449.

[23] This comes from several sources: Cook (1984), Crain (1984), and St. John, N.H. (1975). School desegregation: Outcomes for children. New York: Wiley.

[24] Crain and Mahard (1983).

[25] The Florida study is: Burke, M.A. & Sass, T.R. (2006) “Classroom peer effects and student achievement,” *unpublished manuscript*.

[26] Cooley, J. (2006). “Desegregation and the achievement gap: Do diverse peers help?” *unpublished manuscript*.

[27] The North Carolina study, as well as another in Texas, also found that the race peer effects were largest for African American students who initially had high scores, but the opposite was true for whites. The Texas study is: Hanushek, E.A., Kain, J.F. & Rivkin, S.G. (2002) “New evidence about Brown v. Board of Education: The complex effects of racial composition on achievement,” *Working Paper no.W8741*. Cambridge, MA: National Bureau of Economic Research.

[28] Hanushek et al. (2002); Hoxby, Caroline M. (2000). “Peer effects in the classroom: Learning from gender and race variation.” Working Paper #7867. Cambridge, MA: National Bureau of Economic Research.

[29] Hoxby (2000).

[30] For evidence on the relationship between test scores and wages, see: Neal, D.A. & Johnson, W.R. (1996). “The Role of Premarket Factors in Black-White Wage Differential,” *Journal of Political Economy*, 114, pp. 869-95; and Currie, J. & Thomas, D. (1999). “Early Test Scores, Socioeconomic Status and Future Outcomes,” NBER Working Paper No.6943, Cambridge, MA: National Bureau of Economic Research.

[31] For evidence on the relationship between years of schooling and wages, see: Krueger, A.B. (2003). “Economic Considerations and Class Size,” *Economic Journal*, 113, pp. 34-63.

[32] Wells, A.S. and Crain, R.L. (1994). “Perpetuation theory and the long-term effects of school desegregation,” *Review of Educational Research*, 64(4), 531-555.

[33] Armor, D. (1972). “The evidence on busing,” *Public Interest*, 28, 90-124.
Armor compares the treatment and control groups up to two years after high school graduation. Interestingly, while there were effects on college enrollment, there were no apparent effects on achievement, reinforcing the need to go beyond achievement when looking at the effects of segregation.

[34] Grogger, J.T. (1996). “Does school quality explain the recent black/White wage trend?” *Journal of Labor Economics*, 14(2), 231-253. Rivkin, S.G. (2000). “School desegregation, academic attainment, and earnings.” *Journal of Human Resources*, 35(2), 333-346. The effects in the Rivkin study are only statistically significant in specifications that exclude lagged achievement as an independent variable. It is somewhat unclear why this would be the case.

[35] Most of the excluded states simply provide insufficient data. However, five states for which it is possible to calculate learning gains (Alaska, Hawaii, Idaho, North Dakota, and South Dakota) are dropped from the analysis because they have an insufficient number of minority students to carry out the analysis. Nevada and Oregon are also excluded due to questions about the characteristics and quality of the data.

[36] These national comparisons are based on Clotfelter (2004). In addition to the reasons mentioned in the text, the divergence between the Clotfelter data and the NCLB database is partly due to the fact that Clotfelter's results are five years old and the percentage of students in the United States who are Hispanic is increasing rapidly.

[37] Harris, D.N. (forthcoming b). "High flying schools, student disadvantage and the logic of NCLB," *American Journal of Education*.

[38] Lee, V.E. & Burkham, D.T. (2002). *Inequality at the Starting Gate*. Washington, DC: Economic Policy Institute.

[39] Another more technical way to see why the distinction between gains and levels is to recognize that each student's level of academic learning is determined by a wide range of factors, including their home environments, parent education, parenting practices, and neighborhood characteristics—all factors that are largely outside the control of teachers and school administrators. With learning gains, however, the role of these outside factors largely cancel out, so long as the effect of outside factors on a student's performance in, say, fourth grade is the same as it is when the student reaches fifth grade. Some of these—a divorce in the family or an illness—may change from year to year, but the outside factors generally remain the same. Even to the extent that they do change, they would only bias the results if certain groups experienced disproportionately large changes in a given year; for example, a sudden spike in the divorce rate for parents of African American children.

[40] Lee and Burkham (2002).

[41] While norm-referenced tests are not allowed under NCLB, some states administer them anyway or in addition to criterion-referenced tests.

[42] This calculation is based on a school that is 50 percent minority. See Appendix B for discussion of the specification of the peer achievement variable.

[43] Specifically, the simulations assume that each student is in elementary school for six years (kindergarten plus grades 1-5), in middle school for three years (grades 5-7), and in high school for four years (grades 9-12). The simulations also assume that the effects of segregation are the same within each grade level. As indicated in Table 1A, most of the elementary results are based on data from grades 3-6, which means that the effects are extrapolated to early grades.

[44] No data are available regarding FRL and peer achievement in Louisville and Seattle as discussed earlier. Therefore, I calculated the national averages for these variables in urban schools around the country that had similar racial compositions. The averages were calculated, first, at the actual exposure rates in Table 2 and, again, at the hypothetical exposure rates. The differences in poverty levels between the two comparison groups (by grade level) were 8.0, 5.4, and 6.7 for schools in the Louisville ranges of racial composition and 9.7, 5.4, and 5.5 for schools in the Seattle ranges. Thus, for example, the national average percent poverty for students in elementary schools with 34.7-38.7 percent minority (the actual Louisville average is 36.7) was 8.0 points higher than in schools with 49.7-53.7 percent minority (the hypothetical Louisville range). For peer achievement, the respective differences were 1.1, 0.3, -0.2 (standard deviations) for Louisville and 1.7, 0.2, -0.3 (standard deviations) in Seattle.

[45] The Type #1 and Type #2 analyses are based on the same regression coefficients. However, Type #1 includes all three coefficients (race, poverty, and peer achievement) because these are always jointly significant. The Type #2 analysis sets individual coefficients to zero if they are not statistically significant. See Appendix B for details.

[46] There are three other reasons to think the estimates are too high. First, there are still substantial achievement gaps between white and minority students within the same schools and classrooms. See, for example, the following excellent analysis of Shaker Heights: Ogbu, J. (2003). *Black American Students in*

an Affluent Suburb. London: Lawrence Erlbaum. Second, the simulation described in the text is, what economists would call, "partial equilibrium." This means that the exercise considers the immediate effects of a movement of students, but does not consider the possible indirect or "secondary" effects. For example, it is likely that to the degree that teacher quality is determined by any of the three factors—race, poverty, or peer achievement—the assignment of students across schools based on these factors may change the structure of the education system in important ways. There is considerable debate about the degree to which white students are likely to exit their existing schools as a result of desegregation—"White flight." In addition, teachers appear more willing to work in majority-white schools, which affects school quality. Desegregation may therefore result in a redistribution of teachers that reduces the benefits for minorities. Third, students in high-minority schools, because they come from more disadvantaged family backgrounds, are probably more susceptible to "summer learning loss," referring to the idea that students essentially forget some of what they learn in school during the summer break. The gains calculated in this study are based on changes in tests from one spring to the next and cannot directly account for summer losses.

[47] See Harris (forthcoming) and Harris and Herrington (2006).

[48] There are two possible effects of mobility. First, if mobility is independent of student achievement, then it simply creates measurement error in lagged student achievement. This would place a downward bias on the estimated effect of lagged achievement, but there is no a priori reason to think that that it biases the estimated effect of percent minority. Even though mobility is higher in high-minority schools, this should have an equal influence on both the current and lagged achievement levels and therefore cancel out. This canceling would also occur if mobility is non-random such that lower scoring students are more likely to be mobile. Again, this should have an equal influence on both the current and lagged test scores—regardless of whether these happen to be low-scoring students. In either case, there is not much reason to think that student mobility biases the estimated effects of percent minority. Mobility may also disrupt the classroom environment and result in less learning for all students, but this is part of the effect of segregation and needs to be included as part of the overall segregation effect.