Desegregation and the Achievement Gap: 
Do Diverse Peers Help?*

Jane Cooley†

January 12, 2006

Abstract

Understanding peer effects is critical to evaluating the impact of de facto public school segregation on the achievement of white and nonwhite students. Using a unique panel data set of North Carolina public elementary school students, I estimate a model of achievement production that incorporates heterogeneous responses by students at different points of the achievement distribution, while also allowing for peer spillovers to vary across races and for the formation of different race-based reference groups within the classroom. I find evidence of stronger peer influences within reference groups than across reference groups, the magnitude of which varies substantially across the percentiles of the achievement distribution. I apply my results to evaluate the efficiency and distributional effects of alternative classroom assignment policies. Diversifying peer groups leads to small but fairly uniform improvements in the achievement gap across the achievement distribution.

*I would like to thank Pat Bajari, Han Hong, Tom Nechyba, Peter Arcidiacono, Arie Beresteanu, Charlie Clotfelter, Stephen Ryan, Chris Timmins, Jake Vigdor and participants at Duke’s Applied Microeconomics, Applied Microeconomics lunch groups and Public Economics Workshop for their helpful comments. I am also grateful to the Spencer Foundation for financial support in the form of a dissertation fellowship. All remaining errors are my own.

†Department of Economics, Duke University, Durham, NC 27708; jcooley@econ.duke.edu.
1 Introduction

More than fifty years after Brown v. Board of Education brought to a close the era of de jure segregation in the South, public schools in the United States remain considerably segregated. As of 2000, about 70% of black students still attended predominately nonwhite schools.\(^1\) At the same time, the achievement gap between white and nonwhite students persists at a sizable magnitude. The 2000 National Assessment of Educational Progress reports that by the end of grade 4, black and Hispanic students are already two years behind their white peers.\(^2\) Given marked disparities in the schools white and nonwhite children attend, desegregation is largely advocated as a means of raising nonwhite achievement and closing the gap.\(^3\) The channels through which desegregation could narrow the gap might include the redistribution of resources (for instance, if it leads to nonwhites being exposed to better teachers) or the creation of “better” peer groups (if the achievement of lower-performing nonwhite students increases as a result of placement with higher-achieving white peers.) I isolate the effects of the latter mechanism, quantifying the equity and efficiency of diverse classroom peer groups using a unique panel data set of North Carolina public elementary school students.

A deeper understanding of peer effects is critical to assessing the impact of desegregating peer groups on the achievement of white and nonwhite students. The challenge lies in separating the effect of peer characteristics from the effect of peer actions and the overall effect of peers from unobserved group effects, i.e. Manski (1993)’s reflection problem. Previous literature on peer effects in education production simplifies the problem by minimizing the importance of spillovers from peer actions. Placing achievement production in the context of an equilibrium model clarifies the different sources of peer spillovers and motivates the use of an instrumental variables strategy to identify peer effects. Given anecdotal and empirical evidence that nonwhite students face very different peer pressures than whites,\(^4\) a critical aspect of the analysis is allowing for peer spillovers to vary across races and for the formation of different race-based reference groups within the classroom. Furthermore, given that recent

---

\(^1\)See Clotfelter (2004), Table 2.1.

\(^2\)See Section 2 for a more detailed discussion of evidence regarding desegregation and the achievement gap.

\(^3\)For example, in North Carolina, a state praised for its successful reform efforts, two of the ten “critical reforms” aimed explicitly at narrowing the gap involve a restructuring of school/classroom assignment policies. They strongly encourage districts (1) to eliminate ability grouping (tracking) in elementary schools, a policy that would effectively decrease within-school segregation, and (2) to desegregate schools. (Source: Report of the North Carolina Division of School Improvement.)

\(^4\)For instance, see Simmons (1999) and Fryer and Torelli (2005).
education policies, most notably the No Child Left Behind Act of 2001, aim specifically to raise the achievement of low performers, the heterogeneous responses of students at different ranges of the achievement distribution to any alternative classroom assignment policy are of paramount importance. Applying a quantile estimator to capture distributional effects enables me to assess both the equity and efficiency implications of desegregating peer groups.

Consistent with previous literature on educational achievement, this analysis centers around the educational achievement production function. Generally, the achievement production function takes as inputs student, teacher, and school characteristics, while peer effect studies also incorporate peer characteristics and/or peer achievement among the inputs. In so doing, these models effectively treat students as passive inputs to education production. I deviate from the literature by explicitly treating students as optimizing agents who exercise some control over their achievement through their choices of “effort.” Accounting for strategic choices of students has important implications for the sources of peer spillovers. First, there may be direct spillovers to achievement production from peer effort. For instance, if peers are more attentive in class, this is likely to facilitate learning. Second, a student’s preferences for achievement and effort may be influenced by peer behavior. For instance, there may be considerable psychic costs to deviating from socially expected norms of behavior. I assume that all students simultaneously choose effort, and observed achievement represents an equilibrium. Thus, in the context of peer effects, the achievement production function can be described more intuitively as a best-response function, where students choose actions that maximize utility for a given set of peer actions.

The fundamental identification problem lies in the fact that the econometrician only observes equilibrium outcomes. Identifying the best-response function then relies on finding an exogenous shifter that shifts an individual’s behavior without directly affecting his peers. North Carolina’s student accountability policies, which require that a student perform at a certain level in order to be promoted to the next grade, raise the costs of low achievement for the students for whom the policy is binding, or those “in danger of failing.” Thus, they provide an arguably exogenous shift in the behavior of a subset of peers that enables me to trace out the best-response function. Using this basic strategy, I obtain nonparametric

---

5See Hanushek et al. (2003) and Hanushek (1979) for a detailed discussion of the estimation of educational achievement production functions.

6These “conformity effects” are a popular form of endogenous spillover that are treated in the social interactions literature. (For instance, see Brock and Durlauf (2001b), Akerlof (1997), and Bernheim (1994).) Also, see Fordham and Ogbu (1986) for a description of a particular type of conformity effects that black students may face.
identification results for the effect of contemporaneous peer behavior on achievement, or endogenous effects in the familiar terminology of Manski (1993). Previous studies on peer effects in education have minimized the importance of contemporaneous spillovers from peer behavior and instead focus on spillovers from peer ability, as proxied by lagged peer achievement. However, the costs and benefits of alternative grouping strategies differ substantially if contemporaneous peer spillovers also exist.

Even after breaking the simultaneity in achievement through the use of an instrumental variable, two other well-documented sources of endogeneity may bias peer effect estimates. First, selection into peer groups that leads to similarity in outcomes can be confounded with a direct effect of peers. Though the instrumental variable strategy pursued in this paper alleviates selection concerns given evidence that students are not reassigned to peer groups as a result of student accountability policies, I incorporate teacher fixed effects to control for the most salient source of selection bias. The use of fixed effects to control for selection is also pursued in previous studies, while others benefit from natural experiments with students randomly assigned to peer groups.

Second, omitted variables, particularly unobserved ability, may be a confounding factor. In the present context, this is problematic if the students’ strategic behavior is a function of a characteristic that is unobserved by the econometrician. I use students’ choices regarding free reading time to proxy for the part of their reading ability that is known to the student but unobserved by the econometrician. Because no equivalent measure is available for math ability, this paper focuses on reading achievement. Previous studies have dealt with unobserved heterogeneity by estimating value-added specifications that difference out individual fixed effects. While this is a useful alternative, it is difficult to apply with a large data set and nonlinear framework, as investigated in this paper.

Finally, I use a control function approach to correct for the endogeneity of peer achievement spillovers to education production in a flexible way that permits variation across achievement quantiles and by race. This is particularly important for deriving efficiency implications of alternative assignment policies, since the widely used linear-in-means model

---

7 See Section 5.5 for a more detailed discussion of this literature.
8 For instance, see Graham (2004), Sacerdote (2001), and Vigdor and Nechyba (2004).
9 See Hanushek et al. (2003).
10 The importance of functional form assumptions for deriving implications about optimal classroom assignment in the presence of peer effects is also highlighted in a theoretical study by Arnott and Rowse (1987).
predicts that the gains from reallocating students to raise peer achievement in one classroom are perfectly offset by the losses from lower peer achievement in other classrooms. I make use of the techniques developed in the recent literature on endogenous quantile instrumental variable models, as in Abadie et al. (2002), Chernozhukov and Hansen (forthcoming), Chesher (2003), Honoré and Hu (2004), Imbens and Newey (2003), and Ma and Koenker (2004).11

A central finding of this paper is that whites appear to conform only to the behavior of their white peers, while there is evidence that nonwhites receive positive spillovers from both their white and nonwhite peers. Estimates of the marginal effect of white peers on nonwhites are of considerable magnitude, but very noisy and not generally statistically significant, which is consistent with various studies documenting the particularly complex nature of peer interactions for nonwhites.12 I find evidence of diminishing marginal returns to white peer achievement for whites and weakly increasing marginal returns to nonwhite peer achievement for nonwhites. Finally, endogenous effects appear to be much larger than exogenous effects. This is not surprising given that the implications of endogenous effects differ from exogenous effects due to a social multiplier, whereby increasing the achievement for a given student has positive spillovers for the achievement of his peers which in turn feeds back to his own achievement.

I apply the estimates of peer spillovers to assess the equity and efficiency implications of several alternative classroom assignment policies. First, desegregating North Carolina classrooms by randomly assigning students to any peer group in the state narrows the achievement gap by about 6% of a standard deviation, and this narrowing is fairly uniform across achievement percentiles. To place in context, the desegregated setting is then compared to the controversial alternative assignment policy of academic tracking. Second, I consider a more localized example of merging a predominantly white high-achieving school district (Chapel Hill) with a neighboring lower-achieving racially diverse district (Durham). In this starker example, where one might expect the impact of desegregation to be particularly large, I find that desegregation across the district lines does very little to narrow the achievement gap. Intuitively, this is because white and nonwhite Durham students make comparable gains in achievement on average as a result of the merger, while white and nonwhite Chapel Hill students experience comparable losses. These effects cancel out, leaving

---

11 For more details on this literature, see Section 5.5.
12 See Fordham and Ogbu (1986) and Fryer and Torelli (2005).
the average achievement gap virtually unchanged. This localized example helps to explain why the statewide desegregation experiment has such small effects on the gap, despite large peer spillovers. Finally, I quantify the achievement premium associated with “better” peer groups in a context where general equilibrium sorting effects are unlikely to be important, i.e., for the case of moving a student from Durham to Chapel Hill.

1.1 Literature Review

This paper contributes to the literature on the effects of desegregation on the performance of nonwhite students and to the literature on peer interactions. The desegregation literature frequently takes a historical perspective linking the end of de jure segregation to changes in black outcomes or seeks to link the racial composition of peer groups to educational outcomes. Evidence is mixed, with some concluding that desegregation notably improves black outcomes and others suggesting that desegregation has little to no effect.\(^\text{13}\) The focus of these studies diverges markedly from the present analysis, which, though allowing for direct racial composition effects, focuses on the contemporaneous behavioral spillovers of peers. Furthermore, I identify peer groups by race within classrooms, where much of the achievement-relevant social interaction may occur for elementary school students, and can therefore draw better inference about the different peer influences white and nonwhite students face.

Since within-school segregation occurs in large part in an environment where students are tracked (assigned to classrooms based on ability or prior achievement), evidence on the effects of tracking can shed some light on the potential benefits of integrating classrooms. To the extent that nonwhites are disproportionately placed in low tracks,\(^\text{14}\) any negative consequences of tracking on low-track students can exacerbate the achievement gap between white and nonwhite students. Whether the purported benefits of a specialized curriculum for low- and high-track students overwhelm the potentially negative consequences to low-track students’ expectations and self-esteem, and ultimately their achievement, remains a contentious issue.\(^\text{15}\) The quantile regression approach used in this paper provides new insight

\(^{13}\)For instance, see Card and Rothstein (2005), Cook and Evans (2000), Guryan (forthcoming), Hanushek et al. (2004) and Rivkin (2000).

\(^{14}\)This may occur either through racial discrimination or as a result of initial disparities in achievement (see Fryer and Torelli (2005) and Clotfelter (2004, pp. 137-139)).

\(^{15}\)In a synthesis of the vast literature on ability tracking, Slavin (1990) concludes that tracking is generally ineffective for raising the achievement of both high- and low-ability students. Figlio and Page (forthcoming)
on this debate.

The failure to account for contemporaneous peer effects may explain some of the mixed evidence regarding the effect of segregation and tracking in the literature, since, as Moffitt (2001) discusses in detail, ignoring endogenous effects when they exist will bias estimates of exogenous effects, or in the present context, peer racial composition or ability. Not only is accounting for endogenous effects important for establishing causation, but the policy implications in the presence of endogenous effects are very different than if there are solely exogenous effects due to a social multiplier. In the world of exogenous effects, narrowing the achievement gap involves potentially stark trade-offs. For instance, if all students benefit from having peers with high ability, creating classrooms of mixed ability benefits low-ability students at the expense of high-ability students. In the presence of endogenous achievement effects, the negative effect of detracking on high-ability students' achievement from their lower-ability peers is at least partially offset by the improved achievement of the low-ability peers.

The broader social interactions literature continues to struggle not only with the problem of separating endogenous from exogenous effects, but also the more fundamental problem of separating social effects—the aggregate of endogenous and exogenous effects—from unobserved group effects or correlated effects. Brock and Durlauf (2001a) focus on the identification of endogenous effects in a discrete choice setting. Graham (2004) offers an innovative approach to identifying the social effect in a linear-in-means context using excess variance contrasts between large and small classrooms. He uses a Bayesian informative prior to bound the magnitude of the endogenous effect. My approach differs substantially from these previous studies and identifies the different components of the social effect in a general setting.

---

16Previous results on identification in social interaction models are discussed in greater detail in Section 5.
17See Manski (1993).
18See Brock and Durlauf (2001b) for a detailed discussion of this literature.
2 Background

Though Brown v. Board of Education declared the de jure segregation of public schools unconstitutional in 1954, it was not until the late 1960s to early 1970s that significant progress was made in desegregating schools. One way to measure the extent of public school segregation is through a measure of racial imbalance, which captures the deviation of the observed exposure from the maximal exposure of whites to nonwhites in schools within a district. Formally, let the subscript \( d \) denote the districts and \( s \) the school. Then, \( E_{WN} = \frac{1}{N_s} \sum_s (%NW_s \cdot White_s)/White_d \) is the exposure rate of white to nonwhite students, where \( White_s \) and \( White_d \) denote the total number of white students in the school and district respectively and \( %NW_s \) denotes the percentage of students in school who are nonwhite. The maximum exposure in a district is equivalent to the percentage of nonwhite students in the district and is attained when schools are racially balanced. The segregation index then captures the deviation of the exposure rate from maximum exposure of whites to nonwhites, i.e.,

\[
S = (\%NW_d - E_{WN})/\%NW_d,
\]

where \( \%NW_d \) captures the percentage of students in the district who are nonwhite. The segregation index assigns a value between 0 and 1 to each district, with 0 representing a perfectly integrated district and 1 perfectly segregated.

Segregation of students across schools dropped significantly from 1970 to 2000. Clotfelter (2004) finds that for metropolitan areas, segregation fell from .458 to .326, while it fell more precipitously in the South, from .553 to .303. Clotfelter (2004) documents that the most rapid declines in segregation took place between 1968 and 1973. In the 1974 case of Milliken v. Bradley, the Supreme Court ruled that a federal court could not force integration across district lines, limiting the reach of desegregation efforts.

Furthermore, Figure 1 shows that segregation for grades 3 to 5 within North Carolina school districts has risen steadily since 1994-95 and, if anything, has accelerated in recent years. This can be attributed in large part to a recent succession of court rulings that have declared school districts unitary (i.e., sufficiently integrated) and have thereby released them from the obligation of proactively integrating. For instance, the Fourth Circuit Federal

---

19 See Clotfelter (2004) for a discussion of this and alternative measures of interracial contact.
20 This is only one of several measures of segregation and is influenced by the rising number of nonwhites.
Court of Appeals ruled that the Charlotte-Mecklenburg school district could no longer use race in making school assignments, which led to sharp increases in segregation. Whether the apparent resegregation of North Carolina public schools mirrors a nationwide trend remains a subject of debate.\textsuperscript{22}

While segregation between schools is certainly an important issue, many of the potential benefits of desegregation are muted if students remain segregated within schools. Clotfelter (2004) describes the persistent patterns of segregation within school-sponsored clubs, sports teams and at lunch tables. Though school officials have little to no control over interracial contact in these areas of school life, they can control the degree of classroom segregation. Unfortunately, the data linking students to classrooms are limited. Clotfelter et al. (2003) draw from the same data used in the present study to describe trends in within-school segregation in North Carolina public schools from the academic years 1994-95 to 2000-01. While they find relatively low within-school segregation in elementary schools as compared to middle and high schools, they find that in all cases the degree of within-school segregation increased from 1995 to 2001.

One dominant cause of within-school segregation is the use of ability grouping or tracking, which in elementary schools largely arises through the assignment of students to gifted programs. Using the Early Childhood Longitudinal Study of kindergartners in 1998-99, Fryer and Levitt (2004) find that black kindergartners perform .64 of a standard deviation lower than whites in math and .40 in reading, while Hispanics perform .72 of a standard deviation lower in math and .43 in reading. This finding in itself suggests that ability grouping in schools that is based on prior student achievement would produce greater within-school segregation. However, to further complicate the issue, various studies have found evidence of racial discrimination in track assignment, with black students less likely to be assigned to gifted classrooms.\textsuperscript{23} It is difficult to trace trends in the prevalence of tracking, since tracking occurs in so many different forms and varies by degree. Hallinan (2004) describes a general trend toward detracking that began in the 1980s but lost momentum somewhere along the way.

While the decision to desegregate schools did not hinge on the belief that it would improve

\textsuperscript{22}Using a sample of school districts in the South and Border states, Clotfelter et al. (2005) find that while North Carolina schools are resegregating by all measures, whether this is a general trend depends on the measure of segregation used.

\textsuperscript{23}See Clotfelter (2004, pp. 137-139) for a discussion.
academic outcomes, in the aftermath of the period of intense desegregation efforts (1968-1970), the black-white achievement gap declined precipitously. Figure 2 describes trends in the black-white achievement gap as measured by differences in National Assessment of Education Progress reading scores. Average scores for each subgroup range from 170 to 220. The figure presents the difference in average white achievement and average black achievement over the period. Between 1970 and 1988, there was a marked narrowing of the achievement gap for 9 and 13-year-olds; it dropped from a high of 44 points to a low of 18 points. After 1988, there was a sharp reversal in the trend, with the gap hovering between 30 and 35 points throughout the 1990s.

Studies attempting to link changes in black outcomes to desegregation have met with mixed success. Guryan (forthcoming) credits the desegregation of the 1970s with small declines in black dropout rates, while Cook and Evans (2000) conclude that desegregation does not explain the narrowing of the achievement gap in the 1970s. Hanushek et al. (2004) find that a higher proportion of blacks in a grade has a significantly negative effect on the growth in achievement for black students, while Rivkin (2000) argues that improving school quality is likely to be much more effective than desegregation in improving black achievement and wages. Card and Rothstein (2005) conclude that the effect of neighborhood segregation on the relative test scores of blacks is much larger than the effect of school segregation, and that much of this effect can be explained by differences in socioeconomic characteristics.

Alternative explanations for the achievement gap also have limited success in explaining either its persistence or its magnitude. Fryer and Levitt (2004) find that observable differences in the health and socioeconomic status of mothers can explain disparities in initial achievement at kindergarten. However, they have considerable difficulty explaining why the black-white achievement gap widens over the early years of schooling, rejecting, for instance, observable school differences and behavioral problems. Hedges and Nowell (1998) credit socioeconomic gains for a significant portion of the narrowing of the gap over the ’70s and ’80s. They find that, though both whites and blacks experienced similar trends in socioeconomic gains, blacks appeared to reap higher benefits from these gains. Grissmer et al. (1998) further show that while family characteristics explain almost all of white gains over the period from 1970 to the mid-1990s, they explain very little of black gains.

\footnote{24 For a description of the trends, see Hedges and Nowell (1998) and Grissmer et al. (1998).}
\footnote{25 See Jencks and Phillips (1998) for a more comprehensive look at these trends.}
3 Data

The data used in this analysis are administrative data for North Carolina public school students from the academic years 1994-95 to 2002-03. Beginning in Spring 1995, students in grades 3 to 8 were administered standardized End of Grade exams in reading and math. I focus on reading test scores. The range of test scores varies considerably across grades and years, as does the cutoff for Achievement Level III, the level designated as “passing” the exam. To make scores in a given grade comparable across years, I take the deviation of the raw scores from the cutoff value for passing and normalize them to have mean 0 and standard deviation 1 for each grade over the whole period. Suppose $y_{igt}$ denotes the raw test score for student $i$ in grade $g$ at time $t$, and $y_{gt}^{(3)}$ denotes the cutoff for Achievement Level III. The standardized score, $Y_{igt}$, is constructed as follows:

$$Y_{igt} = \frac{(y_{igt} - y_{gt}^{(3)}) - \frac{1}{T} \frac{1}{N} \sum_{i} \sum_{t} (y_{igt} - y_{gt}^{(3)})}{SD_g(y_{igt} - y_{gt}^{(3)})},$$

where $SD_g(y_{igt} - y_{gt}^{(3)})$ denotes the standard deviation for a given grade over all years.

I define nonwhite students to be black or Hispanic; all other students are white. As Figure 3 shows, reading achievement has steadily risen for both white and nonwhite students in North Carolina over the period of study, but the gap in the average achievement of these groups has remained fairly steady, comparable to the national trend.

A unique feature of these data is that each student record is linked to a teacher identification number. This permits the identification of classroom peer groups for grades where student instruction takes place primarily within self-contained classrooms. Beginning in middle school, it is frequently the case that a given teacher may teach the same subject to different groups of students (i.e., students have different teachers for each subject). I focus on elementary students in grades 3 through 5, where the teacher ID can reliably identify the

---

26 The data come from the North Carolina Education Research Data Center, http://www.pubpol.duke.edu/centers/child/uceddatacenter.html. The individual level student test data are confidential, but some of the aggregate data and enrollment data are publicly available at the North Carolina Public Schools web site, http://www.ncpublicschools.org/reportstats.html.

27 As will become clear in the model section, this is because I have a means of directly estimating reading ability because students choose how much time to spend reading for fun. The data do not contain an equivalent type of leisure choice that reflects math ability. To the extent that reading and math ability are highly correlated, I could use the same estimator to analyze peer effects in math achievement.
classroom peer group. Peer variables are then constructed at the classroom level, where the peer average for an individual student $i$ is for all the students in $i$'s classroom other than $i$.

Another useful feature of this data set is that it tracks students over time as long as they remain within the North Carolina public school system. Each student record is further linked to a grade within an identifiable school in an identifiable district. Included in the data are background characteristics, such as race, sex, parent’s education, and whether the student receives a free or reduced-price lunch, along with some information about student leisure time allocation, such as time spent reading for fun and hours spent watching television. However, since leisure time allocation is not collected prior to the 1998-99 school year, earlier years are dropped from the sample. Even after these restrictions, the sample remains very large with just under 1 million student-year observations.

Table 1 provides summary statistics on the achievement, background characteristics and peer groups for students in the restricted sample. The average standardized score is reported in the first row of the table. After removing the early years, the average standardized reading score for the sample is .14 due to an overall upward trend in test performance from 1995 to 2003. As shown in the second row, there is considerable heterogeneity in the average performance of peers across classrooms, with a standard deviation of .45. The average class size as identified from the teacher ID is about 23. The sample contains a sizable nonwhite population—about 30% black, and 3% Hispanic.

The manner in which data on parental education are collected varies across schools. In some cases, particularly in elementary school, the teacher provides a best guess of the parental education of their students. I assume that parental education is fixed over the period a student is in grades 3 to 8. To correct for potential measurement errors in reporting, I use data from grades 6 to 8 when available under the assumption that middle schoolers are better able to report parental education. Otherwise, I take the most frequently reported value to be the parental education for a given student. I divide parental education into three categories: (1) those who did not obtain a high school degree, (2) those with at least a high school degree, but not a four-year degree (this includes those who received two-year degrees or obtained some post-secondary vocational training) and (3) those with at least a four-year degree (this includes those with graduate and professional degrees). Of the students in the

---

*I drop the bottom and top percentile of class sizes.*
sample, 64% have parents with at least a high school degree but less than a four-year degree, while 30% have parents with at least a four-year degree.

Students who come from households with sufficiently low income can qualify to receive either free or reduced-price lunches when they are attending school. About 40% of the students receive a free/reduced-price lunch. Students report the hours of television watched in a given day and the time spent reading for fun, which can serve as rough indicators of leisure choices. These variables were taken from multiple choice questions. I take the midpoint of the range of hours indicated as an approximation of the hours spent doing each of these activities. In general, students spend significantly more time in a given week watching television (2.5 hours) than leisure reading (.91 hours).

Table 2 reveals that there are notable disparities in the background characteristics and achievement of white and nonwhite students in the sample. On average, whites have higher achievement, .36 compared to -.30. They also have better-educated parents and are less likely to receive free/reduced-price lunches than nonwhites. Furthermore, whites allocate more time to reading for fun and less time to watching television than nonwhites.

While disparities in background characteristics may explain some of the gap in achievement between whites and nonwhites, as discussed in Section 2, another important factor is their classroom peers. As an indication of the extent of classroom segregation, only 24% of the peers of whites are nonwhite, compared to 52% for nonwhites. By all traditional measures, whites are in much “better” peer groups than nonwhites. On average, the classroom peers of whites have better-educated parents, are less likely to receive a free/reduced-price lunch and have higher achievement.

4 Model

Section 4.1 describes the education technology and the preferences of students. Section 4.2 defines and describes the equilibrium to the game and the resulting equilibrium achievement production function, i.e., the achievement realized under utility-maximizing effort.

This is a very noisy measure of income and highly correlated with parental education and race. For these reasons, I do not use it in the regression analysis.
4.1 Primitives

The model involves a simultaneous move game of symmetric information that is played within a peer group, which is defined to be a classroom of students in a particular time period.\footnote{Because the focus is on a particular peer group, I suppress time and classroom subscripts for the moment.} Let \( i = 1, \ldots, N \) index students in a given peer group. Achievement \( Y_i \in \mathbb{R} \) is the standardized reading test score.\footnote{While students are required to take both standardized math and reading exams, the focus is on reading achievement. As discussed further below in Section 5.3, the observation of free reading time offers a means of controlling for heterogeneity in reading ability that is unobserved to the econometrician. Since no equivalent measure exists in the data to proxy for math ability, reading test scores are a natural focus.} The achievement production function for a student \( i \) is as follows:

\[
Y_i = g(e_i, e_{-i}; S_i, \theta_i).
\]  

(4.1)

The choice variable of a student \( i \) is effort, which is chosen on the compact set \( e_i \in [\underline{e}, \bar{e}] \). It encompasses a variety of classroom behavior choices, such as how hard to work on classroom assignments, cooperativeness and attention during lectures. The achievement of \( i \) is determined both by \( i \)'s effort and the effort of his peers, \( e_{-i} = (e_1, \ldots, e_{i-1}, e_{i+1}, \ldots e_N) \). Furthermore, \( i \)'s achievement depends on state variables \( S_i \), which may include individual and peer characteristics as well as classroom inputs such as teacher quality.

This production function admits two types of direct peer spillovers. First, as mentioned above, peers may directly affect an individual’s achievement through their innate characteristics (or exogenous effects), which enter through \( S_i \). Exogenous effects are the focus of the literature on peer effects in education, which has found evidence of spillovers to peers’ race, sex, socioeconomic status and ability. Second, peers may affect achievement through their actions or effort (or endogenous effects). For instance, any one student’s choice to disrupt class takes productive learning time away from all students in the classroom, resulting in lower achievement for all.\footnote{Lazear (2001) presents such a model where the classroom learning environment is treated as a congestible public good.}

Finally, a student cannot perfectly predict his achievement on an exam, even after choosing his own effort and observing the effort of his peers. For instance, the student may not fully know his own ability. This is particularly likely for elementary schools students, given their limited experience taking standardized exams. Furthermore, ability is relative, and while a fourth grader may have some knowledge of his standing in the classroom, it would be difficult for him to know how his ability compares to that of students in other schools.
Also, unpredictable random factors, such as how well he slept the night before, may affect a student’s performance on a given test day. These types of unobservables are captured by $\theta_i$, which can be thought of as an ex post random shock to achievement. I allow for correlation in these random shocks. For instance, construction outside the classroom on test day may provide a distraction that negatively affects the performance of all students.

A student’s utility is defined as follows:

$$U_i = u(Y_i, c_i(e_i, e_{-i}); S_i). \quad (4.2)$$

Students derive utility from achievement and disutility from effort. Exerting effort is costly; the costs are captured by the term $c_i(e_i, e_{-i})$ with $\partial c_i(\cdot)/\partial e_i \geq 0$. Utility is decreasing in $c_i(\cdot)$. Furthermore, preferences for achievement and effort are also determined by state variables $S_i$. A student with highly educated parents may face higher expectations regarding academic performance and thereby derive greater utility from high achievement relative to an otherwise similar student with less educated parents. Education policymakers or teachers may also play a role in determining preferences for achievement, through policies such as imposing achievement standards for promotion to the next grade level or rewarding high performance.\(^{33}\)

Comparable to the case of production, this utility function is sufficiently flexible to allow for the presence of both exogenous and endogenous peer effects. Peer characteristics may enter through the state variable $S_i$. Furthermore, the costs of effort include a “social component,” which captures an alternative source of the endogenous peer effect. Intuitively, peer pressure imposes psychic costs to deviations from the behavioral norm, leading students to seek to conform to the behavior of peers. For instance, a student may prefer not to exert much effort in order to avoid earning the derogatory title of “nerd.” However, being a nerd may entail no psychic costs in a classroom full of nerds.\(^{34}\) This type of peer spillover has received a great deal of attention in the social interactions literature.\(^{35}\)\(^{36}\)

\(^{33}\)Equivalently, one could think of $S_i$ as affecting the cost of effort. “Good” teachers make achievement fun in the sense that effort is less costly.
\(^{34}\)See Bishop et al. (2003) for a discussion of these types of peer spillovers, particularly in the high school setting.
\(^{35}\)See Brock and Durlauf (2001b), Graham (2004), and Sweeting (2004).
\(^{36}\)An alternative model may have the utility from achievement depend on the achievement of peers, i.e, students care more about whether they perform better than others rather than how hard they work relative to others. Ultimately the implications are similar, since this model would suggest that a given student is induced to exert more effort when his peers are exerting more effort in order to maintain his rank in the
The vector of state variables $S = (S_1, ..., S_N)$ is common knowledge to all students in the classroom, while $(\theta_i, \theta_{-i})$ are observed ex post.\footnote{Alternatively, $(\theta_i, \theta_{-i})$ may in fact not be observed, but only imputed after a level of achievement is realized.} Students possess a common prior on $\theta$, $f(\theta|S)$.\footnote{An alternative model is one of private information. It is possible to show that the model of private information has similar implications to the symmetric information model described above.} Suppose $\theta_i$ is defined on the set $\Theta$. Then the expected utility for a given level of effort, $(e_i, e_{-i})$, is denoted as follows:

$$\tilde{U}_i(e_i, e_{-i}; S) \equiv \int_{\Theta} U_i(e_i, e_{-i}; S_i, \theta_i) f(\theta_i|S) d\theta_i.$$ 

4.2 Equilibrium

A student chooses effort to maximize his expected utility conditional on his information set. Let the superscript “∗” denote a utility-maximizing action. The best response $e_i^*(e_{-i}; S)$ of a student $i$ to a given vector of peer effort is then:

$$e_i^*(e_{-i}; S) \in \arg\max_{e_i} \tilde{U}_i(e_i, e_{-i}; S).$$ (4.3)

A pure strategy Nash equilibrium to the game involves everyone playing their best responses.

**Definition 4.1 (Equilibrium).** The vector $e^* \equiv (e_1^*, ..., e_N^*)$ is a pure strategy Nash equilibrium if and only if for every $i$ in a given peer group $e_i^*$ solves (4.3).

To show that an equilibrium exists requires imposing additional assumptions on the structure of the model. I begin with an ordinal concept of complementarities introduced by Milgrom and Shannon (1994), which they show to be a necessary condition to obtain best responses that are non-decreasing. Let $T$ be a partially ordered set and $h : A \times T \mapsto \mathbb{R}$, where $A \equiv \mathbb{R}$. Throughout, I assume that $\leq$ partial orders vectors according to the component-wise order, which is defined as follows:

**Definition 4.2 (Component-wise order).** For the vector $t = (t_1, ..., t_n) \in \mathbb{R}^n$ where $t_i \in \mathbb{R}$ for $i = 1...n$, $t' \leq t''$ in $\mathbb{R}^n$ if $t'_i \leq t''_i$ in $\mathbb{R}^1$ for $i = 1...n$. 

classroom.
**Definition 4.3** (Single Crossing Property). A function $h$ satisfies the single crossing property in $(a; t)$ if for $a' > a''$ and $t' > t''$, $h(a', t') \geq h(a'', t'')$ implies $h(a', t') \geq h(a'', t')$.

**Assumption 4.1** (Single Crossing of Ex Ante Utility). Let $S_i$ be defined on the set $S \subset \mathbb{R}^d$. $\tilde{U}(e_i, e_{-i}; S_i)$ satisfies single crossing in (i.) $(e_i; e_{-i})$ on $[\underline{e}, \overline{e}]$, and (ii.) $(e_i, S_i)$ on $[\underline{e}, \overline{e}] \times S$ for every $i$.\(^{39}\)

Holding the state variables fixed, single crossing in $(e_i; e_{-i})$ implies that if an individual prefers to exert a higher level of effort when peers are exerting $e''_{-i} < e'_{-i}$, then he will still prefer to exert that higher level of effort when peers exert more effort, $e'_{-i}$. A sufficient condition such that this holds is $\partial^2 \tilde{U}_i / \partial e_i \partial e_j \geq 0$ for all $j \neq i$. Holding peer effort fixed, consider the implications of imposing single crossing in the state variables. Single crossing in $(e_i; S_i)$ simply assumes that if the higher level of effort is preferred for a given level of state variables, $S_i = S''_i$, then the higher effort is also preferred for higher values of the state variables, $S_i = S''_i > S'_{i}$.

For example, suppose we compare two students, $A$ and $B$, who are similar in state variables except that $A$ has better-educated parents. If $B$ prefers $e'$ to $e''$ where $e' > e''$, then $A$ will as well. This would follow if the marginal utility of achievement were higher (or if the marginal cost of effort were lower) for the student with better-educated parents.

Because best responses are monotonically increasing under Assumption 4.1, it follows from Milgrom and Shannon (1994, Theorem 4) that an equilibrium to the game exists, which is restated as:

**Theorem 4.1** (Existence of Pure-Strategy Nash Equilibrium). Given that Assumption 4.1 holds, a pure strategy Nash equilibrium, as defined in 4.1, exists in non-decreasing strategies.

If effort were observable to the econometrician, the natural object of interest would be the best response to peer effort. Unfortunately, effort is not observable. However, placing the following monotonicity assumption on the achievement production function ensures that the game in effort maps into a game in achievement that is observable to the econometrician.

**Assumption 4.2** (Monotonicity in effort). Achievement production is strictly monotonic in effort, i.e.,

$$\frac{\partial g(\cdot)}{\partial e_i} > 0.$$  

\(^{39}\)Milgrom and Shannon (1994) show that single crossing may not be preserved under addition. A sufficient condition for this assumption to hold is that the conditional distribution of $\theta_i$ is log-supermodular (i.e., types are affiliated) and that ex post utility satisfies increasing differences.
Definition 4.4 (Achievement Equilibrium). Given Assumption 4.2, the vector \((Y_1^*, ..., Y_N^*)\) is a pure strategy Nash equilibrium if and only if for every \(i\) in a given peer group, \(Y_i^*\) is the achievement realized by exerting the effort \(e_i^*\) that solves (4.3).

Theorem 4.2 (Existence of Achievement Equilibrium). Under Assumption 4.2, the game in effort maps into a game in achievement, which can be represented as

\[
Y_i^* = q(\bar{Y}_i^*, S_i, S_{-i}, \theta_i),
\]

where \(\bar{Y}_i = \int g(e_i, e_{-i}; S_i, \theta_i)f(\theta_i|S)d\theta_i\).

See Appendix A.1 for proof. Equation 4.4 is similar in form to the production functions with peer effects estimated in the literature. Observed achievement is a function of peer achievement, an individual’s own characteristics and classroom inputs \((S_i)\), peer characteristics \((S_{-i})\) and unobservables \((\theta_i)\).

5 Identification

For the purposes of identification, it is useful to distinguish between the different components of the state space. Let \((S_i, S_{-i}) \equiv (X_i^0, X_{-i}^0, K, \mu)\), where \(X_i^0\) captures characteristics of \(i\) such as race, sex, parental education, and ability, while \(X_{-i}^0\) captures the characteristics of \(i\)'s peers. Besides the composition of the peer group, classrooms are differentiated by characteristics \((K, \mu)\), which may capture classroom resources, teacher quality, or overall classroom productivity. I assume that while \(K\) and \(\mu\) are observed by the students and therefore taken into account when choosing effort, \(\mu\) is unobserved to the econometrician. Intuitively, \(\mu\) can be thought of as the part of classroom productivity not captured by teacher fixed effects (or teacher quality), which are included in \(K\). Thus, \(\mu\) could describe the degree to which students get along with the teacher or something more random like a flu epidemic which negatively affects the achievement of all students in the classroom. To use Manski (1993)'s terminology, \(X_{-i}^0\) captures exogenous effects and \(\mu\) correlated effects.

Let \(\bar{Y}_i^* \equiv \frac{1}{N-1} \sum_{j \neq i} \bar{Y}_j^*\) capture the average expected peer achievement and similarly, \(\bar{X}_{-i}^0\), average peer characteristics. The achievement best-response function is simplified to depend on the mean of peer achievement and peer characteristics, rather than the entire
vector as permitted in (4.4), i.e.,

\[ Y_i^* = q(Y_{-i}^*, X_i^0, \bar{X}_{-i}^0, K, \mu, \theta_i). \quad (5.1) \]

An important aspect of this production function is that it is permitted to be nonseparable in the error. The linear model, which is primarily used in the literature, is a special case of this model where \( \theta_i \) enters additively. In the present context, allowing for nonseparability in the residual permits, for instance, the effect of peer achievement to vary across high-achieving and low-achieving students. Those who advocate the elimination of tracking and desegregation of peer groups frequently argue that low-achieving students respond more to increases in peer achievement than high-achieving students, which suggests that racially-mixed/mixed-ability peer groups are both more equitable and more efficient relative to tracking. In contrast, the widely used linear-in-means model predicts that the gains from reallocating students to raise peer achievement in one classroom are perfectly offset by the losses from lower peer achievement in other classrooms, with no efficiency implications.

I assume that \( q(\cdot) \) is strictly increasing in \( \theta_i \), a property that is satisfied by models that are additively separable in the residual. Since the structural function \( q(\cdot) \) is only identified up to positive monotone transformations when the error is nonseparable, I follow the literature on quantile treatment effects in assuming that \( \theta_i \) is independently and identically distributed \( U(0, 1) \). Since \( \theta_i \) is inherently without units, assuming a uniform distribution simply acts as a normalization that pins down \( \theta \). In contrast, the additive model normalizes \( \theta_i \) to have the same units as \( Y_i \). By fixing \( \theta_i = \tau \), Equation 5.1 describes the dependence of the \( \tau \)th quantiles of the achievement distribution on average expected peer achievement and covariates. Identification \( q(\cdot) \) is defined as follows:

**Definition 5.1 (Identification).** The structural function \( q(\cdot) \) is identified on the joint support of \( (Y_i^*, \bar{Y}_{-i}^*, X_i^0, \bar{X}_{-i}^0, K) \), if there exists a unique \( q(\cdot) \) that rationalizes \( F(Y_i^*, \bar{Y}_{-i}^* | X_i, \bar{X}_{-i}, K) \), the observed joint distribution of achievement and peer achievement conditional on exogenous state variables.

In what follows, I describe conditions such that the structural function is identified. There are basically three identification problems described in the social interactions literature: (1) the simultaneity of actions (Manski’s (1993) reflection problem), (2) unobserved

---

\(^{40}\)This simplification is not necessary for identification. The argument follows through with some modification when instead the peer effect is coming through a vector of moments of peer achievement.
ability and (3) non-random assignment to peer groups. Section 5.1 focuses on the reflection problem, under the assumptions that assignment to classrooms is random and that all strategy-relevant student characteristics $X_i^0$ that are observed by the student are also observed by the econometrician. Critical to the identification argument is the presence of a valid exclusion restriction. The particular exclusion restriction used in this paper is discussed in detail in Section 5.2. In Sections 5.3 and 5.4, I consider identification when there exists some strategy-relevant characteristic that is observed by the student but not by the econometrician (i.e., unobserved ability) and when there is non-random assignment to peer groups.

5.1 Simultaneity in Achievement

The central problem in identifying the structural function (5.1) is that the input to production $Y_i^*$ is also a function of the state variable $\mu$, which is unobserved by the econometrician, a classic simultaneity problem. Manski (1993) separates this into two identification problems: (1) separating endogenous effects (the effect of peer achievement) from exogenous effects (the effect of peer characteristics) and (2) separating social effects (endogenous and exogenous effects) from correlated effects ($\mu$). The existence of a valid exclusion restriction, which shifts the optimal behavior of peers independently of $\mu$, offers a solution to both of these identification problems. As I elaborate below, a valid exclusion restriction should enter a student’s utility function but not achievement production directly, i.e., it should affect equilibrium achievement only through the student’s choice of effort. The following assumptions describe the properties of a valid exclusion restriction in the context of the equilibrium model of classroom achievement production described in Section 4.

Assumption 5.1 (Utility Shifter). There exists a variable $P_i$ that affects $i$’s utility from effort, but does not directly affect achievement production. Formally, let $S_i = (S_i', P_i)$. Then, $P_i$ enters (4.2), i.e., $U_i = u(Y_i, c_i(e_i, e_{-i}); S_i', P_i)$, \footnote{An implicit assumption is that $P_{-i}$ is also not an element of $S_i'$.} but not (4.1), i.e., $Y_i = g(e_i, e_{-i}; S_i', \theta_i)$.

Assumption 5.2 (Independence). (i.) $\theta_i$ is independent of $P_i$, and (ii.) $\mu, \theta_i$ are jointly independent of $P_{-i}$.

Under Assumptions 5.1 and 5.2, utility-maximizing effort is a function of $P_i$ and not $P_{-i}$ (i.e., $e_i^* = e_i^*(e_{-i}; S_i', S_{-i}', P_i)$). Therefore, $P_i$ affects $i$’s equilibrium achievement only through...
his optimal choice of effort. The independence of \( \theta_i \) and \( P_{-i} \) ensures that \( P_{-i} \) does not enter expected utility through the distribution of \( \theta_i \), i.e., \( f(\theta_i|S_i', S_{-i}', P_i, P_{-i}) = f(\theta_i|S_i', S_{-i}') \). Otherwise, \( P_{-i} \) would enter the student’s utility-maximizing effort.

Let \((S_i', S_{-i}') = (X_i, X_{-i}, K, \mu)\). Then, Equation 5.1 is modified as follows:

\[
Y_i^* = q(\bar{Y}_{-i}^*, X_i, \bar{X}_{-i}, P_i, K, \mu, \theta_i). \tag{5.2}
\]

The system of response functions for students \( i = 1, ..., N \) in a given peer group is then as follows:

\[
Y_1^* = q(\bar{Y}_{-1}^*, X_1, \bar{X}_{-1}, P_1, K, \mu, \theta_1),
\]
\[
Y_2^* = q(\bar{Y}_{-2}^*, X_2, \bar{X}_{-2}, P_2, K, \mu, \theta_2),
\]
\[
\vdots
\]
\[
Y_N^* = q(\bar{Y}_{-N}^*, X_N, \bar{X}_{-N}, P_N, K, \mu, \theta_N).
\]

I assume that there exists some function \( h(\cdot) \) that approximates the average expected value of peer achievement, such that

\[
\bar{Y}_{-i}^* = h(X_i, \bar{X}_{-i}, P_i, \bar{P}_{-i}, K, \mu). \tag{5.3}
\]

Intuitively, expected peer achievement is a function of the state variables that are common knowledge to all students in the peer group, including \( \mu \), which is unobservable to the econometrician. If \( q(\cdot) \) were linear-in-means, then I could solve explicitly for \( \bar{Y}_{-i}^* \) as a function of individual characteristics, average peer characteristics, and the shared components \((K, \mu)\). With \( q(\cdot) \) nonlinear, this assumption, while more restrictive, still offers a fairly flexible approximation of average expected peer achievement.

Equations 5.2 and 5.3 form a triangular system of equations. If the structural function is restricted to be linear-in-means, these equations are comparable to the second and first stages, respectively, of a two-stage least squares regression. For the peer effect to be identified, there needs to exist a valid exclusion restriction that enters Equation 5.3 but not Equation 5.2, and is plausibly independent of the unobserved components. These criteria are described in Assumptions 5.1 and 5.2. The reason that the exclusion restriction cannot enter achievement production directly (Assumption 5.1) is because of the direct spillovers.
from effort in achievement production. Intuitively, if \( P_j \) had a direct effect on achievement production for student \( j \), it would affect the achievement of his classmate \( i \neq j \) because expected peer achievement net of state variables serves as a proxy for direct spillovers from unobserved peer effort in achievement production. In other words, \( \bar{P}_{-i} \) would also enter Equation 5.2 and would no longer be a valid instrument.

Two types of utility shifters, as described in Assumption 5.1, seem viable. First, since parents are likely to have considerable influence in determining how their child values achievement, parental characteristics are a natural starting point for thinking about exclusion restrictions. However, the argument that certain parental characteristics affect how a student values achievement but not achievement directly is a difficult one to make. For instance, students with better-educated parents are likely to value education more and to have better help on homework, suggesting that parental education is also a direct input to achievement.

Second, education policymakers may affect the utility a student derives from achievement by either raising the rewards for high achievement or increasing the costs of low achievement. To be a valid exclusion restriction, the policy must not affect teacher incentives (because then \( P_i \) would affect \( i \)'s achievement directly) and must also have different effects on students within the same classroom. Policy variation that plausibly meets these criteria exists in North Carolina in its student accountability policy, which was enacted during the period of study. It imposed stricter, or at least more transparent, standards for grade retention and promotion, namely that a student perform above the level deemed as sufficient mastery of material (Achievement Level III) on the standardized End of Grade (EOG) tests in order to be automatically promoted to the next grade. The standards are absolute, in the sense that they are not set such that a certain percentage of students fails in a given year. Student accountability began to take effect for all fifth graders in the North Carolina public school system in 2001.\(^{42}\)

Identification relies on the intuition that student accountability policies had a direct effect on the effort of students “in danger of being retained,” but no direct effect on the effort of those well above the threshold for passing. Students who performed below Achievement Level III in the year before the standards were put in place are induced to exert more effort to meet the requirement, due to the increased cost of low achievement. On the other hand,

\(^{42}\)These policies had a significant effect on retention rates. I find that the retention rate for fifth graders increased by 50\% after student accountability policies were enacted (from .010 to .015). Over the same period, retention of fourth graders only increased by 7\% (from .015 to .016).
high achievers could effectively disregard these policies because the new standards are not binding for them, i.e., they would probably meet the cut-off for passing even with minimal effort.

In the present context, Assumption 5.2 requires that the percentage of students in danger of failing under the new standards for promotion is independent of $\mu$ and $\theta_i$. Independence of $\theta_i$ and $\bar{P}_{-i}$ is fairly intuitive given that $\theta_i$ is unobserved at the time that students choose their effort. However, independence of $\mu$ and $\bar{P}_{-i}$ merits more focused attention. A particular concern is that teachers responded in some way to these new standards (for instance, by shifting resources to low achievers) and that the unobserved classroom productivity $\mu$ and the instrument are therefore correlated. This will be examined in more detail in Section 5.2. The requirement of full independence is stronger than the typically assumed mean independence, but is a necessary trade-off for identifying the production function under weaker functional form assumptions.

For the structural function to be nonparametrically identified, I need to place one further restriction on the structure of the reduced-form equation for average expected peer achievement (5.3), namely that it is strictly monotonic in the unobserved group effect. Note that an important special case where this property is satisfied is in models that assume additive separability in the unobserved components.

**Assumption 5.3 (Monotonicity of Peer Achievement).** With probability one, $h(X_i, \bar{X}_{-i}, P_i, \bar{P}_{-i}, K, \mu)$ is strictly monotonic in $\mu$.

To fix a value for $\mu$, I assume that it is distributed $\mathcal{U}(0, 1)$. Under the above assumptions, $\mu$ can be recovered from the first-stage regression as shown in Imbens and Newey (2003, Theorem 1) and stated formally in the following theorem.

**Theorem 5.1 (Identification of $\mu$).** Given Assumptions 5.2, 5.3 and $\mu \sim \mathcal{U}(0, 1)$,

$$F_{\bar{Y}_{-i}^*|X_i, \bar{X}_{-i}, P_i, \bar{P}_{-i}}(\bar{Y}_{-i}^*|X_i, \bar{X}_{-i}, P_i, \bar{P}_{-i}) = \mu.$$  

See Appendix A.1 for details.

Given that $\mu$ can be recovered from (5.3) by Theorem 5.1, it remains to be shown that the structural function, $q(\cdot)$, is identified. This requires imposing the additional assumption, that the unobserved group effect is independent of the individual type.
**Assumption 5.4** (Independence of Residuals). \( \theta_i \) is independent of \( \mu \).

Recall that the individual shocks \( \theta_i \) can be correlated within the classroom. This assumption just requires that the state variable that is unobserved to the econometrician is independent of the individual shock, which is fairly straightforward given that \( \theta_i \) is realized ex post.

Under Assumption 5.4, for values of \( \theta_i = \tau \), the structural function \( q(\cdot; \tau) \) can be interpreted as a conditional quantile function that describes the dependence of the \( \tau^{th} \) quantile of achievement on peer achievement conditional on observed state variables \( (X_i, \bar{X}_i, K) \) and the common component \( \mu \).

**Theorem 5.2** (Identification of the Structural Function). Given Assumptions 5.2, 5.3, and 5.4, \( q(Y^*_i, X_i, \bar{X}_i, P_i, K, \mu, \theta_i) \) is identified on the joint support of \( (Y^*_i, X_i, \bar{X}_i, P_i, \mu, \theta_i) \).

See Appendix A.1 for details. The proof of this result follows from Imbens and Newey (2003, Corollary 6), similarly to the proof of Theorem 5.1 above. Intuitively, conditioning on the unobserved group effect \( \mu \) controls for the endogeneity of peer achievement, and the structural function is identified.

### 5.2 Exclusion Restriction

In this section, I provide a little more background on North Carolina’s student accountability policies and carefully discuss the implications of Assumption 5.2. One important aspect of the identification argument is the assumption that student accountability only affects the incentives of a subset of the students in the classroom, namely those “in danger of failing.” The actual effect of student accountability on the distribution of achievement for fifth graders in the largest North Carolina school district is shown in Figure 4. Comparing the year prior to accountability (2000) to the first year of accountability (2001), we see that the lower tail of the distribution shifted toward the center while the upper tail remained about the same, suggesting that low achievers responded to the threat of retention. In contrast, the distribution of achievement for fourth graders, who were not held to the new accountability standards in either year, remains almost identical across the two years. This supports the argument that schools did not respond generally to student accountability by shifting resources to low achievers in all grades.
Note that if accountability itself were the instrument, one could plausibly identify the peer effect for high achievers (those not in danger of failing), but it would not be possible to separate the direct effect of accountability on the achievement of those in danger of failing from the endogenous peer effect. Given that much of the current policy debate is centered around improving the performance of low achievers, this would inhibit addressing many policy-relevant questions. However, the model suggests using an aggregate measure of the policy effects on peers ($\bar{P}_{-i}$) in the classroom as the instrument. Intuitively, the magnitude of the peer effect will vary by the percentage of peers in danger of failing, since larger percentages imply a larger shift in peer achievement due to the policy. This is illustrated in Figure 5, which compares the distribution of achievement in two districts with different concentrations of low achievers (based on previous scores). Though the size of the shift is different at all percentiles of the achievement distribution, there was a large shift in both the upper and lower tail of the distribution for the district with a high percentage of low achievers, while the shift in the upper tail of the distribution was smaller for the district with a lower percentage of low achievers. This evidence is consistent with the story of positive spillovers from improving the effort of low performers. This strategy permits the identification of the peer effect for all students, assuming that the percentage of students affected by the accountability policy is independent of $\theta_i$ and $\mu$.

Independence of $\bar{P}_{-i}$ and $\mu$ (Assumption 5.2) may not hold if teachers or administrators redistribute resources to low achievers as a result of student accountability. For instance, a teacher may choose to teach more to the low achievers as a result of the policy, suggesting a direct effect of student accountability on achievement production. If this occurs, then classes with higher percentages of low achievers may experience shifts in the achievement distribution simply as a result of teachers devoting more time to low achievers and less time to high achievers. This is only problematic if the shift occurs for the grades in which student accountability policies are in place and not for other grades. Otherwise, fourth grade acts as a control group. More importantly, any shifts in resources to low achievers are likely to have occurred prior to the advent of student accountability with the introduction of school accountability in 1996, under the School Based Management and Accountability Program.

---

43 For instance, see the federal government’s No Child Left Behind Act of 2001 and the ensuing debates.
44 Of course, the problem is solved if one is willing to assume that the marginal effect of peer achievement is constant across percentiles of the achievement distribution. This severely limits the ability to derive efficiency implications from alternative assignment policies, and further, as this paper illustrates, is not a very plausible assumption.
45 In the high percentage district, 25% of students were in danger of failing, as compared to only 14% in the low percentage district.
This policy included bonuses for schools and teachers when the target percentage of students performed above Achievement Level III on the standardized EOG exams, so that teachers had strong incentives to meet these standards. Given that student accountability policies did not take effect until Spring 2001, it is likely that any resource shifts or changes in teaching style were fairly well established prior to student accountability. That said, to help ensure that the change in achievement of the peer group is not due to shifts the unobserved state variable $\mu$, I control for a direct effect of accountability on high- and low-performers by including a quantile-specific location shift for achievement of students after accountability.$^{46}$

A second related concern is that students may be reassigned to different classrooms as a result of accountability. This might occur if parents become more concerned about the threat of retention, and thus try to have their children placed in better classrooms, or if school administrators determine that redistributing low achievers across classrooms can help lower the probability of retention. Given that any reassigning is likely to occur through the choice of teachers, controlling for teacher fixed effects alleviates this concern.$^{47}$

5.3 Ability

The identification of peer spillovers is further complicated if the student’s strategic effort choice is based on a characteristic that is unobserved to the econometrician but observed to the student and his peers. I describe this characteristic as “observed ability” ($A_i$) or the persistent component of ability, since it is observed by the student, to differentiate it from the unobserved component of ability that is included in $\theta_i$. To clarify the distinction

$^{46}$As will be noted in Section 7, I find that accountability has only a small direct effect on achievement in the linear-in-means model, but has the predicted positive and significant effect on those in danger of failing, consistent with the intuition that accountability acts primarily as a preference shifter rather than a direct input to production.

$^{47}$If low achievers are being reassigned to classrooms in response to student accountability, then we should observe a shift in the dispersion of prior-year test scores in peer groups. If parents or schools thought they could improve a low-performing student’s achievement by placement with higher-achieving peers, then classrooms would become more unequal, in the sense that there would be greater variation in prior test scores within the class. As a measure of inequality, I calculate the Gini coefficient for fifth grade classrooms in 2000 and 2001. Then, I conduct a Kolmogorov-Smirnov test for the equality of distributions of the Gini coefficients before and after accountability. Recovering an approximate p-value of .19, I fail to reject that the distributions of the Gini coefficients are the same in the two years. It may be that shifts occurred at the school level, which I control for with a school fixed effect. After accounting for changes in inequality at the school level by taking the ratio of the Gini coefficient at the classroom level to that at the school level, I also fail to reject that the distributions of the ratios are the same across the two years with an approximate p-value of .62. Thus, I do not find evidence of significant regrouping as a result of student accountability.
between these two components, observed ability may be learned by a student over time through consistent high performance in school or the relative ease in reading books. On the other hand, students—especially elementary school students—do not have perfect knowledge of their ability to perform on tests nor how they “measure up” to other students, thus suggesting an unobserved component to ability. If \( A_i \) is not observed to the econometrician, then Assumption 5.1 may not hold. For instance, students who perceive themselves to be more capable may respond better to accountability policies. By similar logic, the fact that peer “observed ability,” \( A_{-i} \), is unobserved to the econometrician brings into doubt Assumption 5.2.

Previous studies have dealt with the problem of unobserved ability by controlling for lagged achievement.\(^{48}\) This strategy is not desirable in the present setting because lagged achievement may also capture important strategic aspects such as the previous effort of the student and his peers, particularly when peer groups do not vary much from one grade to the next, as is the case when students stay in the same elementary school.\(^{49}\) I use an alternative proxy for observed reading ability, namely the portion of leisure time spent in reading for fun. This is somewhat analogous to the strategy used in Olley and Pakes (1996) and Levinsohn and Petrin (2003) that recovers the unobserved state variable (in their context firm productivity rather than ability) from observed inputs to production.

Denote the time student \( i \) spends reading for fun and watching television by \( R_i \) and \( TV_i \). Partition \( X_i \) into two components, observed ability \( A_i \) and observed characteristics \( X_{1i} \), i.e., \( X_i = (X_{1i}, A_i) \). Similarly partition \( \bar{X}_{-i} \) into two components, observed elements \( \bar{X}_{1i} \) and unobserved ability \( \bar{A}_{-i} \), such that \( \bar{X}_{-i} = (\bar{X}_{1i}, \bar{A}_{-i}) \). I assume that free reading time does not enter the production function because it is an activity that takes place outside of the classroom.\(^{50}\) Partitioning individual and peer characteristics into the components that are observed and those that are unobserved to the econometrician, the achievement

\(^{48}\)See Hanushek et al. (2003) for a detailed discussion.

\(^{49}\)In the linear-in-means context when panel data are available, as in the present setting, an obvious strategy for dealing with the unobserved individual heterogeneity is to include individual fixed effects. I do not pursue this strategy for several reasons. It does not generalize very well to the nonparametric quantile context. (See Koenker (2004) for a description of quantile regressions using longitudinal data.) Second, even if I were to assume the location shift model as in Koenker (2004), one cannot difference out the fixed effect as would be feasible in the linear-in-means case and therefore must compute the model with a large set of dummy variables. This is not computationally feasible in the present context due to the large sample size and the excessive memory requirements needed to run such a specification.

\(^{50}\)This assumption is not necessary, but does affect the interpretation of the estimates. If free reading time does enter achievement production directly, then the interpretation of the marginal effect of free reading time would partially include an indirect effect coming from the effect of ability on achievement.
best-response function for a student \( i \) is as follows:

\[
Y_i^* = q(Y_{-i}^*, X_1^i, \tilde{X}_1^i, P_i, K, A_i, \tilde{A}_{-i}, \mu, \theta_i).
\]

The residual now consists of four unobserved components, \((A_i, \tilde{A}_{-i}, \mu, \theta_i)\). Similarly, expected peer achievement, as represented in (5.3), becomes:

\[
\bar{Y}_{-i}^* = h(X_1^i, \bar{X}_{-i}, P_i, \bar{P}_{-i}, K, A_i, \tilde{A}_{-i}, \mu).
\]

For the control function approach to work in this context, it would be necessary to assume that this residual can be written as a scalar composite of \((A_i, \tilde{A}_{-i}, \mu)\). This seems unnecessarily restrictive, since these components may have important effects on peer achievement in ways that do not necessarily move together. Furthermore, Assumption 5.2 would need to be modified such that the instrument is independent of observed ability and peer ability, as well as the unobserved group effect. Since both of these assumptions seem implausible, I propose an alternative solution below.

Students differ in their leisure time \((T_i \equiv TV_i + R_i)\), which is the sum of time spent watching television \((TV_i)\) and reading for fun \((R_i)\). For instance, some may face longer bus rides to/from school; others may participate more in after-school activities. Some teachers may assign more homework, and some students may take longer to complete homework assignments. All of these factors can detract from the amount of leisure time at a student’s disposal. The portion of leisure time spent reading for fun is assumed to take the following form:

\[
R_i/T_i = \rho_i(X_1^i, A_i).
\]

Observed leisure time allocation may follow from either student optimization or parental mandate. Recall that \(X_1^i\) includes observable characteristics of the parent (such as education) that may make them more or less likely to promote reading at home.

Assumption 5.5. The portion of leisure time spent reading for fun is strictly increasing in ability, \(\partial(R_i/T_i)/\partial A_i > 0\).

Identification of \(A_i\) relies on the intuition that the portion of leisure time spent reading for fun and ability are positively correlated. When parents are the decision makers, the positive correlation arises simply because parents who encourage their children to spend more leisure time reading for fun will have more capable readers relative to parents who do
not. When the student is the decision maker, more capable readers may choose to spend more time reading for fun because the activity is less costly or they derive greater intrinsic benefits relative to less capable readers. It is important to emphasize that I focus on free reading time as a portion of total leisure time because the total amount of leisure time may vary greatly across students in ways unrelated to ability.

Under Assumption 5.5, ability can be expressed as a function of observable characteristics—free reading time, leisure time and individual characteristics, i.e.,

$$A_i = \rho_i^{-1}(R_i/T_i, X^1_i).$$

(5.4)

Similarly, peer ability can be approximated as $$\bar{A}_{-i} = (1/(N-1)) \sum_{j \neq i} \rho_j^{-1}(R_j/T_j, X^1_j) \equiv \bar{\rho}_{-i}^{-1}(R_{-i}/T_{-i}, X^1_{-i}).$$ Effectively, I am able to control for the portion of ability that is observed to the individual and his peers through the leisure-time choices that reveal his type.

Plugging in for observed ability and peer ability in the achievement equation,

$$Y_i^* = q(\bar{Y}_{-i}, X^1_i, X^1_{-i}, P_i, K, \rho_i^{-1}(R_i/T_i, X^1_i), \bar{\rho}_{-i}^{-1}(R_{-i}/T_{-i}, X^1_{-i}), \mu, \theta_i).$$

### 5.4 Non-Random Assignment

A well-known problem in the identification of peer effects is non-random assignment to classrooms. The primary source of selection bias arises from parents selecting their children’s schools, often through their choice of residence. This is further complicated by non-random assignment to classrooms within schools. Policies for classroom placement diverge widely across schools. In some cases, parents have little to no control over their child’s classroom assignment, while in others they can request certain teachers. When parents exercise control over classroom assignment, the particular concern is that more “attentive” parents, those who select the better teachers, may at the same time have higher-achieving children. By selecting better teachers, they effectively select into better peer groups. This might lead one to mistakenly conclude that positive peer achievement spillovers exist, when in fact the positive correlation in outcomes stems from selection. Furthermore, school administrators may not assign students randomly to classrooms, but may instead employ some sort of ability tracking. This latter type of non-random assignment is less of a concern in the present setting given that the production function includes ability proxies.
Selection in the present context is problematic only if students are assigned to classrooms based on the unobserved \( \mu \). However, the primary type of selection that occurs within schools is likely to be on perceived teacher quality. Further assuming that teachers stay in similar “quality” schools over the period of the data, including teacher fixed effects controls for non-random assignment to classrooms and schools. Formally, partition classroom level inputs into the teacher fixed effect \((Tch)\) and other inputs, \( K = (Tch, K^1) \). I modify (5.2) as follows:

\[
Y_i^* = \tilde{q}(\bar{Y}_{-i}^*, X_i, \bar{X}_{-i}, P_i, K^1, \mu, \theta_i) + \beta(\theta_i)Tch,
\]

where the teacher fixed effect enters flexibly in the form of a quantile-specific location shift, with different effects on different types of students.

### 5.5 Discussion

A growing body of research treats the difficulties associated with identifying peer effects.\(^{51}\) One issue lies in what Manski (1993) calls the reflection problem. He identifies three reasons why members of a group may behave similarly or experience similar outcomes and finds that there are significant challenges inherent in separating these effects. First, unobserved group-level effects might lead to similar outcomes. Particular to the context of education production, teacher quality has proven difficult to measure\(^{52}\) and in the absence of adequate controls could lead one to mistakenly attribute similarity in outcomes within the classroom to peer effects. Second, peer characteristics may influence outcomes. In particular, previous studies have found that students benefit from more able peers and peers with better-educated or wealthier parents.\(^{53}\) Finally, peer behavior may directly influence individual behavior. For instance, a student may be more likely to disrupt the class if placed in a classroom with more disruptive peers. Disruptive behavior, in turn, is likely to result in lower achievement for the whole class.\(^{54}\)

In particular, the pioneering work of Manski (1993) presents two negative identification results for the linear-in-means model in the absence of imposing exogenous identifying assumptions. First, social effects (exogenous and endogenous) cannot be separated from un-

\(^{51}\)See Brock and Durlauf (2001b) for an overview.
\(^{52}\)See Hanushek et al. (1998).
\(^{53}\)See Betts and Zau (2002), Hanushek et al. (2003), and Vigdor and Nechyba (2004).
\(^{54}\)See Lazear (2001) for a theoretical model and Figlio (2003) for an empirical application.
observed group effects, or correlated effects. Second, exogenous effects cannot be separated from endogenous effects. The literature on peer effects in education simplifies the problem by focusing on exogenous peers effects, while minimizing the importance of endogenous peer effects in the context of achievement production.\textsuperscript{55} Graham and Hahn (2004) recast the first problem as the panel data problem of identifying the coefficient on time-invariant regressors after including a fixed effect. They interpret the cross-sectional dimension of social interaction models to be the peer groups and the time dimension as the individuals. The problem reduces to finding an instrument that provides exogenous between-group variation. Graham (2004) uses excess variation in random assignment to large and small classrooms to identify the presence of social effects and the magnitude of the social multiplier in a linear-in-means setting. However, he needs a Bayesian informative prior on the exogenous effects parameter to compute the posterior distribution of endogenous effects.

While identification in the linear-in-means framework is a natural starting point, in reality many models are non-linear in nature. Furthermore, the linear-in-means model has limited implications for important policy questions where distributional effects matter. Brock and Durlauf (2001a) consider peer effects in a discrete choice setting and find that the endogenous peer effect can be identified. As they explain, their positive identification result relies on assumed nonlinearities between the relative probability of the outcome and its determinants. Sweeting (2004) shows that the multiplicity of equilibria introduced in a nonlinear setting can aid in identification.

The intuition for my approach to identification follows more along the lines of Moffitt (2001), who formulates the reflection problem as a classic simultaneous equations problem. This approach is particularly relevant when one moves beyond the familiar concept of education production functions as relating fixed inputs to some measure of output to production functions as best responses, as described in Section 4. Moffitt (2001) considers the context of two individuals in a linear-in-means framework and discusses potential exclusion restrictions that would permit identification of both endogenous and exogenous peer effects.\textsuperscript{56} I dis-

\textsuperscript{55} For example, Hanushek et al. (2003), Betts and Zau (2002), and Vigdor and Nechyba (2004) use variation in grade and class peer composition respectively to study the effect of observed peer characteristics—race, income and ability as measured by lagged achievement. Hanushek et al. (2003) discount the importance of contemporaneous effects and provide a detailed discussion of the conditions under which fixed effects can be used in combination with lagged achievement and a value-added specification to attain identification of exogenous effects.

\textsuperscript{56} As he discusses, one difficulty in determining the validity of any given exclusion restriction is our limited understanding of the underlying behaviors that lead to the production functions that we seek to estimate. This, in part, explains my attempt to carefully model the incentives of students in Section 4 to shed light...
cuss conditions for identification in a more general, nonparametric setting. While my paper contributes most directly to the identification of peer effects in education and in a social interactions context, models of strategic interactions among agents are also important in the industrial organization literature, particularly in models of entry\textsuperscript{57} and auction models.\textsuperscript{58}

Recent papers on the instrumental variable identification of quantile structural functions that are not restricted to be additively separable in the unobserved component provide the basis for the identification result presented in this paper. I describe the conditions that permit quantile instrumental variables techniques (Imbens and Newey (2003), among others) to be applied to the identification of peer effects. Chernozhukov and Hansen (forthcoming) present an alternative non-linear quantile IV approach, which allows for identification of the quantile structural function for systems of the form in (5.2) and (5.3) without the control function assumption. In this version of the paper, I use the control function approach because it allows me to say something about the correlated effect from the first stage, which is an important aspect of the peer effect model.

6 Estimation

Suppose time is indexed $t = 1, ..., T$ and classrooms $c = 1, ..., C$. Estimation of the QSF proceeds in two steps. First, I recover the residual from Equation (5.3), the reduced form regression predicting the ex ante expected value of peer achievement. I then estimate the QSF defined in Equation (5.5). Each of these steps is discussed in detail in the following sections.

6.1 Reduced Form

The first stage of the estimator is analogous to the first stage in 2SLS estimation. I recover the unobserved classroom productivity, the correlated effect, from a reduced-form equation for peer achievement, as described in (5.3). By Theorem 5.1, I can recover $\mu$, which captures the sources of exclusion restrictions.

\textsuperscript{57}For instance, see Ackerberg and Gowrisankaran (2005), Aguirregabiria and Mira (2004), Bresnahan and Reiss (1991), Ciliberto and Tamer (2004), Pakes et al. (2005), Pesendorfer and Schmidt-Dengler (2003), and Rysman (2004).

\textsuperscript{58}See Athey and Haile (Forthcoming) for an overview of this literature.
the correlated effect in the model.

As discussed previously, an important dimension of this analysis involves allowing the spillovers to vary across races and for different race-based reference groups. Let $NW_i$ be an indicator for a nonwhite student, and the superscripts $k \in \{W, NW\}$ indicate white and nonwhite respectively. Then, $\bar{Y}_{\cdot NW}^{\cdot ICT} = \frac{1}{\sum_j NW_j} (\sum_j NW_j Y_j^* - NW_i Y_i^*)$ denotes the observed mean achievement of student $i$’s nonwhite classroom peers and similarly $\bar{Y}_{\cdot ICT}^{W}$ for white peers. The reduced-form equation for peer achievement is approximated as follows:

$$\bar{Y}_{\cdot ICT}^k = \alpha_0 + X_{it} \alpha_1 + \bar{X}_{\cdot ICT} \alpha_2 + \alpha_3 P_{it} + \bar{P}_{\cdot ICT} \alpha_4 + K_{ct} \alpha_5 + Tch_{it} + Year_t + \mu_{ct} + \delta_{ict}, \quad (6.1)$$

where dependence of the parameters on the race subgroup $k$ is suppressed. The covariates $X_{it}$ include sex, parental education, a low-performing dummy defined as whether the student performed below the 30th percentile in the previous year, and the portion of leisure time spent reading for fun as an ability proxy. The mean characteristics of $i$’s peers are captured by $\bar{X}_{\cdot ICT}$, which includes the peer average for each of the above, i.e., the percentage of peers with college-educated or high-school-educated parents, the percentage of peers who are low performing, the average portion of leisure time peers devote to free reading, and the percentage of nonwhite students in the classroom. Other than teacher fixed effects, $Tch_{it}$, classroom level-inputs $K_{ct}$ include an indicator for years/grades when student accountability policies are in place and a dummy for classrooms with no peers of the other race. Finally, $P_{it}$ is the utility shifter; it indicates students for whom Student Accountability policies are binding. Formally, $P_{it}$ is equal to 1 for fifth graders in years 2001 and beyond who are low performers, i.e., if they performed below the 30th percentile in reading in the prior year. The percentage of peers of each race who are held accountable are the instruments for peer achievement, i.e., $\bar{P}_{\cdot ICT} = \{\bar{P}_{\cdot ICT}^W, \bar{P}_{\cdot ICT}^{NW}\}$. The identification strategy relies on the intuition that students in classrooms with larger percentages of low achievers of a given race experience larger shifts in the achievement for peers of that race after student accountability.

The remaining residual $\delta_{ict}$ can be thought of as measurement error and captures the fact that the sample average of observed peer achievement is only an approximation for ex ante expectations of average peer achievement, i.e., $\bar{Y}_{\cdot ICT} = \bar{Y}_{\cdot ICT}^* + \delta_{ict}$. Given that classes are sufficiently large, about 23 students on average, $\delta_{ict}$ should be of relatively small magnitude.

I estimate the two first-stage regressions for white and nonwhite peer achievement separately for students of each race. From these regressions, I recover four estimates of the
correlated effect $\hat{\mu}_{ct} = \mu_{ct} + \delta_{ict}$ as the residual from OLS estimates of (6.1) and four values of the predicted teacher fixed effects, $\hat{\text{Tch}}_{it}$. The triangular structure in (6.1) implicitly approximates peer achievement for multiple peer groups flexibly. An important case when the approximation becomes exact is when there are no cross-subgroup spillovers.

6.2 Quantile Structural Function

In the second stage, I estimate the structural function (5.5), which describes a student’s achievement as a function of peer characteristics and peer achievement at different points of the conditional achievement distribution. Estimating the second stage as a mean regression would obscure potentially important distributional effects of peers. The quantile estimator allows for the bottom tails of the achievement distribution to respond differently to peers than the upper tails. If low-ability students benefit relatively more from increases in peer achievement than high-ability students, tracking students into homogeneous classrooms based on prior achievement or ability is less efficient than mixing students, as in the desegregated setting. In the linear-in-means context, regrouping students has no efficiency implications because the losses to one student are perfectly offset by the gains to another. Thus, the quantile estimator provides a much richer picture of the equity/efficiency trade-offs associated with alternative classroom assignment policies.

Previous studies have pursued alternative strategies to capture nonlinearities in achievement production, categorizing students as high- or low-ability based on prior test scores and estimating mean regressions on different subsets of the sample or including interactions of these dummies with exogenous peers effects.\footnote{See Hanushek et al. (2004) and Hanushek et al. (2003).} Effectively, these strategies provide evidence of the marginal effects at different points of the unconditional achievement distribution. Alternatively, the quantile regression provides evidence of the marginal effects at different points in the conditional distribution. It is difficult to relate the evidence from the unconditional distribution to the parameters of the structural response function. Furthermore, the quantile estimator offers a lot of flexibility and can be estimated for a large number of quantiles.\footnote{This is emphasized in Chernozhukov and Hansen (forthcoming).} Finally, an important benefit of the quantile estimator is that it is not sensitive to outliers.

While it is feasible to estimate the quantile structural function without assuming a para-
metric form, in this setting it is useful to assume a parametric approximation for the system of equations because of the large number of covariates. Therefore, I approximate (5.5) as follows:

\[
Y^*_{ict} = \beta_0 + \beta_1 Y^W_{-ict} + \beta_2 Y^NW_{-ict} + X_{it}\beta_3 + X_{-ict}\beta_4 + \beta_5 P_{it} + K^1_{ct}\beta_6 + \beta_7 Tch^W_{it} + \beta_8 Tch^NW_{it} + Year_t + \beta_9 \hat{\mu}^W_{ct} + \beta_{10} \hat{\mu}^{NW}_{ct} + u_{ict},
\]

(6.2)

where dependence of the parameters on the quantile \(\beta(\theta_i)\) and race is suppressed to simplify notation. To emphasize, an important feature of the structural function is that it permits heterogeneity in responses by student type. Furthermore, teacher fixed effects are allowed to vary by race, capturing the fact that effectiveness of the teacher may vary across races. This would control for discrimination, if, for instance, teachers form lower expectations of the capability of nonwhite students that in turn inhibits these students’ performance. The \(\hat{\mu}_{ct}\)'s capture the unobserved classroom characteristics, or correlated effects, that simultaneously affect the achievement of an individual and his peers. These enter achievement in a flexible way, with the marginal effect permitted to vary both by race and quantile. Finally, it is worth noting that this approximation of the achievement best-response predicts a unique equilibrium.

For each subgroup and a given quantile \(\theta_{it} = \tau\), parameter estimates \(\hat{\beta}(\tau)\) solve the following optimization problem:

\[
\arg\min_{\hat{\beta}(\tau)} \frac{1}{NCT} \sum_i \sum_{c} \sum_{t} \rho_{\tau}(u_{ict}),
\]

where \(\rho_{\tau}(u_{ict}) = \tau u^+_{ict} + (1 - \tau) u^-_{ict}\).

Before proceeding to the results, the interpretation of the exogenous effects (\(\hat{\beta}_4\)) and correlated effects (\(\hat{\mu}_{ct}\)) merits further discussion. First, consider exogenous effects. If there were no contemporaneous peer spillovers and students did not choose effort, then \(\bar{X}_{-ict}\) would only affect \(i\)'s achievement through the state variables that enter his achievement directly.

---


63 Alternatively, for the typical 2SLS estimator, one could plug in the fitted values of peer achievement from the first stage in place of observed peer achievement and remove \(\hat{\mu}_{ct}\). However, this is not consistent with the model, in which students make a best respond to \(\hat{Y}_{-ict}\) not \(\hat{Y}_{-ict}\).
This is the way that exogenous effects are generally thought about in the literature. In this case, we would expect that increasing, say, the percentage of peers with high parental education would have a positive effect on $i$’s achievement. However, when student $i$ is able to choose effort, it is unclear whether increasing $\bar{X}_{ict}$ will have a positive or a negative effect. For any level of effort, students with higher $\bar{X}_{ict}$ would attain a higher level of achievement, suggesting that raising $\bar{X}_{ict}$ could lower effort, i.e., effort and peer characteristics could act as substitutes. On the other hand, any amount of effort may also be more productive as a result of the “better” peer group, suggesting a higher level of optimal effort, in which case effort and peer characteristics are complementary. Finally, when there are spillovers from peer effort, $\bar{X}_{ict}$ has a direct effect on $i$’s achievement through peer effort. Holding $\bar{Y}_{ict}$ fixed, higher $\bar{X}_{ict}$ would predict lower peer effort and therefore have a negative effect on $i$’s achievement. Given these three countervailing effects, the sign of $\hat{\beta}_4$ is indeterminate.\footnote{Appendix A.2 illustrates two of these opposing forces using a simple parametric form of the production and utility functions.}

A similar conclusion holds for classroom productivity, $\mu$. It is generally assumed that OLS estimates of the peer effect are biased upwards due to these unobserved correlated effects. Similar to the argument for peer characteristics above, this makes sense if $\mu$ is only an element in $S_i$, students do not choose effort, and there are no spillovers from peer effort in production. Otherwise, higher $\mu$ could negatively affect achievement through its inverse relation with peer effort. Secondly, higher $\mu$ could predict higher or lower utility-maximizing effort. Thus, careful consideration of the strategic behavior of students suggests that the assumption of an upward bias from unobserved correlated effects may not hold.

7 Results

This section presents results on the magnitude of peer spillovers to achievement and distributional effects. To provide a baseline comparable to previous estimates in the literature, Section 7.1 describes estimates of the endogenous peer effect in a linear-in-means model, comparing 2SLS to OLS with contemporaneous and lagged peer achievement. In Section 7.2, I describe results for the two-stage quantile estimator described in Section 6, and allow peer spillovers to vary by race. In Section 7.3, I compare the relative magnitude of exogenous and endogenous peer effects.
### 7.1 Baseline: Linear-in-Means

Before exploring the distributional effects of peers, it is useful to begin with two-stage least squares estimates of the linear-in-means model, which are most readily comparable to previous estimates in the literature. I modify the first stage of the 2SLS estimator so that the dependent variable is peer achievement for the whole class (not broken out by race subgroups). The second step is as follows:

\[
Y^*_\text{ict} = \beta_0 + Y_{-\text{ict}}\beta_1 + X_{it}\beta_2 + X_{-\text{ict}}\beta_3 + \beta_4 P_{it} + K_{ct}\beta_5 + Tch_{it} + Year_t + \beta_6 \hat{\mu}_{ct} + \theta_{ict},
\]

where all exogenous effects \(X_{-\text{ict}}\) are defined at the class level.\(^{65}\) Parameter estimates from the first and second stages of this estimator are presented in the second two columns of Table 3. I provide two comparison cases: (1) the “naive” case, an OLS regression that does not account for simultaneity in achievement and (2) OLS using twice-lagged peer achievement to break the simultaneity—shown in the first two columns of Table 3. Justification for the latter estimation strategy relies on the assumption that contemporaneous endogenous peer effects are negligible.\(^{66}\)

I begin by comparing estimates of the endogenous effect \((\beta_1)\) across the three estimators. The naive estimates predict an average peer effect of -.07, in comparison to .48 for the 2SLS estimator. Using twice-lagged peer achievement to proxy for the endogenous peer effect more severely underestimates peer spillovers with a coefficient of -.15. If the direct effect of unobserved group productivity \(\mu\) on achievement dominates, one would expect the naive estimates of the peer effect to be biased upwards relative to 2SLS. However, I find that this is not the case. The downward bias of OLS can be explained if the dominant effect of \(\mu\), after controlling for teacher fixed effects, is coming through peer effort spillovers to production or strategic effort choices rather than a direct effect.\(^{67}\) To interpret the 2SLS estimate, it predicts that a one standard deviation increase in peer achievement (.45) leads to about 22% of a standard deviation increase in achievement.\(^{68}\)

---

\(^{65}\)Note that in the linear-in-means case, it is easy to control for the teacher fixed effect directly and it is also no longer necessary to include the controls for classrooms without a given subgroup in \(K_{ct}\).

\(^{66}\)These estimators skip the first-stage regression (i.e., \(\hat{\mu}_{ct} = 0\) in the above equation).

\(^{67}\)These results of downward bias in OLS estimates are very sensitive to changes in specification due to the countervailing forces acting through utility-maximizing effort and peer effort as described in Section 6.2. For instance, with school by year fixed effects (rather than teacher fixed effects), the naive OLS results are biased upward relative to 2SLS and lagged OLS results give a small but positive coefficient on peer achievement.

\(^{68}\)As an aside, it is worth noting that this estimate is within the 95% Bayesian credibility set, .0827 to .6976, calculated in Graham (2004), using a similar linear in means model.
Estimates of the exogenous peer effects also differ considerably across the estimators, with marginal effects frequently opposite in sign. For instance, lagged OLS predicts relatively small but positive effects from increasing peer parental education, while 2SLS predicts a larger negative effect. Intuitively, this negative effect follows if the spillovers from peers’ parental education arise primarily through peer effort, which is ignored in the twice-lagged specification. In other words, suppose that the parental education of $i$’s peers does not directly affect student $i$’s achievement. Then, holding peer achievement fixed, higher parental education of $i$’s peers suggests lower peer effort. This intuition can be extended to the other exogenous effect parameters to help explain the marked differences between the lagged estimator that ignores contemporaneous peer spillovers and the 2SLS estimator.

Finally, the first-stage regression in Column 3 presents evidence regarding the validity of the exclusion restriction. The percentage of students held accountable is a statistically significant predictor of peer achievement, with a sizable marginal effect of .17. Thus, students in classes with higher percentages of low performers witnessed greater increases in peer achievement from accountability. Furthermore, the second-stage regression reveals that the advent of student accountability policies appears to have limited effects on high achievers (-.01), but a much larger effect (.10) on those who are designated “low reading” students by prior years’ test scores. These estimates are consistent with the story that student accountability did not entail a redistribution of resources away from high achievers, but rather induced low performers to work harder and boost their achievement.\footnote{Note that if resources were redistributed from high to low performers, this would suggest that the endogenous spillovers are even larger.}

### 7.2 Heterogeneous Reference Groups and Distributional Effects

Turning to the main results of the paper, I consider two relatively unexplored aspects of peer effects that are particularly relevant to policymakers. First, I test whether students respond differently to peers more similar to themselves in observable dimensions, namely race. Then, I proceed to consider distributional effects, i.e., how the marginal effect of peer achievement varies across percentiles of the achievement distribution.

Table 4 presents estimates of the endogenous effects parameter using the procedure de-
scribed in Section 6 and comparing the mean and the median case $(\beta_1(.5), \beta_2(.5))$. Each column corresponds to a separate regression for the given race, while the rows describe the achievement spillovers from peers of each race. Both the 2SLS and median two-stage quantile estimators predict that white students receive positive spillovers from their white peers, .40 and .39, but minimal spillovers from their nonwhite peers, .02. On the other hand, in the mean case nonwhite students receive positive, but not statistically significant, spillovers from both their white and nonwhite peers, .20 and .16, while in the median case they receive positive and significant spillovers from their nonwhite peers, .27, and smaller insignificant spillovers from their white peers, .09. The standard errors are particularly large for the effect of white peers on nonwhite students, which may be a result of two countervailing influences affecting the preferences of nonwhites. On the one hand, if nonwhites perceive achievement to be associated with the negative stereotype of “acting white,” this disutility from achievement may be further reinforced when they are placed with higher-achieving white peers. On the other hand, if the dominant effect is coming through a classroom disruption model, or congestion effect, rather than the effect of peers on preferences, there should be positive spillovers from both white and nonwhite peer achievement. At least for the median case, the finding of spillovers only from “like” peers is more consistent with a conformity effect story, i.e., that the dominant effect of peers is coming through the psychic costs of deviating from peer-defined norms of behavior.

While Table 4 presents compelling evidence that peer spillovers vary by race, Figure 6 provides a richer understanding of the distributional effect of peers, i.e., how the marginal effect of average peer achievement varies across quantiles for each race. The story is fairly consistent with that for the median case where white students respond almost entirely to white peers, while nonwhite students respond to both white and nonwhite peers. Interestingly, I find evidence of strongly diminishing marginal returns to white peer achievement for whites, from a high of .73 to a low of .22. In contrast, the marginal effect of nonwhite peers appears to be weakly increasing for nonwhites, from a low of about .17 to a high of .30. This suggests that at least within the same race, detracking classrooms would be more efficient (i.e., would raise overall achievement) for white students and slightly less efficient for nonwhites, while creating more equitable outcomes relative to tracking students into high- and low-achieving classes. The effect of white peers on nonwhites is somewhat U-shaped,\textsuperscript{71} Fryer and Torelli (2005) find that the prevalence of “acting white” depends on the racial composition of schools, with predominantly black schools displaying no evidence of these negative peer effects. I try interacting peer achievement with the percent nonwhite and breaking the sample into predominantly nonwhite and predominantly white schools, but the results are not statistically significantly different.

\textsuperscript{71} Fryer and Torelli (2005) find that the prevalence of “acting white” depends on the racial composition of schools, with predominantly black schools displaying no evidence of these negative peer effects. I try interacting peer achievement with the percent nonwhite and breaking the sample into predominantly nonwhite and predominantly white schools, but the results are not statistically significantly different.
with sizable effects on the lower and upper tails of the achievement distribution of .31 to .27 and smaller effects in the center, .09. Thus, there is some evidence that nonwhites receive positive spillovers from white peers, however the standard errors are fairly large, so that the effects are only statistically significant for the upper tail of the achievement distribution. This is indicative of the particularly complex nature of spillovers from whites to nonwhites documented in the literature.

7.3 Endogenous vs Exogenous Effects

Exogenous peer effects have served as the focal point of the literature on peer effects in education, though the estimates in the previous section provide compelling evidence that endogenous effects do indeed exist. In this section, I compare the relative magnitude of endogenous and exogenous effect parameters and consider the implications of exogenous spillovers for desegregating classrooms.

Tables 5 and 6 present marginal effects for whites and nonwhites of a one standard deviation increase in each of the peer variables using the estimates from the two-stage quantile regression (2SQR) corresponding to those shown in Figure 6. The first column presents the average over quantiles within a given race, while the remaining columns present the marginal effect for a given quantile and race. The first two rows describe the marginal effect of peer achievement in each subgroup. The marginal effects of peer achievement are stronger for whites than for nonwhites. The effect of a one standard deviation increase in white peer achievement is .18 for whites and .11 for nonwhites, while the effect of a one standard deviation increase in nonwhite peer achievement is 0 for whites and .11 for nonwhites. Thus, for nonwhites the marginal effect of white and nonwhite peer achievement is about the same on average.

This table also reveals that increasing the percentage of nonwhite students has a small negative effect on whites and nonwhites, -.02. This is consistent with evidence in Hanushek et al. (2004) that high concentrations of nonwhites have a negative effect on the achievement of both whites and nonwhites. The effect of racial composition is further illustrated in Figure 7, which compares the quantile derivatives of the portion of peers who are nonwhite on white and nonwhite achievement. The effect of percentage nonwhite on nonwhites is diminishing across quantiles, with nonwhites in the highest quantiles receiving only small, and statistically insignificant, negative effects. It is worth pointing out that since I do not
include any income controls, the effect of the higher concentration of nonwhites could be picking up an income effect. Furthermore, as discussed previously, it is generally difficult to interpret the exogenous effects. This is further illustrated in considering the other exogenous effects.

Seemingly contrary to intuition, the marginal effect of peers’ parental education is negative for whites and nonwhites. This negative effect would arise if, say, i’s parental education has little direct effect on his peers’ achievement, but the exogenous effect is operating primarily through spillovers from peer effort. It is more difficult to come up with a story of a strong direct effect of peer parental education on achievement in the present context, since the estimates already control for selection with teacher fixed effects. Similarly, the percentage of peers who are low performing has a seemingly counterintuitive positive effect on achievement. Whites are only affected by the percentage of white peers who are low performing, a marginal effect of .06 for a one standard deviation increase in the percentage of white peers who are low performing. Nonwhites have a smaller but positive marginal effect from the percentage of both white and nonwhite peers who are low performing, .04, but this is insignificant. The fact that endogenous peer effects are much larger in magnitude than exogenous peer effects suggests that the former will play the dominant role in any changes to achievement from desegregating peer groups. Nonetheless, it is important also to account for the role of exogenous effects. Section 8.1 paints a clearer picture of the effect of desegregating peer groups taking into account both endogenous and exogenous effects.

8 Counterfactual Classroom Assignment Policies

With parameter estimates of the achievement production function now in hand, I turn to the central question of the paper: What is the effect of desegregating peer groups on the achievement gap between white and nonwhite students? To reiterate the findings of the previous section, the lack of cross-racial spillovers for whites suggests minimal negative consequences to white achievement from being grouped with lower-achieving nonwhite peers. Nonwhites at all percentiles of the achievement distribution stand to gain by being grouped with higher-achieving white and nonwhite peers. While the dominant effect of peers appears to be coming through these endogenous effects, exogenous effects will also play a role. Since a higher concentration of nonwhites is negatively correlated with achievement (perhaps picking up an income effect), this would suggest that creating more diverse peer groups would
raise nonwhite achievement while lowering white achievement and narrowing the achievement gap. Disparities in parental education and free reading time across races will also enter into the overall effect of any student reassignment policy. Thus, while it appears that desegregating peer groups may indeed help to narrow the achievement gap, it remains difficult to conjecture the magnitude of the effect and the efficiency implications.

To elucidate the equity/efficiency trade-offs of desegregating peer groups, Section 8.1 simulates achievement outcomes in an unconstrained setting where students can be assigned to any classroom in the state of North Carolina. In particular, I compare the outcomes under desegregated peer groups to those under observed peer groups and the controversial alternative of tracking. Clearly, this abstracts away from important issues of residential sorting, proximity constraints and the potential to select out of public schools. However, the experiment provides evidence regarding an upper bound on the effects of desegregation on the achievement gap, holding the student population fixed.

In reality, school desegregation policies are almost exclusively restricted to be within district lines, a result of the Supreme Court’s ruling in the 1974 case of Milliken v. Bradley that a federal court could not force integration across district lines. As a result, it is not uncommon to find wealthy, high-achieving, predominantly white school districts next to poorer, lower-achieving, racially-mixed school districts. With these observations in mind, I also consider a more localized experiment of desegregating schools across the lines of two such school districts in North Carolina.

In choosing where to live, parents place a high premium on the quality of the local public schools, and an important measure of the quality of the school is the quality of the peer group. In turn, the measure of school quality, in large part, lies in the social composition of peer groups, a finding that received a great deal of attention with the publication of the Coleman Report of 1966 (Coleman et al. (1966)). In Section 8.2, I estimate the achievement premium associated with moving a student from a low-income, racially-mixed public school district to an adjoining high-income, predominantly white district. This provides evidence on the returns to “better” peer groups in a context where general equilibrium sorting effects are unlikely to be important.
8.1 Desegregating Peer Groups

In this section, I estimate the treatment effects for alternative classroom assignment policies. North Carolina policymakers advocate desegregated/mixed-ability classrooms as most conducive to narrowing the achievement gap. To test this, first I randomly assign students to classrooms, thereby creating racially desegregated and mixed-ability peer groups. I compare the outcomes under diverse peer groups to the controversial alternative policy of tracking, where students are placed into classrooms with peers who performed at a similar level based on standardized test scores in the previous year. While various efficiency and equity measures are relevant, I begin by comparing the average achievement for whites and nonwhites, the gap between average white and nonwhite achievement and the overall average achievement under these alternative assignment policies. I restrict attention to fifth graders in the 2002-03 academic year and assign to every student the average teacher fixed effect and average correlated effect to ensure that variation in school productivity across race does not obscure the impact of altering peer groups. Furthermore, I assume that students can be assigned to any classroom/school in the state of North Carolina.

The first three rows of Table 7 present simulated achievement, using the preferred two-stage quantile regression described in Section 6. Comparing the first two rows of Table 7 reveals that desegregating classrooms does help to narrow the achievement gap, but only by .06 of a standard deviation relative to the observed assignment policy. White achievement does not appear to be affected much by desegregation on average, and the overall improvement is driven almost entirely by gains to nonwhite achievement. Therefore, random assignment is marginally more efficient and more equitable than the observed assignment policy.

While desegregation does help narrow the achievement gap, the improvements are small. Given the large magnitude of peer spillovers, it is interesting to consider whether alternative regrouping strategies may be more effective. Since comparing all possible assignment poli-

---

72See footnote 3.
73In the simulations, it is necessary to assign a value of $\theta_i$ to each student. Consistent with the story that the residual captures some persistent, unobserved ability, these simulations use the residual recovered from observed outcomes. Alternatively, if $\theta_i$ captures more of a random shock, I could assign students randomly to quantiles. Since the reality is likely to be somewhere in between, I try the simulations both ways and using random rather than recovered residuals does not affect the conclusions.
74Note that in calculating achievement under the observed grouping, I also use the average teacher fixed effect and average correlated effect so that the comparisons in this table isolate peer group effects.
cies presents a much larger (though interesting) problem, for the moment I compare both the observed and random assignment policies to the controversial alternative of tracking. Generally, tracking is viewed as being unfavorable to nonwhite students and favorable to whites, but the results in row 3 of Table 7 suggest otherwise. I find losses to nonwhites of .03, but even larger losses to whites, .07. Contrary to conventional wisdom, I find that tracking also narrows the average achievement gap by .02 of a standard deviation, about a third of the improvement under random assignment. To provide a richer understanding of the distributional impact of these alternative assignment policies, I compare the gains from random assignment and tracking relative to the observed grouping for white and nonwhite students at different percentiles of the achievement distribution.

Figure 8 shows that while the percentile gains are small, the lower quantiles benefit more from random assignment than the upper quantiles, .05 compared to -.01. The change in achievement for whites hovers around 0 at all quantiles, while nonwhites make fairly uniform gains across the quantiles, ranging from .04 to .06. In contrast, tracking has fairly large, negative effects on the achievement of students in the lower quantiles and positive effects on the upper quantiles, ranging from -.23 to .16 for the overall population. The lower quantiles of whites sustain larger losses than nonwhites, -.33 compared to -.15, whereas whites in the upper quantiles generally have higher gains than nonwhites. The larger losses for whites at the lower quantiles is most likely driven by their particularly large spillovers from white peers; they stand to lose the most by being placed with lower-achieving white peers.

An innovation of this paper is allowing for heterogeneity in peer effects by estimating quantile rather than mean regressions. To provide evidence on the importance of this dimension, I compare predictions from the 2SLS estimator, where the marginal effects are allowed to vary by race, to the predictions of the 2SQR estimator. The comparable average treatment effects from the 2SLS estimator are shown in the last 3 rows of Table 7. The 2SLS estimator predicts a smaller narrowing in the achievement gap, .03, which is half of that predicted by the 2SQR estimator. This is because 2SLS predicts small gains to white achievement on average and slightly smaller gains to nonwhite achievement than 2SQR. Furthermore, as Figure 9 shows, if one is concerned about alternative measures of the achievement gap, such as the percentile gap, the predictions of the 2SLS and 2SQR estimators are quite different. The 2SLS estimator severely overpredicts the achievement gap for the lower to middle quantiles.

Furthermore, Figure 10 compares the predicted change in the white-nonwhite achieve-
ment gap in moving from the observed peer groupings to either random assignment or tracking across the two estimators. The 2SQR estimator predicts that tracking increases the achievement gap at the upper percentiles of the achievement distribution and decreases the gap at the lower percentiles, while random assignment decreases the achievement gap at all percentiles. The predicted change in the gap is very different under the 2SLS estimator. It predicts that tracking increases the gap at the lower percentiles, while random assignment increases the gap at some of the lower percentiles and decreases the gap at the upper percentiles.

Next, I consider a more localized case of merging two neighboring school districts—one that is predominantly white and high-achieving, the other predominantly nonwhite and lower-achieving. In particular, this experiment provides implications of limiting desegregation efforts to within school districts. The neighboring public school districts of Durham County (home of Duke University) and Chapel Hill City (home of the University of North Carolina-Chapel Hill) provide an excellent example of such a contrast in North Carolina. As illustrated in Table 8, nonwhite students’ reading test scores are .10 of a standard deviation higher on average in Chapel Hill, while whites perform .40 of a standard deviation higher. The disparity in test scores across the districts stems in part from the sharp contrast in income and education of the parents of students who attend the local public schools. Only 16% of the students in Chapel Hill public schools receive free/reduced-price lunch, compared to 43% in Durham. Similarly, while a majority of parents in Durham (61%) have a high school degree but no four-year degree, a striking majority of parents in Chapel Hill (79%) have at least a four-year degree. Furthermore, Durham public schools have a much larger minority population—68% of fourth and fifth graders are nonwhite, as compared to only 23% in Chapel Hill. That said, Durham public schools are also much more segregated, with a segregation index of .28 as compared to only .04 in Chapel Hill.

Again, I restrict attention to fifth graders in the 2002-03 academic year. It is worth noting that the Durham public school district is much larger—2099 students versus only 667 in Chapel Hill. Table 9 compares the achievement of Durham and Chapel Hill students under observed groupings to the counterfactual achievement realized under (1) desegregation (random assignment to classrooms) within the district and (2) desegregation across district lines. In both districts, white students are marginally better off as a result of within-district desegregation, while nonwhites make larger gains of .05 in both districts. As a result, the achievement gap decreases a little, by .03 in both Durham and Chapel Hill.
Not surprisingly, desegregation across districts produces gains for Durham students and losses to Chapel Hill students. Nonwhites in Durham gain .05 of a standard deviation relative to within-district desegregation, while nonwhites in Chapel Hill lose .02. Whites in Durham make comparable gains to nonwhites, .05, while whites in Chapel Hill lose slightly more than nonwhites, .03. Whereas desegregation within districts lowers the overall achievement gap by .03, desegregation across districts has minimal additional effects on the gap. This is because gains (losses) to nonwhite achievement are offset by gains (losses) to white achievement for students in each of the districts. Figure 11 provides further evidence on the effects of the merger on the gap and achievement at different points of the achievement distribution. The gap decreases at almost all percentiles of the achievement distribution as a result of the merger, and overall gains in achievement relative to the observed grouping are somewhat higher for the lower percentiles than for the upper percentiles. Furthermore, nonwhite achievement increases as a result of the merger at all percentiles of the achievement distribution, with gains ranging from .06 to .15, while whites make smaller gains, ranging from 0 to .08. Finally, the overall variance in achievement does fall some as a result of the merger but not by much, from a standard deviation of .87 under the observed groupings to .84 in the merged/desegregated district.

This localized example provides some intuition for the limited impact of the statewide desegregation experiment on the achievement gap, despite large spillovers from peer achievement. It may be that whites in lower-performing peer groups are benefitting from desegregation almost as much as nonwhites, and this limits the effect on the gap. Furthermore, random assignment in the statewide experiment has small effects on average achievement of each race because the losses from students in “better” peer groups may be offsetting, to some degree, the gains to students in “worse” peer groups.

8.2 Durham to Chapel Hill

While the previous section provides estimates of the benefits to desegregation, there are considerable barriers to actually implementing such a large-scale desegregation plan. For instance, if the districts of Durham and Chapel Hill were to merge, it is unlikely that the population of students currently attending the public schools would remain constant. In fact, the intense desegregation of the 1970s was accompanied by “white flight” from urban school districts to the suburbs, thus hampering efforts to integrate schools. In choosing where to live,
parents often face the trade-off of higher housing prices and “better” schools/peer groups versus lower housing prices and “worse” schools/peer groups. In this section, I quantify the achievement benefits of moving a student from Durham to Chapel Hill, isolating the effect of “better” peer groups. Since Chapel Hill parents are unlikely to select out of public schools as a result of the introduction of one new Durham student, this experiment may be more realistic in that general equilibrium sorting effects are unlikely to be important in this context.

In particular, I conduct a partial equilibrium experiment to estimate the effect on a Durham student’s achievement of randomly assigning him to a classroom in a Chapel Hill public school in comparison to the losses to a Chapel Hill student from random assignment to a Durham school. I compute results for fifth graders in the 2002-03 academic year. The results in Section 6 suggest that moving a nonwhite student from Durham to Chapel Hill may raise his achievement in two ways—from higher-achieving white peers and higher-achieving nonwhite peers. Moving a white student from Durham to Chapel Hill is likely to entail much larger benefits because the pool of white peers is much higher performing.

Table 10 compares the achievement of students in Durham to the simulated achievement for each Durham student from random placement in a Chapel Hill classroom. On average Durham students benefit from placement in Chapel Hill schools, with achievement rising from .34 to .42. Furthermore, whites gain slightly more than nonwhites from the move, .09 compared to .07. To emphasize the negative consequences of concentrating nonwhite students in low-performing schools, the second three rows of Table 10 show gains to achievement for students in Durham schools with high concentrations of low-performing nonwhites. Intuitively, the gains for these students are likely to be higher, given that their peers are generally reinforcing a low-achievement equilibrium. I identify predominantly nonwhite, low-equilibrium Durham schools as those with a student population that is more than 90% nonwhite and where average achievement was below 0 in 2003. There are six such schools in Durham, with a total student population of 297 (11% of Durham’s fifth graders in 2003). On average the gain is much higher for nonwhite students in these low-equilibrium schools relative to the nonwhite students in the other schools, .15, but the gain for whites is even larger, .33.

What are the losses from moving a student from a higher-performing to a lower-performing peer group? Table 10 reveals the significant losses to achievement on average from moving a Chapel Hill student to a Durham public school. Whites lose .17 of a standard deviation
in achievement, while nonwhites lose .07. Thus, the gains to achievement from moving a white Durham student to Chapel Hill are dominated by the large losses to achievement from moving a white Chapel Hill student to Durham, whereas the benefit of moving a nonwhite Durham student to Chapel Hill are comparable to the losses of moving a nonwhite student from Chapel Hill to Durham. Overall, these estimates suggest that there is a significant achievement premium associated with residing in a “better” school district.

9 Conclusion

In answer to the central question posed in this paper, it appears that desegregating peer groups helps to narrow the achievement gap, though only by 6% of a standard deviation. I find evidence of strong spillovers from peer achievement. White students appear to conform only to white peers. While there is some evidence that nonwhites benefit from both white and nonwhite peers, spillovers from nonwhite peers are more robust. Given the absence of cross-racial spillovers for whites, the convergence in white and nonwhite achievement in the desegregated setting is driven primarily from improvements in nonwhite achievement. However, the achievement gap remains sizable even after desegregating all classrooms in North Carolina, suggesting that this may not be the most effective way to close the gap. A more localized example of merging a predominantly white, high-achieving school district with a neighboring lower-performing, racially-mixed school district, where one might expect the returns to be particularly large, has minimal effects on the achievement gap. It appears that the gains for white students are working to offset the gains to nonwhites in lower-achieving districts and similarly the losses for each race are offsetting each other in the higher-achieving districts. On net, the achievement gap does not change much under desegregation, despite large peer spillovers.

A secondary contributor is the racial composition of the peer group. I find that a higher percentage of nonwhites in the classroom is negatively correlated with achievement. I do not control for income, so this may very well be picking up an income effect, especially since schools with large minority populations are often in the poorest neighborhoods. Because nonwhites are typically in peer groups with larger percentages of nonwhite peers than whites, from the standpoint of purely exogenous effects the achievement gap could be narrowed by desegregating peer groups, which would raise nonwhite achievement and lower white achievement. This effect is relatively small, however, since on average white achievement
does not decrease under desegregation.

To assess the impact of segregated schools on achievement, it is worth emphasizing that this paper does not address the benefits of desegregation that may follow from a more equitable distribution of resources across whites and nonwhites. Furthermore, it seems likely that over time the greater interracial contact that results from desegregated schools could itself foster larger cross-racial spillovers in the classroom. Though beyond the scope of this paper, such an effect could eventually serve to narrow the achievement gap.

Beyond presenting evidence on policy concerns close to the heart of both civil rights advocates and education policymakers, this paper contributes to the broader social interactions literature on the identification of peer effects and even more generally to the identification of spillovers to production. I discuss the conditions needed to identify peer effects under minimal functional form assumptions. My estimates show that contemporaneous peer effects are very important to achievement and that in ignoring them, the literature to date has severely underestimated the impact of peers in education production.

To the best of my knowledge, my paper is also the first to apply a flexible nonparametric estimator to the context of peer effects in education or to estimate an educational production function more generally. With evidence of significant variation in responses to peers across race and achievement quantiles, moving from the typical linear-in-means model to a more flexible model leads to a much richer understanding of the impact of peers and alternative grouping strategies. Furthermore, the simulations illustrate that allowing for sufficient flexibility in responses to peers across races and percentiles of the achievement distribution is critical to accurately assessing the equity and efficiency of alternative grouping strategies and in particular the effect on the lower quantiles of the achievement distribution.

Many other pertinent policy questions related to peer effects may be addressed with these production function estimates. One interesting extension of this paper would be to use the production function estimates to recover the objective function of school administrators that is implied by their allocation of students to classrooms. This would help policymakers better understand the trade-offs between equity and efficiency. Another extension would be to use the production function estimates to address questions about the effect of competition on schools after controlling for the peer effect component. Finally, it would be interesting to place these production function estimates in the larger context of a general equilibrium framework that models the school/neighborhood selection mechanism.
References


A Appendix

A.1 Proofs

Proof of Theorem 4.2. In what follows, I illustrate how the game in effort maps into a game in achievement. Given Assumption 4.2, ex ante expected achievement can proxy for effort. Denote the ex ante expected value of achievement ($\tilde{Y}_i$) as follows:

$$\tilde{Y}_i = \tilde{g}(e_i, e_{-i}; S) \equiv \int_{\Theta} g(e_i, e_{-i}; S_i, \theta_i)f(\theta_i|S)d\theta_i.$$  

The following system describes the effort for all students in the classroom as a function of ex ante achievement and peer effort:

$$e_1 = \tilde{g}^{-1}(\tilde{Y}_1, e_2, ..., e_N; S)$$

$$\vdots$$

$$e_N = \tilde{g}^{-1}(\tilde{Y}_N, e_1, ..., e_{N-1}; S).$$

I assume that the solution to this system is unique and is captured by the function $G(\cdot)$, i.e.,

$$e_i = G(\tilde{Y}_i, \tilde{Y}_{-i}; S) \text{ for } i = 1, ..., N.$$  

The vector of peer effort as a function of the vector of achievement and state variables is as follows:

$${e_{-i}} = (...)G(\tilde{Y}_{i-1}, \tilde{Y}_{-(i-1)}; S), G(\tilde{Y}_{i+1}, \tilde{Y}_{-(i+1)}; S), ...)$$

$$\equiv G_{-i}(\tilde{Y}_i, \tilde{Y}_{-i}; S).$$

Therefore, the effort best response can be written as a function of peer achievement, i.e.,

$$e_i^*(e_{-i}; S) = e_i^*(G_{-i}(\tilde{Y}_i, \tilde{Y}_{-i}; S); S)$$

$$= e_i^*(\tilde{Y}_i^*, \tilde{Y}_{-i}; S).$$

Plugging utility-maximizing effort into ex ante expected achievement, we have the
achievement best response of a student $i$ to any level of peer achievement $\bar{Y}_{-i}$:

$$\bar{Y}^*_i = g(e_i(\bar{Y}^*_i, \bar{Y}_{-i}; S), G_{-i}(\bar{Y}^*_i, \bar{Y}_{-i}; S); S).$$

Let $\bar{q}(\cdot)$ represent an explicit solution for $\bar{Y}^*_i$ as follows:

$$\bar{Y}^*_i = \bar{q}(\bar{Y}_{-i}, S_i, S_{-i}).$$

The ex post achievement realized by $i$ under his best response is as follows:

$$Y^*_i = q(\bar{Y}^*_{-i}, S_i, S_{-i}, \theta_i).$$

\[\Box\]

**Proof of Theorem 5.1.** Following the proof in Imbens and Newey (2003):

$$F_{Y^*_{-i}|X_i, X_{-i}, K, P_i, \bar{P}_{-i}}(\bar{Y}^*_{-i}|X_i, \bar{X}_{-i}, K, P_i, \bar{P}_{-i}) \overset{(1)}{=} Pr(\bar{Y}^*_{-i} \leq \bar{y}_0|x_0, \bar{x}_0, k_0, p_0, \bar{p}_0)$$

$$\overset{(2)}{=} Pr(h(X_i, \bar{X}_{-i}, K, P_i, \bar{P}_{-i}, \mu) \leq \bar{y}_0|x_0, \bar{x}_0, k_0, p_0, \bar{p}_0)$$

$$\overset{(3)}{=} Pr(\mu \leq h^{-1}(x_0, \bar{x}_0, k_0, p_0, \bar{p}_0, \bar{y}_0)|x_0, \bar{x}_0, k_0, p_0, \bar{p}_0)$$

$$\overset{(4)}{=} Pr(\mu \leq h^{-1}(x_0, \bar{x}_0, k_0, p_0, \bar{p}_0, \bar{y}_0))$$

$$\overset{(5)}{=} F_{\mu}(h^{-1}(x_0, \bar{x}_0, k_0, p_0, \bar{p}_0, \bar{y}_0)).$$

The first equality follows by definition; the second by the representation of peer achievement in (5.3). The third follows by Assumption 5.3, and the fourth by Assumption 5.2. Therefore, $\mu = h^{-1}(x_0, \bar{x}_0, k_0, p_0, \bar{p}_0, \bar{y}_0)$ is identified by the joint distribution of $(\bar{Y}^*_{-i}, X_i, \bar{X}_{-i}, P_i, \bar{P}_{-i}, K)$. \[\Box\]
Proof of Theorem 5.2. Following the proof in Imbens and Newey (2003, Corollary 6):

\[
F_{Y^*_i | Y^*_i, X_i, \bar{X}_i, P_i, K, \mu}(Y^*_i | Y^*_i, X_i, \bar{X}_i, P_i, K, \mu) \\
= Pr(Y^*_i \leq y_0 | \bar{y}_0, x_0, \bar{x}_0, p_0, k_0, \mu_0) \\
= Pr(q(Y^*_i, X_i, \bar{X}_i, P_i, K, \mu, \theta_i) \leq y_0 | \bar{y}_0, x_0, \bar{x}_0, p_0, k_0, \mu_0, y_0) \\
= Pr(\theta_i \leq q^{-1}(\bar{y}_0, x_0, \bar{x}_0, p_0, k_0, \mu_0, y_0)) \\
= F_{\theta_i}(q^{-1}(\bar{y}_0, x_0, \bar{x}_0, p_0, k_0, \mu_0, y_0)) = q^{-1}(\bar{y}_0, x_0, \bar{x}_0, p_0, k_0, \mu_0, y_0).
\]

Since the inverse of the structural function is identified, the function itself is also identified on the joint support of \((\bar{Y}^*_i, X_i, \bar{X}_i, P_i, \mu, \theta_i)\).

A.2 Illustration of Exogenous Effects

To illustrate the ambiguity in exogenous effects in Equation 6.2, consider the following simple case. Suppose there are only 2 players, \(i = \{1, 2\}\) and that utility (4.2) takes the following form:

\[
U_i = \gamma_1 Y_i - \frac{\gamma_2}{2} (e_i - \gamma_3 e_{-i})^2.
\]

The first term captures the utility from achievement. The cost of effort, \(c_i(e_i, e_{-i}) = \frac{\gamma_2}{2} (e_i - \gamma_3 e_{-i})^2\), takes the familiar form of conformity effects used in Brock and Durlauf (2001a). The cost of effort is diminishing in the effort of peers, i.e., \(\partial^2 U_i / \partial e_i \partial e_{-i} \geq 0\). Let achievement production (4.1) take the following simple form:

\[
Y_i = \rho_1 e_i + S_i \rho_2 + \mu + \theta_i.
\]

For simplicity of exposition, I assume that there is no direct effect of peer effort. The story is more complicated with direct spillovers from peer effort, but this simplified model is sufficient to illustrate the potential ambiguity in the exogenous effects that are estimated in the model. The ex ante expected value of achievement is then captured by \(\bar{Y}_i = \rho_1 e_i + S_i \rho_2 + \mu\), where \(E(\theta_i | S_i, S_{-i}) = 0\). Each student chooses effort to maximize his ex ante expected utility,

\[
\max_{e_i} \gamma_1 (\rho_1 e_i + S_i \rho_2 + \mu) - \frac{\gamma_2}{2} (e_i - \gamma_3 e_{-i})^2;
\]

57
where $S_1, S_2$ are common knowledge. Utility-maximizing effort for each student is then:

$$e^*_i = \frac{\gamma_1 \rho_1}{\gamma_2} + \gamma_3 e_{-i}.$$  

Effort is increasing in peer effort due to the conformity effect. Furthermore, it is increasing in $\gamma_1$, the marginal utility of achievement, and decreasing in $\gamma_2$, the cost parameter.

To map the game in effort into achievement, I solve for effort as a function of ex ante expected achievement, i.e.,

$$e_i = \frac{1}{\rho_1} (\tilde{Y}_i - S_i \rho_2 - \mu).$$

Plugging effort as a function of ex ante achievement into the effort best-response function, we have:

$$Y^*_i = \rho_1^2 \frac{\gamma_1}{\gamma_2} + \gamma_3 \tilde{Y}_{-i} + S_i \rho_2 - S_{-i} \gamma_3 \rho_2 - \mu (1 - \gamma_3) + \mu + \theta_i.$$  

Thus, observed, equilibrium achievement is a function of the ex ante peer achievement, the state variables of the individual and his peers.

Now, I partition the vector of state variables into a student’s own characteristics and peer characteristics, $S_i \equiv (X_i, X_{-i})$. Similarly, I partition the parameters such that $\rho_2 \equiv (\rho_{21}, \rho_{22})$ corresponding to the marginal effect of individual and peer characteristics. Achievement can then be written explicitly as a function of individual and peer characteristics as follows:

$$Y^*_i = \rho_{21}^2 \frac{\gamma_1}{\gamma_2} + \gamma_3 \tilde{Y}_{-i} + X_i (\rho_{21} - \gamma_3 \rho_{22}) + X_{-i} (\rho_{22} - \gamma_3 \rho_{21}) + \mu (1 - \gamma_3) + \theta_i.$$  

Intuitively, we would expect the marginal effect of a student $i$’s characteristics on his own achievement would be greater than the marginal effect of student $i$’s characteristics on his peer’s achievement, i.e. $\rho_{21} > \gamma_3 \rho_{22}$. However, the marginal effect of peer characteristics, $\rho_{22} - \gamma_3 \rho_{21}$, is ambiguous and depends on the magnitude of $\gamma_3$, the conformity parameter. For large $\gamma_3$, where peer effort has a larger effect on utility-maximizing effort, the marginal effect of $X_{-i}$ is negative. Similarly, if the direct effect of peer characteristics on achievement, $\rho_{22}$, is sufficiently small, the marginal return to peer characteristics is negative. Finally, if the marginal return to a student’s own characteristics, $\rho_{21}$, is sufficiently large, the effect of $X_{-i}$ is negative. Thus, the intuition for exogenous peer effects having a positive effect on achievement follows only when the dominant effect of peer characteristics is coming from the direct effect on peer achievement, and not through the strategic component, peer achievement/effort.
Turning to correlated effects, the marginal effect of $\mu$ in this simple case only depends on whether $\gamma_3 \geq 1$. Allowing for a direct effect of peer effort on achievement, complementarities between effort and $\mu$, or diminishing marginal returns to achievement makes the marginal effect of $\mu$ much more difficult to predict or interpret.

A.3 Figures and Tables

Table 1: Summary Statistics: Class Level Peer Groups (N=876,176)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading (standardized)</td>
<td>0.1441</td>
<td>0.9500</td>
<td>-3.346</td>
<td>3.064</td>
</tr>
<tr>
<td>Avg. peer reading</td>
<td>0.1228</td>
<td>0.4541</td>
<td>-2.385</td>
<td>2.070</td>
</tr>
<tr>
<td>Class size</td>
<td>22.53</td>
<td>3.53</td>
<td>9</td>
<td>29</td>
</tr>
<tr>
<td>Male</td>
<td>0.4990</td>
<td>0.5000</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Black</td>
<td>0.2952</td>
<td>0.4561</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.0343</td>
<td>0.1819</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Parent HS/some post-sec.</td>
<td>0.6421</td>
<td>0.4794</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Parent 4-year degree+</td>
<td>0.2956</td>
<td>0.4563</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Free/reduced price lunch</td>
<td>0.3946</td>
<td>0.4888</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Free reading hours</td>
<td>0.9076</td>
<td>0.6266</td>
<td>0</td>
<td>2.5</td>
</tr>
<tr>
<td>TV hours</td>
<td>2.4643</td>
<td>1.6618</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Avg. white peer reading</td>
<td>0.3162</td>
<td>0.4758</td>
<td>-3.236</td>
<td>2.329</td>
</tr>
<tr>
<td>Avg. nonwhite peer reading</td>
<td>-0.2732</td>
<td>0.5534</td>
<td>-3.026</td>
<td>2.737</td>
</tr>
</tbody>
</table>

Source: Author’s calculations using North Carolina Education Research Data Center, End of Grade exams. Sample restricted to grades 4 and 5 and academic years 1998-99 to 2002-03.
Table 2: Summary Statistics by Race

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>White</th>
<th>Nonwhite</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Reading (standardized)</td>
<td>0.3642 0.9021</td>
<td>-0.3038 0.8853</td>
</tr>
<tr>
<td>Class size</td>
<td>22.84 3.43</td>
<td>21.90 3.64</td>
</tr>
<tr>
<td>Male</td>
<td>0.5044 0.5000</td>
<td>0.4882 0.4999</td>
</tr>
<tr>
<td>Parent HS/some post-sec.</td>
<td>0.5852 0.4927</td>
<td>0.7580 0.4283</td>
</tr>
<tr>
<td>Parent 4-year degree+</td>
<td>0.3664 0.4818</td>
<td>0.1514 0.3585</td>
</tr>
<tr>
<td>Free/reduced price lunch</td>
<td>0.2431 0.4290</td>
<td>0.7029 0.4570</td>
</tr>
<tr>
<td>Free reading hours</td>
<td>0.9436 0.6281</td>
<td>0.8344 0.6170</td>
</tr>
<tr>
<td>TV hours</td>
<td>2.214 1.497</td>
<td>2.974 1.852</td>
</tr>
</tbody>
</table>

Characteristics of Classroom Peer Groups

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>White</th>
<th>Nonwhite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. peer reading</td>
<td>0.2049 0.4286</td>
<td>-0.0444 0.4588</td>
</tr>
<tr>
<td>Avg. white peer reading</td>
<td>0.3490 0.4435</td>
<td>0.2450 0.5322</td>
</tr>
<tr>
<td>Avg. nonwhite peer reading</td>
<td>-0.2395 0.5955</td>
<td>-0.3324 0.4646</td>
</tr>
<tr>
<td>% nonwhite</td>
<td>0.2384 0.2099</td>
<td>0.5244 0.2643</td>
</tr>
<tr>
<td>% parents HS degree</td>
<td>0.6174 0.2169</td>
<td>0.6857 0.1959</td>
</tr>
<tr>
<td>% parents 4-year degree</td>
<td>0.3168 0.2377</td>
<td>0.2465 0.2081</td>
</tr>
<tr>
<td>% FRP lunch</td>
<td>0.3410 0.2084</td>
<td>0.5191 0.2473</td>
</tr>
<tr>
<td>% without peers of other race</td>
<td>0.1468 0.3539</td>
<td>0.0637 0.2442</td>
</tr>
<tr>
<td>N</td>
<td>587,483</td>
<td>288,693</td>
</tr>
</tbody>
</table>

Source: Author’s calculations using North Carolina Education Research Data Center, End of Grade exams. Sample restricted to grades 4 and 5 and academic years 1998/99 to 2002/03.
Figure 1: Segregation within North Carolina Public School Districts

![Figure 1: Segregation within North Carolina Public School Districts](image)

Source: Author’s calculations. Sample restricted to grades 3 to 5.

Figure 2: Black-White Achievement Gap

![Figure 2: Black-White Achievement Gap](image)

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1971 to 2004 Long-Term Trend Reading Assessments. The black-white gap is measured as the average of the reading scale score for white students minus the average for blacks, where the reading trend scale was constructed by the NCES based on the 1984 assessment and included all previous reading assessments.
Figure 3: Reading Achievement in North Carolina Public Schools by Race

Source: Author’s calculations. Sample restricted to grades 3 to 5.
Figure 4: Student Accountability in Charlotte-Mecklenburg Schools

Figure 5: Comparison of Two Districts with Different Concentrations of Low Achievers
Table 3: Comparison of Contemporaneous to Lagged Achievement

<table>
<thead>
<tr>
<th></th>
<th>Naive Twice-Lagged</th>
<th>Contemporaneous 1st Stage</th>
<th>2nd Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. peer reading</td>
<td>-0.0715***</td>
<td>-0.1454***</td>
<td>0.4794***</td>
</tr>
<tr>
<td></td>
<td>[0.0088]</td>
<td>[0.0081]</td>
<td>[0.0511]</td>
</tr>
<tr>
<td>Accountability</td>
<td>0.0033</td>
<td>0.1353***</td>
<td>-0.0200***</td>
</tr>
<tr>
<td></td>
<td>[0.0039]</td>
<td>[0.0200]</td>
<td>[0.0053]</td>
</tr>
<tr>
<td>Low reading</td>
<td>-1.1135***</td>
<td>-1.1472***</td>
<td>-0.0096***</td>
</tr>
<tr>
<td></td>
<td>[0.0024]</td>
<td>[0.0044]</td>
<td>[0.0006]</td>
</tr>
<tr>
<td>Accountable*low reading</td>
<td>0.1081***</td>
<td>0.1414***</td>
<td>0.0043***</td>
</tr>
<tr>
<td></td>
<td>[0.0041]</td>
<td>[0.0055]</td>
<td>[0.0008]</td>
</tr>
<tr>
<td>% low reading</td>
<td>-0.2234***</td>
<td>-0.3098***</td>
<td>-1.1831***</td>
</tr>
<tr>
<td></td>
<td>[0.0143]</td>
<td>[0.0174]</td>
<td>[0.0110]</td>
</tr>
<tr>
<td>Accountable*% low reading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonwhite</td>
<td>-0.2642***</td>
<td>-0.2535***</td>
<td>-0.0039***</td>
</tr>
<tr>
<td></td>
<td>[0.0020]</td>
<td>[0.0028]</td>
<td>[0.0005]</td>
</tr>
<tr>
<td>Male</td>
<td>-0.0632***</td>
<td>-0.0520***</td>
<td>-0.0026***</td>
</tr>
<tr>
<td></td>
<td>[0.0015]</td>
<td>[0.0021]</td>
<td>[0.0006]</td>
</tr>
<tr>
<td>Parent HS degree</td>
<td>0.2139***</td>
<td>0.2159***</td>
<td>-0.0030***</td>
</tr>
<tr>
<td></td>
<td>[0.0033]</td>
<td>[0.0046]</td>
<td>[0.0008]</td>
</tr>
<tr>
<td>Parent 4-year+</td>
<td>0.5672***</td>
<td>0.5591***</td>
<td>-0.0012</td>
</tr>
<tr>
<td></td>
<td>[0.0037]</td>
<td>[0.0052]</td>
<td>[0.0009]</td>
</tr>
<tr>
<td>Free reading hours</td>
<td>0.2266***</td>
<td>0.2512***</td>
<td>0.0071***</td>
</tr>
<tr>
<td></td>
<td>[0.0035]</td>
<td>[0.0051]</td>
<td>[0.0009]</td>
</tr>
<tr>
<td>% nonwhite</td>
<td>-0.0939***</td>
<td>-0.1002***</td>
<td>-0.3520***</td>
</tr>
<tr>
<td></td>
<td>[0.0108]</td>
<td>[0.0163]</td>
<td>[0.0096]</td>
</tr>
<tr>
<td>% male</td>
<td>-0.0559***</td>
<td>-0.0599***</td>
<td>-0.1245***</td>
</tr>
<tr>
<td></td>
<td>[0.0122]</td>
<td>[0.0172]</td>
<td>[0.0113]</td>
</tr>
<tr>
<td>% HS degree</td>
<td>-0.0555***</td>
<td>0.0169</td>
<td>0.2021***</td>
</tr>
<tr>
<td></td>
<td>[0.0170]</td>
<td>[0.0242]</td>
<td>[0.0157]</td>
</tr>
<tr>
<td>% 4-year degree</td>
<td>0.0109</td>
<td>0.1478***</td>
<td>0.6110***</td>
</tr>
<tr>
<td></td>
<td>[0.0185]</td>
<td>[0.0269]</td>
<td>[0.0167]</td>
</tr>
<tr>
<td>Avg. free reading hrs.</td>
<td>0.1479***</td>
<td>0.0858***</td>
<td>0.3493***</td>
</tr>
<tr>
<td></td>
<td>[0.0187]</td>
<td>[0.0277]</td>
<td>[0.0171]</td>
</tr>
<tr>
<td>N</td>
<td>876,176</td>
<td>421,787</td>
<td>876,176</td>
</tr>
<tr>
<td>R²</td>
<td>0.5356</td>
<td>0.5541</td>
<td>0.9042</td>
</tr>
</tbody>
</table>

*significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in brackets, clustered at the peer group level. Sample restricted to 4th and 5th graders, academic years 1998-99 to 2002-03. Teacher and year fixed effects and constant also included.
### Table 4: Heterogeneous Reference Groups

<table>
<thead>
<tr>
<th></th>
<th>2SLS Median</th>
<th>2SLS Nonwhite Median</th>
<th>Median White</th>
<th>Median Nonwhite Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. white peer reading</td>
<td>0.3957***</td>
<td>0.1973***</td>
<td>0.3867***</td>
<td>0.0857</td>
</tr>
<tr>
<td></td>
<td>[0.0890]</td>
<td>[0.1358]</td>
<td>[0.0908]</td>
<td>[0.1438]</td>
</tr>
<tr>
<td>Avg. nonwhite peer reading</td>
<td>0.0157</td>
<td>0.1585</td>
<td>0.0182</td>
<td>0.2729**</td>
</tr>
<tr>
<td></td>
<td>[0.0446]</td>
<td>[0.1558]</td>
<td>[0.0533]</td>
<td>[0.1211]</td>
</tr>
<tr>
<td>Accountability</td>
<td>-0.0026</td>
<td>0.0049</td>
<td>-0.0150***</td>
<td>-0.0184**</td>
</tr>
<tr>
<td></td>
<td>[0.0036]</td>
<td>[0.0085]</td>
<td>[0.0042]</td>
<td>[0.0088]</td>
</tr>
<tr>
<td>Low reading</td>
<td>-1.1832***</td>
<td>-0.9991***</td>
<td>-1.1525***</td>
<td>-0.9788***</td>
</tr>
<tr>
<td></td>
<td>[0.0037]</td>
<td>[0.0036]</td>
<td>[0.0045]</td>
<td>[0.0044]</td>
</tr>
<tr>
<td>Accountable*Low reading</td>
<td>0.1251***</td>
<td>0.0606***</td>
<td>0.1542***</td>
<td>0.1076***</td>
</tr>
<tr>
<td></td>
<td>[0.0056]</td>
<td>[0.0062]</td>
<td>[0.0067]</td>
<td>[0.0068]</td>
</tr>
<tr>
<td>% white low reading</td>
<td>0.3673***</td>
<td>0.1822</td>
<td>0.3510***</td>
<td>0.0474</td>
</tr>
<tr>
<td></td>
<td>[0.1047]</td>
<td>[0.1549]</td>
<td>[0.1076]</td>
<td>[0.1617]</td>
</tr>
<tr>
<td>% NW low reading</td>
<td>-0.0037</td>
<td>0.1032</td>
<td>-0.0037</td>
<td>0.2205*</td>
</tr>
<tr>
<td></td>
<td>[0.0437]</td>
<td>[0.1501]</td>
<td>[0.0523]</td>
<td>[0.1136]</td>
</tr>
<tr>
<td>% nonwhite</td>
<td>-0.0646***</td>
<td>-0.1050***</td>
<td>-0.0851***</td>
<td>-0.0769***</td>
</tr>
<tr>
<td></td>
<td>[0.0103]</td>
<td>[0.0209]</td>
<td>[0.0120]</td>
<td>[0.0190]</td>
</tr>
<tr>
<td>Male</td>
<td>-0.0495***</td>
<td>-0.0881***</td>
<td>-0.0429***</td>
<td>-0.0860***</td>
</tr>
<tr>
<td></td>
<td>[0.0016]</td>
<td>[0.0023]</td>
<td>[0.0021]</td>
<td>[0.0029]</td>
</tr>
<tr>
<td>Parent HS degree</td>
<td>0.2636***</td>
<td>0.1531***</td>
<td>0.2633***</td>
<td>0.1497***</td>
</tr>
<tr>
<td></td>
<td>[0.0043]</td>
<td>[0.0045]</td>
<td>[0.0054]</td>
<td>[0.0059]</td>
</tr>
<tr>
<td>Parent 4-year +</td>
<td>0.6087***</td>
<td>0.4174***</td>
<td>0.6218***</td>
<td>0.4107***</td>
</tr>
<tr>
<td></td>
<td>[0.0045]</td>
<td>[0.0068]</td>
<td>[0.0057]</td>
<td>[0.0079]</td>
</tr>
<tr>
<td>Free reading hours</td>
<td>0.2745***</td>
<td>0.0739***</td>
<td>0.2973***</td>
<td>0.0592***</td>
</tr>
<tr>
<td></td>
<td>[0.0042]</td>
<td>[0.0051]</td>
<td>[0.0049]</td>
<td>[0.0062]</td>
</tr>
<tr>
<td>% male</td>
<td>-0.0039</td>
<td>0.0026</td>
<td>-0.0141</td>
<td>0.0092</td>
</tr>
<tr>
<td></td>
<td>[0.0141]</td>
<td>[0.0289]</td>
<td>[0.0179]</td>
<td>[0.0333]</td>
</tr>
<tr>
<td>% HS degree</td>
<td>-0.1392***</td>
<td>-0.1209***</td>
<td>-0.1488***</td>
<td>-0.1705***</td>
</tr>
<tr>
<td></td>
<td>[0.0237]</td>
<td>[0.0424]</td>
<td>[0.0231]</td>
<td>[0.0290]</td>
</tr>
<tr>
<td>% 4-year degree</td>
<td>-0.2654***</td>
<td>-0.2023***</td>
<td>-0.2727***</td>
<td>-0.2185***</td>
</tr>
<tr>
<td></td>
<td>[0.0572]</td>
<td>[0.0901]</td>
<td>[0.0563]</td>
<td>[0.0745]</td>
</tr>
<tr>
<td>Avg. free reading hrs.</td>
<td>-0.0232</td>
<td>0.0390</td>
<td>-0.0263</td>
<td>0.0495</td>
</tr>
<tr>
<td></td>
<td>[0.0424]</td>
<td>[0.0599]</td>
<td>[0.0427]</td>
<td>[0.0574]</td>
</tr>
<tr>
<td>No peers of other race</td>
<td>0.0008</td>
<td>0.0576</td>
<td>0.0009</td>
<td>0.0214</td>
</tr>
<tr>
<td></td>
<td>[0.0092]</td>
<td>[0.0487]</td>
<td>[0.0111]</td>
<td>[0.0445]</td>
</tr>
<tr>
<td>N</td>
<td>587,484</td>
<td>288,693</td>
<td>587,483</td>
<td>288,693</td>
</tr>
<tr>
<td>R²</td>
<td>0.4895</td>
<td>0.4941</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in brackets, clustered at the peer group level. Standard errors calculated using bootstrap with sample size of 100. Year and teacher fixed effects and constant also included.
Figure 6: Effect of Average Peer Achievement: 2SQR

![Graph showing the effect of average peer achievement on achievement quantile for Whites and Nonwhites.]

Figure 7: Exogenous Peer Group Composition Effects: % Nonwhite

![Graph showing the effect of exogenous peer group composition on achievement quantile for Whites and Nonwhites.]

66
Table 5: Average Marginal Effects of Peers for Whites

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>.1 Quantile</th>
<th>Median</th>
<th>.9 Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. white reading</td>
<td>0.1841***</td>
<td>0.3220***</td>
<td>0.1715***</td>
<td>0.0971</td>
</tr>
<tr>
<td>Avg. nonwhite reading</td>
<td>0.0032</td>
<td>-0.0216</td>
<td>0.0101</td>
<td>-0.0056</td>
</tr>
<tr>
<td>% white low performing</td>
<td>0.0586***</td>
<td>0.1156***</td>
<td>0.0526***</td>
<td>0.0238</td>
</tr>
<tr>
<td>% NW low performing</td>
<td>-0.0044</td>
<td>-0.0188</td>
<td>-0.0012</td>
<td>-0.0070</td>
</tr>
<tr>
<td>% nonwhite</td>
<td>-0.0179***</td>
<td>-0.0127***</td>
<td>-0.0179***</td>
<td>-0.0215***</td>
</tr>
<tr>
<td>% male</td>
<td>-0.0002</td>
<td>0.0020</td>
<td>-0.0012</td>
<td>-0.0015</td>
</tr>
<tr>
<td>% parents with HS degree</td>
<td>-0.0310***</td>
<td>-0.0343***</td>
<td>-0.0323***</td>
<td>-0.0311***</td>
</tr>
<tr>
<td>% parents with 4-year degree</td>
<td>-0.0660***</td>
<td>-0.0973***</td>
<td>-0.0648***</td>
<td>-0.0468**</td>
</tr>
<tr>
<td>Avg. free reading</td>
<td>-0.0022</td>
<td>-0.0092**</td>
<td>-0.0019</td>
<td>0.0031</td>
</tr>
</tbody>
</table>

*significant at 10%; ** significant at 5%; *** significant at 1%. The marginal effects are for a one standard deviation increase in the peer variable using the 2SQR regression broken out by subgroup, i.e., that depicted in Figures 6 and 7. Marginal effects are averaged over quantiles for the first column.
Table 6: Average Marginal Effects of Peers for Nonwhites

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>.1 Quantile</th>
<th>Median</th>
<th>.9 Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. white reading</td>
<td>0.1095</td>
<td>0.1589*</td>
<td>0.0444</td>
<td>0.1413</td>
</tr>
<tr>
<td></td>
<td>[0.0794]</td>
<td>[0.0945]</td>
<td>[0.0746]</td>
<td>[0.0929]</td>
</tr>
<tr>
<td>Avg. nonwhite reading</td>
<td>0.1056*</td>
<td>0.0777</td>
<td>0.1265**</td>
<td>0.1285**</td>
</tr>
<tr>
<td></td>
<td>[0.0598]</td>
<td>[0.0820]</td>
<td>[0.0561]</td>
<td>[0.0649]</td>
</tr>
<tr>
<td>% white low performing</td>
<td>0.0426</td>
<td>0.0687</td>
<td>0.0103</td>
<td>0.0564</td>
</tr>
<tr>
<td></td>
<td>[0.0373]</td>
<td>[0.0448]</td>
<td>[0.0350]</td>
<td>[0.0435]</td>
</tr>
<tr>
<td>% NW low performing</td>
<td>0.0420</td>
<td>0.0345</td>
<td>0.0508*</td>
<td>0.0504*</td>
</tr>
<tr>
<td></td>
<td>[0.0279]</td>
<td>[0.0385]</td>
<td>[0.0262]</td>
<td>[0.0303]</td>
</tr>
<tr>
<td>% nonwhite</td>
<td>-0.0194***</td>
<td>-0.0331***</td>
<td>-0.0203***</td>
<td>-0.0084</td>
</tr>
<tr>
<td></td>
<td>[0.0055]</td>
<td>[0.0069]</td>
<td>[0.0050]</td>
<td>[0.0056]</td>
</tr>
<tr>
<td>% male</td>
<td>0.0014</td>
<td>0.0011</td>
<td>0.0009</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>[0.0033]</td>
<td>[0.0043]</td>
<td>[0.0031]</td>
<td>[0.0034]</td>
</tr>
<tr>
<td>% parents with HS degree</td>
<td>-0.0352***</td>
<td>-0.0433***</td>
<td>-0.0334***</td>
<td>-0.0299***</td>
</tr>
<tr>
<td></td>
<td>[0.0057]</td>
<td>[0.0076]</td>
<td>[0.0057]</td>
<td>[0.0056]</td>
</tr>
<tr>
<td>% parents with 4-year degree</td>
<td>-0.0583***</td>
<td>-0.0786***</td>
<td>-0.0455**</td>
<td>-0.0584***</td>
</tr>
<tr>
<td></td>
<td>[0.0169]</td>
<td>[0.0215]</td>
<td>[0.0155]</td>
<td>[0.0186]</td>
</tr>
<tr>
<td>Avg. free reading</td>
<td>0.0012</td>
<td>-0.0014</td>
<td>0.0036</td>
<td>-0.0045</td>
</tr>
<tr>
<td></td>
<td>[0.0045]</td>
<td>[0.0054]</td>
<td>[0.0042]</td>
<td>[0.0049]</td>
</tr>
</tbody>
</table>

*significant at 10%; ** significant at 5%; *** significant at 1%. The marginal effects are for a one standard deviation increase in the peer variable using the 2SQR regression broken out by subgroup, i.e., that depicted in Figures 6 and 7. Marginal effects are averaged over quantiles for the first column.

Table 7: Alternative Assignment Policies (5th graders 2002-03, N=89,389)

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th>Nonwhite</th>
<th>Gap</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2SQR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0.7003</td>
<td>0.0377</td>
<td>0.6626</td>
<td>0.4710</td>
</tr>
<tr>
<td>Random</td>
<td>0.6962</td>
<td>0.0939</td>
<td>0.6023</td>
<td>0.4877</td>
</tr>
<tr>
<td>Tracking</td>
<td>0.6334</td>
<td>-0.0057</td>
<td>0.6392</td>
<td>0.4122</td>
</tr>
<tr>
<td><strong>2SLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0.7167</td>
<td>-0.0940</td>
<td>0.8106</td>
<td>0.4361</td>
</tr>
<tr>
<td>Random</td>
<td>0.7321</td>
<td>-0.0535</td>
<td>0.7856</td>
<td>0.4602</td>
</tr>
<tr>
<td>Tracking</td>
<td>0.7326</td>
<td>-0.0871</td>
<td>0.8197</td>
<td>0.4489</td>
</tr>
</tbody>
</table>
Figure 8: Achievement Gains from State-Wide Desegregation and Tracking

Figure 9: Comparison of Predictions using 2SQR and 2SLS

Figure 10: Changes in Achievement Gap from Alternative Assignment Policies
Table 8: Summary Statistics for Two Public School Districts  
(Grades 4 and 5, Academic Year 2002-03)

<table>
<thead>
<tr>
<th></th>
<th>Durham</th>
<th>Chapel Hill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. nonwhite reading</td>
<td>-0.0190</td>
<td>0.0848</td>
</tr>
<tr>
<td>Avg. white reading</td>
<td>0.8389</td>
<td>1.2361</td>
</tr>
<tr>
<td>Segregation index</td>
<td>0.2806</td>
<td>0.0358</td>
</tr>
<tr>
<td>% Free/reduced price lunch</td>
<td>0.4310</td>
<td>0.1592</td>
</tr>
<tr>
<td>% Nonwhite</td>
<td>0.6789</td>
<td>0.2285</td>
</tr>
<tr>
<td>% Parents with HS degree</td>
<td>0.6140</td>
<td>0.1936</td>
</tr>
<tr>
<td>% Parents with 4-year degree</td>
<td>0.3446</td>
<td>0.7911</td>
</tr>
<tr>
<td>N</td>
<td>4,780</td>
<td>1,501</td>
</tr>
</tbody>
</table>

Table 9: Merger of Durham and Chapel Hill Districts (5th graders 2002-03, N=2,766)

<table>
<thead>
<tr>
<th></th>
<th>Nonwhite</th>
<th>White</th>
<th>Gap</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durham Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0.0623</td>
<td>0.9358</td>
<td>0.8736</td>
<td>0.3390</td>
</tr>
<tr>
<td>Desegregation within district</td>
<td>0.1134</td>
<td>0.9596</td>
<td>0.8462</td>
<td>0.3815</td>
</tr>
<tr>
<td>Desegregation across district</td>
<td>0.1618</td>
<td>1.0091</td>
<td>0.8473</td>
<td>0.4302</td>
</tr>
<tr>
<td>Chapel Hill Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0.1909</td>
<td>1.2968</td>
<td>1.1059</td>
<td>1.0597</td>
</tr>
<tr>
<td>Desegregation within district</td>
<td>0.2443</td>
<td>1.3231</td>
<td>1.0788</td>
<td>1.0918</td>
</tr>
<tr>
<td>Desegregation across district</td>
<td>0.2229</td>
<td>1.2888</td>
<td>1.0659</td>
<td>1.0603</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0.0739</td>
<td>1.0949</td>
<td>0.9296</td>
<td>0.5128</td>
</tr>
<tr>
<td>Desegregation within district</td>
<td>0.1252</td>
<td>1.1198</td>
<td>0.9023</td>
<td>0.5528</td>
</tr>
<tr>
<td>Desegregation across district</td>
<td>0.1673</td>
<td>1.1323</td>
<td>0.9000</td>
<td>0.5821</td>
</tr>
</tbody>
</table>
Figure 11: Merger of Durham and Chapel Hill

Table 10: Durham to Chapel Hill

<table>
<thead>
<tr>
<th></th>
<th>Nonwhite</th>
<th>White</th>
<th>Gap</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Durham 5th graders 2002-03 (N=2,099)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0.0623</td>
<td>0.9358</td>
<td>0.8736</td>
<td>0.3390</td>
</tr>
<tr>
<td>Counterfactual</td>
<td>0.1346</td>
<td>1.0210</td>
<td>0.8864</td>
<td>0.4154</td>
</tr>
<tr>
<td>Change</td>
<td>0.0723</td>
<td>0.0852</td>
<td>0.0129</td>
<td>0.0764</td>
</tr>
<tr>
<td><strong>Low Equilibrium Durham Schools</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>-0.1155</td>
<td>0.3193</td>
<td>0.4348</td>
<td>-0.0950</td>
</tr>
<tr>
<td>Counterfactual</td>
<td>0.0314</td>
<td>0.6518</td>
<td>0.6203</td>
<td>0.0607</td>
</tr>
<tr>
<td>Change</td>
<td>0.1470</td>
<td>0.3325</td>
<td>0.1855</td>
<td>0.1557</td>
</tr>
<tr>
<td><strong>Chapel 5th graders 2002-03 (N=667)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0.1909</td>
<td>1.2968</td>
<td>1.1059</td>
<td>1.0597</td>
</tr>
<tr>
<td>Counterfactual</td>
<td>0.1224</td>
<td>1.1277</td>
<td>1.0053</td>
<td>0.9122</td>
</tr>
<tr>
<td>Change</td>
<td>-0.0685</td>
<td>-0.1691</td>
<td>-0.1006</td>
<td>-0.1475</td>
</tr>
</tbody>
</table>