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

Jane Cooley[†]

March 6, 2007

Abstract

Understanding peer effects is critical to evaluating the effect of persistent 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, allows peer spillovers to vary across races and allows students to form different race-based reference groups within the classroom. I find evidence of stronger peer influences within reference groups than across reference groups and the magnitude of these peer influences varies substantially across percentiles of the achievement distribution. I apply my results to evaluate the distributional effects of alternative classroom assignment policies. Desegregating peer groups leads to small improvements in the achievement gap across the achievement distribution.

Keywords: racial achievement gap, peer effects, desegregation

JEL: I20, I21, J15

[*]I would like to thank Pat Bajari, Han Hong, Tom Nechyba, Peter Arcidiacono, Arie Beresteanu, Charlie Clotfelter, Steven Durlauf, Stephen Ryan, Karl Scholz, Chris Timmins, Jake Vigdor and seminar participants at Duke, Brown, University of Chicago, Clemson, the Ford School, University of Illinois-Urbana Champaign, Princeton, Northwestern, Notre Dame, Penn State, Virginia Tech, University of Wisconsin-Madison, and Yale 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, University of Wisconsin-Madison, Madison, WI 53706; jcooley@ssc.wisc.edu.

1 Introduction

Despite decades of reform that helped to equalize educational opportunities across races, a sizable racial achievement gap persists. For instance, 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. Though the determinants of this striking gap remain a puzzle, the persistent segregation of public schools is frequently considered to be a contributing factor. Despite the mixed evidence in the literature,[1] policy makers often advocate desegregation as a means of narrowing the gap.[2] This could occur through several channels, such as the redistribution of resources or the creation of "better" peer groups. Despite notable differences in predominantly white and predominantly nonwhite public schools, the evidence suggests that resource disparities explain little, if any, of the achievement gap.[3] Thus, I focus on the latter channel, isolating the effect of racially diverse classroom peer groups on the achievement of white and nonwhite students.

Creating racially diverse peer groups could help narrow the gap if lower-achieving nonwhite students benefit from being grouped with higher-achieving white peers. However, benefits could be muted if students form different race-based reference groups within the classroom.[4] Therefore, analyzing the effects of racial diversity requires a deeper understanding of peer effects in the classroom and, in particular, how these peer effects may vary by race and across the percentiles of the achievement distribution. The first challenge, which Manski (1993) describes as the *reflection problem*, lies in separating the effect of peer characteristics from the effect of peer actions and the overall effect of peers from unobserved group effects. The second challenge lies in developing an identification and estimation strategy that is sufficiently flexible to capture heterogeneity in peer spillovers by race and across percentiles of the achievement distribution. I address these challenges in the context of a unique panel data set of standardized achievement scores for North Carolina public elementary school

[1] For instance, see Card and Rothstein (2005), Cook and Evans (2000), Guryan (2004), Hanushek et al. (2004), Hoxby and Weingarth (2005) and Rivkin (2000).

[2] 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. (Report of the North Carolina Division of School Improvement.)

[3] For instance, see Jencks and Phillips (1998) for an overview.

[4] For instance, see Simmons (1999) and Fryer and Torelli (2005) for existing anecdotal and empirical evidence.

students in which peer groups are identified at the classroom level.

I begin by writing down a model of student achievement. The first order conditions from this model show that a student's own achievement is a function of the achievement of his peers and exogenous characteristics of the student, his peers and the classroom. This effectively generates what has traditionally been referred to as an "achievement production function" with peer spillovers and reinterprets it as being an achievement best response function.[5] Placed in the context of optimizing behavior, identification then relies on finding an exogenous shifter that affects peer achievement through students' choices of "effort" that are independent of a student's own achievement.

I argue that the introduction of *student accountability* policies, which require that students perform above a certain level in order to be automatically promoted to the next grade, provide such an exogenous shift. Student accountability acts as a utility shifter by imposing an additional cost on low performance, thus shifting the effort of those who perceive themselves to be in danger of failing under the new policy. Intuitively, peer spillovers are then identified by contrasting classrooms with varying compositions of low-performers in grades with and without student accountability policies. Put differently, I compare classrooms with different portions of students that are *held accountable* to classrooms of similar composition where students are not held accountable. After controlling for arguably the most important source of unobserved group effects, teacher quality, through teacher fixed effects, the central identifying assumption is then that the percentage of students held accountable is independent of the remaining unobserved group effect. I take considerable care to describe the conditions under which this assumption holds.

Extending the identification strategy to allow peer effects to vary by race is relatively straightforward since the percentage of peers of a given race that are held accountable varies across classrooms. However, a remaining concern is heterogeneity in responses to peers based on individual characteristics that are unobserved to the econometrician, such as ability. To illustrate, suppose low-achieving students respond relatively more to increases in peer achievement than high-achieving students. Abstracting from race-specific peer effects for the moment, this would suggest that a linear-in-means model would understate the effect of desegregating peer groups since nonwhite students are more heavily concentrated in the lower tails of the achievement distribution. In fact, by assuming a common treatment effect,

[5]See Hanushek et al. (2003) for a detailed discussion of educational achievement production functions with peer effects.

the linear-in-means model predicts that no matter how students are assigned to peer groups average achievement remains the same. Thus, capturing heterogeneity in responses to peers across the percentiles of the achievement distribution provides a much richer picture of the equity and efficiency tradeoffs associated with alternative grouping strategy. [6]

The complication introduced by allowing peer spillovers to vary by the percentiles of the achievement distribution is nontrivial since I need to allow the best response function to be nonseparable in the residual. Nonseparability presents challenges at each stage, particularly in specifying an informational structure for the model that is both reasonable and yields a tractable estimator.[7] I make use of the techniques developed in the recent literature on endogenous quantile instrumental variable models.[8] In particular, to capture the distributional effect of peers I make assumptions so that Imbens and Newey (2003)'s control function approach to estimating quantile treatment effects can be applied to a simultaneous-move game of incomplete information. The quantile approach also provides new insight on academic tracking, isolating the distributional effect of peers from resource effects that are often confounded in tracking studies.[9]

A central finding of this paper is that peer spillovers are stronger within race than across races and vary considerably across the percentiles of the achievement distribution. I find diminishing marginal returns to white peer achievement for whites and weakly increasing marginal returns to nonwhite peer achievement for nonwhites. Endogenous effects appear to be much larger than exogenous effects. This is not surprising given that endogenous effects include 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 effect of desegregating peer group on the achievement of different student types, particularly developing implications for the

[6]The importance of functional form assumptions for determining optimal classroom assignment in the presence of peer effects is also highlighted in a theoretical study by Arnott and Rowse (1987).

[7]On the other side, it is worth pointing out that only a very narrow set of assumptions yield a linear-in-means best response function that is additively separable in the residual.

[8]See Abadie et al. (2002), Chernozhukov and Hansen (2005), Chesher (2003), Honoré and Hu (2004), Imbens and Newey (2003), and Ma and Koenker (2004).

[9]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 (2002) find that after correcting for selection into schools, there is no evidence that low-ability students are harmed by being grouped together.

achievement gap. First, I consider implications of merging a predominantly white high-achieving school district (Chapel Hill) with a neighboring lower-achieving racially diverse district (Durham). Desegregating peer groups across the two districts leads to small improvements in the achievement gap across the percentiles of the achievement distribution, with slightly larger gains at the middle percentiles. Then, 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.

2 Data

I use administrative data for North Carolina public school students from the academic years 1994-95 to 2002-03.[10] Beginning in Spring 1995, students in grades 3 to 8 were administered standardized exams in reading and math. I focus on reading test scores.[11] 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:

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

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*. 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

[10] I would like to thank the North Carolina Education Research Data Center (http://www.pubpol.duke.edu/centers/child/nceddatacenter.html) for providing the data. 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.

[11] As discussed in Section 4.3, I have data on free reading time, which 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.

takes place primarily within self-contained classrooms. Beginning in middle school, it is frequently the case that students have different teachers (and potentially different peers) for each subject. I focus on elementary students in grades 3 through 5, where the teacher ID can reliably identify the classroom peer group.[12] 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,[13] and whether the student receives a free or reduced-price lunch,[14] along with some information about student leisure time allocation, such as time spent reading for fun and hours spent watching television.[15] 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.

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 than nonwhites, .36 compared to -.30. They also have better-educated parents

[12] I drop the bottom and top percentile of class sizes. Because classes are not to exceed 25 students, the top percentile of class sizes (30 or more) are likely to have additional teachers. The bottom percentile (smaller than 9) may be small groups of students taking makeup exams or special education classes. However, the results are robust to including the top and bottom percentile.

[13] Data on parental education are collected differently across schools. In some cases, particularly in elementary school, the teacher provides a best guess of the parental education of the 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).

[14] 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.

[15] Students are asked how many hours of television they watch in a given day and how much time they spend reading for fun, which can serve as rough indicators of leisure choices. I take the midpoint of the range of hours indicated as the response to multiple choice questions, and use this to approximate the hours spent doing each of these activities.

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, another potentially 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. Furthermore, 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.

3 Model

The literature on peer spillovers in education production posits many different sources of peer influence but ultimately focuses on a reduced-form setting in which a student's achievement is assumed to be a function of peer achievement and peer characteristics along with the typical individual, teacher and resource inputs. In what follows, I describe a particular model of peer spillovers that encompasses many of these channels and use it to motivate a new approach to identification. I recast students as optimizing agents whose decisions are influenced by their peers and these decisions, in turn, determine achievement in a given peer group. I describe informational assumptions such that the first order conditions from this model yield a reduced-form achievement best response function that is a more general form of the familiar achievement production with peer spillovers that has traditionally been estimated in the literature.

Section 3.1 describes the education technology and the preferences of students. Students are assumed to play a simultaneous-move game of symmetric information. Section 3.2 defines and describes the equilibrium to the game and the resulting equilibrium achievement production function, i.e., the achievement realized under utility-maximizing effort.

3.1 Primitives

Define a peer group to be a classroom of students in a particular time period.[16] Let $i = 1, ..., N$ index students in a given peer group. Achievement $Y_i \in \mathbb{R}$ is the standardized reading test score. The achievement production function is

$$Y_i = g(e_i, e_{-i}; S_i, \theta_i). \tag{3.1}$$

The choice variable of a student i is *effort*, which is chosen on the compact set $e_i \in [\underline{e}, \overline{e}]$. 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 his effort and the effort of his peers, $e_{-i} = (e_1, ..., e_{i-1}, e_{i+1}, ...e_N)$. 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 allows two types of direct peer spillovers. First, as mentioned above, peers may affect directly 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 or prior achievement, often thought to proxy for ability or other unobservables. 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.[17]

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

[16] Because the focus is on interactions within a particular peer group, I suppress time and classroom subscripts for the moment.

[17] Lazear (2001) presents such a model where the classroom learning environment is treated as a congestible public good.

student's performance on a given test day. These types of unobservables are captured by θ_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

$$U_i = u(Y_i, c_i(e_i, e_{-i}); S_i). \tag{3.2}$$

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 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.[18]

The utility function (like the production function) permits 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, e_{-i}. 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.[19] This type of peer spillover has received a great deal of attention in the social interactions literature.[20]

The vector of state variables $S = (S_1, ..., S_N)$ is common knowledge to all students in

[18]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.

[19]See Bishop et al. (2003) for a discussion of these types of peer spillovers, particularly in the high school setting.

[20]See Brock and Durlauf (2001*b*), Graham (2004), and Sweeting (2004). 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 classroom.

the classroom, while (θ_i, θ_{-i}) are observed ex post.[21] Students possess a common prior on θ, $f(\theta_i|S)$.[22] Suppose θ_i is defined on the set Θ. Then the expected utility for a given level of effort, (e_i, e_{-i}), is denoted as follows:

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

3.2 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:

$$e_i^*(e_{-i}; S) \in \text{argmax}_{e_i} \tilde{U}_i(e_i, e_{-i}; S). \tag{3.3}$$

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

Definition 3.1 (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 (3.3).*

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 in peer effort and state variables. 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 3.2 (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 $i = 1...n$.*

Definition 3.3 (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 $h(a^{'}, t^{'}) \geq h(a^{''}, t^{'})$.*

[21] Alternatively, (θ_i, θ_{-i}) may in fact not be observed, but only inferred by the student after a level of achievement is realized.

[22] An alternative model is one of private information. I can show that a simple model of private information has similar implications to the symmetric information model described above.

Assumption 3.1 (Single Crossing of Ex Ante Utility)**.** *Let S_i be defined on the set $\mathbb{S} \subset \mathbb{R}^d$. $\tilde{U}(e_i, e_{-i}; S_i)$ satisfies single crossing in (i.) $(e_i; e_{-i})$ on $[\underline{e}, \overline{e}]^N$, and (ii.) (e_i, S_i) on $[\underline{e}, \overline{e}] \times \mathbb{S}$ for every i.*[23]

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 3.1, it follows from Milgrom and Shannon (1994, Theorem 4) that an equilibrium to the game exists, which is restated as:

Theorem 3.1 (Existence of Pure-Strategy Nash Equilibrium)**.** *Given that Assumption 3.1 holds, a pure strategy Nash equilibrium, as defined in 3.1, 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 3.2 (Monotonicity in effort)**.** *Achievement production is strictly monotonic in effort, i.e.,*

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

[23] 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 θ_i is log-supermodular (i.e., types are affiliated) and that ex post utility satisfies increasing differences.

Definition 3.4 (Achievement Equilibrium)**.** *Given Assumption 3.2, 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 (3.3).*

Theorem 3.2 (Existence of Achievement Equilibrium)**.** *Under Assumption 3.2, the game in effort maps into a game in achievement, which can be represented as*

$$Y_i^* = q(\tilde{Y}_{-i}^*, S_i, S_{-i}, \theta_i), \tag{3.4}$$

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

See Appendix A.1 for proof. Equation 3.4 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 (θ_i).

4 Identification

A growing body of research treats the difficulties associated with identifying peer effects.[24] The pioneering work of Manski (1993) presents two negative identification results for the linear-in-means model. First, the effect of peer characteristics (exogenous peer effects) cannot be separated from the effect of peer behaviors (endogenous peer effects). Second, social effects (exogenous and endogenous) cannot be separated from unobserved group effects, or correlated 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.[25] In contrast, the model presented in Section 3 illustrates a potentially important role for contemporaneous spillovers in achievement production.

If the social interactions take a nonlinear form as posited in this paper, Manski (1993) first identification problem is alleviated, as illustrated by Brock and Durlauf (2001*a*) in the

[24] See Brock and Durlauf (2001*b*) for an overview.

[25] 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.

discrete choice setting. However, while the discrete choice setting is nonlinear in nature, assuming nonlinearities to obtain identification in the continuous choice setting considered here is less palatable. Furthermore, even when the social effect is nonlinear, the second problem of distinguishing unobserved group effects from social effects remains. Intuitively, the problem is that similarity in outcomes could be a result of coordination on the part of students or unobserved classroom inputs, such as teacher quality, which simultaneously improve everyone's outcomes.[26]

The reflection problem can be recast as a simultaneity problem. Because individual and peer achievement are simultaneously determined in equilibrium, in the absence of an exclusion restriction that shifts the achievement of the individual independently of his peers, I cannot identify a causal effect of peers. In Section 4.1, I describe conditions for the existence of a valid instrument in the context of the equilibrium model of achievement. Assumptions are presented such that the peer effect is identified in a nonlinear setting, though these are also sufficient to guarantee identification in the linear-in-means framework. Furthermore, while the instrument helps alleviate endogeneity of peer achievement due to unobserved group effects, I also include a direct control for teacher quality, one of the most salient unobserved group effects in achievement production, by including teacher fixed effects.

Section 4.2 provides some background information and presents evidence in support of the particular exclusion restriction used in this paper. Section 4.3 expands the argument to consider identification when there exists a strategy-relevant individual characteristic that is observed to the student and his peers, but not the econometrician (i.e., unobserved ability). Finally, in Section 4.4, I discuss the implications of non-random assignment.

4.1 Simultaneity in Achievement

To discuss identification, it is useful to distinguish between the different components of the state space. Let $(S_i, S_{-i}) \equiv (X_i, X_{-i}, P_i, K, \mu)$, where X_i captures characteristics of i such as race, sex, parental education, and ability, while X_{-i} captures the characteristics of i's peers (i.e., *exogenous effects*). P_i captures an education policy that affects i's utility from

[26] Graham (2004) presents an innovative solution to this problem of unobservables, exploiting excess variation in random assignment to large and small classrooms to identify the presence of social effects in a linear-in-means setting, relying on an experimental setting where unobserved group effects are plausibly random.

achievement. It is discussed in Assumption 4.1 below. Besides the composition of the peer group, classrooms are differentiated by characteristics (K, μ), which may capture classroom resources, teacher quality, or overall classroom productivity. I assume that while K and μ are observed by the students and therefore taken into account when choosing effort, μ is unobserved to the econometrician (i.e., a *correlated effect*). Intuitively, μ 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, μ could describe the degree to which students get along with the teacher or something more random like a flu outbreak that negatively affects the achievement of all students in a classroom.

Let $\bar{Y}^*_{-i}$ capture expected average peer achievement and similarly, $\bar{X}_{-i}$, average peer characteristics. The achievement best-response function is simplified to depend on the expected average peer achievement and average peer characteristics, rather than the entire vector as permitted in (3.4),[27] i.e.,

$$Y_i^* = q(\bar{Y}^*_{-i}, X_i, \bar{X}_{-i}, P_i, K, \mu, \theta_i). \tag{4.1}$$

As mentioned previously, an important aspect of this production function is that it is permitted to be nonseparable in the error, which provides a much richer picture of the distributional achievement tradeoffs that can be obscured in the linear context. I assume that $q(\cdot)$ is strictly increasing in θ_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 θ_i is independently and identically distributed $\mathcal{U}(0,1)$. Since θ_i is inherently without units, assuming a uniform distribution simply pins down θ. In contrast, the additive model normalizes θ_i to have the same units as Y_i. By fixing $\theta_i = \tau$, Equation 4.1 describes the dependence of the τ^{th} quantiles of the achievement distribution on average expected peer achievement and covariates. Identification of $q(\cdot)$ is defined as follows:

Definition 4.1 (Identification)**.** *The structural function $q(\cdot)$ is* identified *on the joint support of $(Y_i^*, \bar{Y}^*_{-i}, X_i, \bar{X}_{-i}, P_i, K)$, if there exists a unique $q(\cdot)$ that rationalizes $F(Y_i^*, \bar{Y}^*_{-i} | X_i, \bar{X}_{-i}, P_i, K)$, the observed joint distribution of achievement and peer achievement conditional on exogenous state variables.*

[27]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.

The central problem in identifying the structural function (4.1) is that Y_i^* and $\bar{Y}_{-i}^*$ are simultaneously determined in equilibrium. Furthermore, they are a function of the the state variable μ, which is unobserved by the econometrician. The solution that follows relies on the existence of a valid exclusion restriction, which shifts the optimal behavior of peers independently of μ. The following assumptions describe the properties of a valid exclusion restriction and the other assumptions needed for identification of peer spillovers in the context of the equilibrium model of classroom achievement production described in Section 3.

Assumption 4.1 (Utility Shifter). *There exists a variable P_i that affects i's utility from effort (3.2), but does not directly affect achievement production (3.1).*

Assumption 4.2 (Independence). *Conditional on $(X_i, \bar{X}_{-i}, K, \mu)$, θ_i is independent of $(P_i,\ \bar{P}_{-i})$.*

Assumption 4.3 (Exclusion Restriction). *Conditional on $(X_i, \bar{X}_{-i}, K)$, μ, θ_i are jointly independent of $\bar{P}_{-i}$.*

Assumption 4.4 (Monotonicity of Peer Achievement). *With probability one, $h(X_i, \bar{X}_{-i}, P_i, \bar{P}_{-i}, K, \mu)$ is strictly monotonic in μ.*

Assumption 4.5 (Independence of Residuals). *Conditional on $(X_i, \bar{X}_{-i}, K)$, θ_i is independent of μ.*

Assumptions 4.1 and 4.2 ensure that there is no direct effect of P_j on the equilibrium achievement of i for any peer $j \neq i$, i.e., Equation (4.1) includes P_i but not P_j. To ensure this, P_i cannot enter i's achievement production directly (Assumption 4.1) 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 other inputs serves as a proxy for direct spillovers from unobserved peer effort in achievement production.

I argue that such a utility shifter exists in North Carolina's accountability policies, which require that students perform above a certain level on standardized End of Grade (EOG) exams (Achievement Level III) in order to be automatically promoted to the next grade. To be a valid exclusion restriction, the policy must not affect teacher incentives (because then P_i would affect i's achievement directly). This assumption seems reasonable in the present context. Under the School Based Management and Accountability Program of 1996, bonuses for schools and teachers were awarded based on the criteria that not too many students perform below Achievement Level III on the standardized EOG exams, which suggests that teachers and schools had strong incentives to shift attention to low achievers well before the introduction of student accountability.

To have any identifying power, the policy must also have different effects on students within the same peer group. Intuitively, student accountability is likely to have 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, high achievers can effectively disregard the new standards because they would probably meet the cut-off for passing even with minimal effort.

The independence of θ_i and $P_i, \bar{P}_{-i}$ (Assumption 4.2) ensures that P_{-i} does not enter i's expected utility through the distribution of θ, i.e., $f(\theta_j|X_j, \bar{X}_{-j}, K, \mu, P_j, \bar{P}_{-j}) = f(\theta_j|X_j, \bar{X}_{-j}, K, \mu)$. This ensures that i's utility-maximizing effort is a function of P_i and not P_{-i} (i.e., $e_i^* = e_i^*(e_{-i}; X_i, X_{-i}, K, \mu, P_i)$). Otherwise, $\bar{P}_{-i}$ would enter i's utility-maximizing effort through his prediction of peer utility-maximizing effort. In the present context, this means simply that the effect of the policy on the individual and his peers is independent of the unobserved individual shock after controlling for characteristics that are observable at the time the student makes his strategic choice. In other words, students in danger of being retained under the policy draw from the same distribution of θ as similar students in similar peer groups for whom the policy does not apply.

Given Assumptions 4.1 and 4.2, the system of response functions for students $i = 1, ..., N$ in a given peer group is then as follows:

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

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

$$\bar{Y}^*_{-i} = h(X_i, \bar{X}_{-i}, P_i, \bar{P}_{-i}, K, \mu). \tag{4.2}$$

Intuitively, expected peer achievement is a function of the state variables that are common knowledge to all students in the peer group, including μ, 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, μ). With $q(\cdot)$ nonlinear, this assumption, while more restrictive, still offers a fairly flexible approximation of average expected peer achievement.

Equations 4.1 and 4.2 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 4.2 but not Equation 4.1 and is plausibly independent of the unobserved components, as described in Assumption 4.3. The requirement of full independence is stronger than the mean independence required for the linear-in-means context, but is a necessary trade-off for identification of the production function under weaker functional form assumptions.

In the present context, Assumption 4.3 requires that the percentage of students in danger of failing under the new standards for promotion be independent of μ and θ_i. Independence of θ_i and $\bar{P}_{-i}$ seems fairly straightforward. However, a particular concern is that μ and $\bar{P}_{-i}$ are not independent if teachers responded in some way to these new standards (for instance, by shifting resources to low achievers) and that the unobserved classroom productivity μ and the instrument are therefore correlated. I argued above that because student accountability was preceded by school accountability, which already provided incentives for teachers to focus on low achievers, the teacher incentives would not be affected. I discuss this assumption further in Section 4.2.

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 (4.2), namely that it is strictly monotonic in the unobserved group effect as stated in Assumption 4.4. Note that an important special case where this property is satisfied is in models that assume additive separability in the unobserved components. To fix a value for μ, I assume that it is distributed $\mathcal{U}(0, 1)$. Under the above assumptions, μ 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 4.1 (Identification of μ)**.** *Given Assumptions 4.3, 4.4 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 μ can be recovered from (4.2), it remains to be shown that the structural function, $q(\cdot)$, is identified. This requires imposing the additional assumption, Assumption 4.5, that the unobserved group effect is independent of the individual type. Recall that the individual shocks θ_i can be correlated within the classroom. This assumption just requires that the state variable that is unobserved to the econometrician but observed to the students is independent of the individual shock, which is fairly straightforward given that θ_i is realized ex post.

Under Assumption 4.5, 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 τ^{th} quantile of achievement on peer achievement conditional on observed state variables $(X_i, \bar{X}_{-i}, K, P_i)$ and the common component μ.

Theorem 4.2 (Identification of the Structural Function)**.** *Given Assumptions 4.2, 4.4, and 4.5,* $q(\bar{Y}^*_{-i}, X_i, \bar{X}_{-i}, P_i, K, \mu, \theta_i)$ *is identified on the joint support of* $(\bar{Y}^*_{-i}, X_i, \bar{X}_{-i}, P_i, K, \mu, \theta_i)$*.*

See Appendix A.1 for details. The proof of this result follows from Imbens and Newey (2003, Corollary 6). Intuitively, conditioning on the unobserved group effect μ controls for the endogeneity of peer achievement, thus identifying the structural function.

4.2 Exclusion Restriction

In this section, I provide a little more background on North Carolina's student accountability policy and discuss the implications of Assumption 4.3. Student accountability began to take effect for all fifth graders in the North Carolina public school system in 2001. Thus, fifth graders prior to 2001 and fourth graders effectively act as two types of control groups. The standards are absolute, in the sense that they are not set such that a certain percentage of students fails in a given year.[28]

As mentioned before, and important aspect of the identification argument is that accountability has a different effect on students within the same classroom, i.e., affecting 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 1. 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.

Given the differential effect, if accountability itself were the instrument, I could plausibly identify the peer effect for higher 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. Because much of the current policy debate centers around improving the performance of low achievers, such an instrument would not be useful for answering some of the more pertinent questions.[29] However, the model suggests using an aggregate measure of the policy effect on peers ($\bar{P}_{-i}$) as the instrument. Intuitively, classrooms with a larger percentage of students in danger of failing will experience a larger shift in peer achievement as a result of student accountability. Such an effect is illustrated in Figure 2, which compares the distribution of achievement in two districts with different concentrations of low achievers. 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 tails 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.

Independence of $\bar{P}_{-i}$ and μ (Assumption 4.3) may not hold if teachers or administrators redistribute resources to low achievers as a result of student accountability. For instance, a

[28] In the short term, it is sufficient that low-achievers feel like there is an increased threat of retention and not that retention rates actually increases. However, I do 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).

[29] 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. As this paper illustrates, making this assumption is not very plausible assumption.

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. As mentioned before, policies that were in place 5 years earlier already provided teachers with incentives to shift attention to low achievers.

To provide further suggestive evidence in support of this argument, given that teachers and schools are forward looking, we might expect these types of resource shifts to occur at all grades. This does not appear to be the case. Returning to Figure 1, 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. That said, to help ensure that the change in achievement of the peer group is not due to shifts in the unobserved state variable μ, I control for a direct effect of accountability on high- and low-performers by including a quantile-specific shift in the achievement of students after accountability.[30]

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. If low achievers are being reassigned to classrooms in response to student accountability, 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.[31]

[30] As will be noted in Section 6, 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.

[31] It could also be the case that shifts occurred at the school level, i.e., districts reassigned students across schools. 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

4.3 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 loosely term this characteristic "observed ability" (A_i) or accumulated human capital, since it is observed by the student, to differentiate it from the unobserved component of ability that is included in θ_i. To clarify the distinction 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" next to other students, thus suggesting an unobserved component to ability. If A_i is not observed to the econometrician and students who perceive themselves to be more capable respond more to accountability policies, then P_i is not independent of the θ_i which includes observed ability (i.e., Assumption 4.2 may not hold). By similar logic, the fact that peer observed ability, A_{-i}, is unobserved to the econometrician also brings into doubt Assumption 4.3, since then $\bar{P}_{-i}$ may no longer be independent of the μ that includes peer observed ability.

Previous studies attempt to control for unobserved ability using lagged achievement.[32] While this is feasible in the present setting, a drawback is that it is unclear what lagged achievement is capturing. Prior achievement is a function of previous strategic effort choices, which may be fairly persistent when peer groups do not vary much from one grade to the next.[33] I use an alternative proxy for observed reading ability, namely the portion of leisure time spent reading for fun. This is somewhat analogous to the strategy used in Olley and Pakes (1996) and Levinsohn and Petrin (2003) in that it recovers the unobserved state variable (in their context firm productivity rather than ability) from observed inputs to production.

Partition X_i into two components, observed ability A_i and observed characteristics X_i^1, i.e., $X_i = (X_i^1, A_i)$. Similarly partition $\bar{X}_{-i}$ into two components, such that $\bar{X}_{-i} = (\bar{X}_{-i}^1, \bar{A}_{-i})$. Denote the time student i spends reading for fun and watching television by R_i and TV_i.

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.

[32] See Hanushek et al. (2003) for a detailed discussion of this approach and Todd and Wolpin (2003) for a description of some of the drawbacks.

[33] It is worth noting that the estimates of the endogenous peer spillover are robust to including lagged achievement in the specification.

Students differ in their leisure time ($T_i \equiv TV_i + R_i$). The portion of leisure time spent reading for fun is as follows:

$$R_i/T_i = \rho_i(X_i^1, A_i).$$

Observed leisure time allocation may follow from either student optimization or parental mandate. Recall that X_i^1 includes observable characteristics of the parents (such as education) that may make them more or less likely to promote reading at home. 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. This is captured in the following assumption:

Assumption 4.6. *The portion of leisure time spent reading for fun is strictly increasing in ability,* $\partial(R_i/T_i)/\partial A_i > 0$.

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. 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.

Under Assumption 4.6, ability can be expressed as a function of observable characteristics—free reading time, leisure time and individual characteristics, i.e.,

$$A_i = \rho_i^{-1}(R_i/T_i, X_i^1). \tag{4.3}$$

Similarly, peer ability can be approximated as $\bar{A}_{-i} = (1/(N-1)) \sum_{j \neq i} \rho_j^{-1}(R_j/T_j, X_j^1) \equiv \bar{\rho}_{-i}^{-1}(R_{-i}/T_{-i}, X_{-i}^1)$. 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, I get,

$$Y_i^* = q(\bar{Y}_{-i}^*, X_i^1, \bar{X}_{-i}^1, P_i, K, \rho_i^{-1}(R_i/T_i, X_i^1), \bar{\rho}_{-i}^{-1}(R_{-i}/T_{-i}, X_{-i}^1), \mu, \theta_i).$$ [34]

An alternative strategy to deal with unobserved individual heterogeneity in the linear-in-means context when panel data are available, as in the present setting, would be 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.[35] Second, even if I were to assume the location shift model as in Koenker (2004), I could not difference out the fixed effect as would be feasible in the linear-in-means case. Alternatively, it is not computationally feasible to estimate the model with a large set of dummy variables due to the large sample size and the excessive memory requirements needed to estimate the parameters.

4.4 Non-Random Assignment

A well-known problem in the identification of peer effects is non-random assignment to peer groups. In the present context, the primary source of selection bias arises from parents selecting their child'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. Non-random assignment suggests that observable and unobservable characteristics of the individual and his classroom peers may be correlated. To the extent that observable characteristics are correlated, this simply affects our interpretation of exogenous effects in our achievement production estimates. Similarly, the above arguments do not restrict the unobservable characteristics to be uncorrelated, and I account for potential correlation of θ_i in estimation.

A particular concern is that when parents exercise control over classroom assignment, more "attentive" parents, those who select the better teachers, may at the same time have

[34] Note that I assume that free reading time does not enter the production function because it is an activity that takes place outside of the classroom. However, this assumption is not necessary and only affects the interpretation of the estimates. In fact, I conduct some specification checks using forward values of free reading time, which could not directly affect prior achievement, and the results are unchanged.

[35] See Koenker (2004) for a description of quantile regressions using longitudinal data.

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. To some extent, the concern for certain types of selection are alleviated by including ability controls.

However, selection in the present context remains problematic if students are assigned to classrooms based on the unobserved μ. The primary type of selection that occurs within schools, be it at the principal or parent level, is likely to be on perceived teacher quality. Further assuming that teachers stay in similar "quality" schools with similar "quality" classrooms 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 (4.1) as follows:

$$Y_i^* = \tilde{q}(\bar{Y}_{-i}^*, X_i, \bar{X}_{-i}, P_i, K^1, \mu, \theta_i) + \beta(\theta_i)Tch, \tag{4.4}$$

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 Estimation

Suppose time is indexed $t = 1, ..., T$ and classrooms $c = 1, ..., C$. Estimation of the quantile structural function proceeds in two steps. First, I recover the residual from Equation (4.2), the reduced form regression predicting the ex ante expected value of peer achievement. I then estimate the quantile structural function defined in Equation (4.4), controlling for the residual from the first stage which captures the unobserved group effect. Each of these steps is discussed in detail in the following sections.

5.1 Reduced Form

The first stage of the estimator is analogous to the first stage in 2SLS. I recover the unobserved classroom productivity, the correlated effect, from a reduced-form equation for peer achievement, as described in (4.2). As discussed previously, allowing the spillovers to vary across races and to vary across different race-based reference groups is an important feature of this analysis. Let NW_i be an indicator for a nonwhite student, and the superscripts $k \in \{W, NW\}$ indicate white and nonwhite respectively. Then, $\bar{Y}^{NW}_{-ict} = \frac{1}{\sum_j NW_j - NW_i}(\sum_j NW_j Y^*_j - NW_i Y^*_i)$ denotes the observed mean achievement of student i's nonwhite classroom peers and similarly $\bar{Y}^{W}_{-ict}$ for white peers.

The reduced-form equation for peer achievement is approximated as

$$\bar{Y}^k_{-ict} = \alpha_0 + X_{it}\alpha_1 + \bar{X}_{-ict}\alpha_2 + \alpha_3 P_{it} + \vec{\bar{P}}_{-ict}\alpha_4 + K^1_{ct}\alpha_5 + Tch_{it} + Year_t + \mu_{ct} + \delta_{ict}, \quad (5.1)$$

where dependence of the parameters on the race subgroup k is suppressed. The covariates X_{it} include the sex of the student, parental education, a dummy for low-performance defined as whether the student performed below the 30^{th} 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 $\bar{X}_{-ict}$, 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^1_{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 30^{th} 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., $\vec{\bar{P}}_{-ict} = \{\bar{P}^W_{-ict}, \bar{P}^{NW}_{-ict}\}$.

The remaining residual δ_{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., $\bar{Y}_{-ict} = \bar{Y}^*_{-ict} + \delta_{ict}$. Given that classes are sufficiently large, about 23 students on average, δ_{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 (5.1) and four values of the predicted teacher fixed effects, $\hat{T}ch_{it}$. The triangular structure in (5.1) 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.

5.2 Quantile Structural Function

In the second stage, I estimate the structural function (4.4), which describes a student's achievement as a function of peer characteristics and peer achievement at different points of the conditional achievement distribution. If the mean effect of peers were sufficient, I could proceed simply using the familiar 2SLS estimator, estimating the second stage also as a mean regression. However, as mentioned before, regrouping students does not affect average achievement in the linear-in-means context because the losses to one student are perfectly offset by the gains to another. Previous studies have also acknowledged the importance of capturing these types of nonlinearities, but pursued alternative strategies, such as 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.[36] 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 approach offers considerable flexibility, can be estimated for a large number of quantiles,[37] and is not sensitive to outliers.

While it is feasible to estimate the quantile structural function without assuming a parametric form,[38] 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 (4.4) as

$$\begin{aligned} Y^*_{ict} = \beta_0 + \beta_1 \bar{Y}^W_{-ict} + \beta_2 \bar{Y}^{NW}_{-ict} + X_{it}\beta_3 + \bar{X}_{-ict}\beta_4 + \beta_5 P_{it} \\ + K^1_{ct}\beta_6 + \beta_7 \hat{Tch}^W_{it} + \beta_8 \hat{Tch}^{NW}_{it} + Year_t + \beta_9 \hat{\mu}^W_{ct} + \beta_{10}\hat{\mu}^{NW}_{ct} + u_{ict}, \end{aligned} \quad (5.2)$$

where dependence of the parameters on the quantile ($\beta(\theta_i)$) and race is suppressed to sim-

[36]See Hanushek et al. (2004), Hanushek et al. (2003) and Hoxby and Weingarth (2005).
[37]This is emphasized in Chernozhukov and Hansen (2005).
[38]See Imbens and Newey (2003) for a discussion of the fully nonparametric estimator.

plify notation. Teacher fixed effects are allowed to vary by race, capturing the fact that the effectiveness of the teacher may vary across races.[39] 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.[40] 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 $\hat{\vec{\beta}}(\tau)$ solve the following optimization problem:

$$\text{argmin}_{\vec{\beta}(\tau)} \frac{1}{NCT} \sum_i \sum_c \sum_t \rho_\tau(u_{ict}),$$
$$\text{where } \rho_\tau(u_{ict}) = \tau u_{ict}^+ + (1-\tau) u_{ict}^-.$$

Before proceeding to the results, 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. 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

[39] 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. For instance, see Ferguson (1998).

[40] 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 $\bar{Y}_{-ict}$ not $\hat{\bar{Y}}_{-ict}$.

achievement. Given these three countervailing effects, the sign of $\hat{\beta}_4$ is indeterminate.[41]

A similar conclusion holds for classroom productivity, μ. 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 μ is only an element in S_i, students do not choose effort, and there are no spillovers from peer effort in production. Otherwise, higher μ could negatively affect achievement through its inverse relation with peer effort. Secondly, higher μ 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.

6 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 6.1 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 6.2, I present results for the two-stage quantile estimator described in Section 5, and allow peer spillovers to vary by race. In Section 6.3, I compare the relative magnitude of exogenous and endogenous peer effects.

6.1 Baseline: Linear-in-Means

Before exploring the distributional effects of peers, it is useful to begin with 2SLS 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 stage is then

$$Y^*_{ict} = \beta_0 + \bar{Y}_{-ict}\beta_1 + X_{it}\beta_2 + \bar{X}_{-ict}\beta_3 + \beta_4 P_{it} + K^1_{ct}\beta_5 + Tch_{it} + Year_t + \beta_6\hat{\mu}_{ct} + \theta_{ict},$$

[41] Cooley (2006) expands on the intuition for the interpretation of exogenous effects in the context of various peer spillover mechanisms.

where all exogenous effects $\bar{X}_{-ict}$ are defined at the class level.[42] Parameter estimates from the first and second stages of this estimator are presented in the second two columns of Table 3. The first two columns correspond to 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.[43]

I begin by comparing estimates of the endogenous effect (β_1) across my estimator and the comparison cases. 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 even more severely underestimates peer spillovers, as the coefficient on peer test scores is -.15. If the direct effect of unobserved group productivity μ 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 μ, after controlling for teacher fixed effects, is coming through peer effort spillovers to production or strategic effort choices rather than a direct effect.[44] 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.[45]

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,[46] while 2SLS predicts a larger negative effect. Intuitively, as discussed in Section 5.2 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

[42]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^1_{ct}.

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

[44]These results suggesting downward bias in OLS estimates are very sensitive to changes in specification. 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.

[45]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.

[46]Estimates using twice-lagged peer achievement with school by year fixed effects give positive effects from peer's parental education, suggesting that the spillovers from parents in the classroom may be driven primarily from some type of teacher selection mechanism within schools.

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 instrument. 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.[47]

6.2 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 described in Section 5 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 signficant, 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 larger for nonwhite students and particularly large for the effect of white peers on nonwhites. One potential explanation is

[47]Note that if resources were redistributed from high to low performers, this would suggest that the endogenous spillovers are even *larger*.

that the designation "nonwhite" groups Hispanic and black students together, and the within peer group effects may not come through as strongly as a result. Another explanation is that this is indicative of the particularly complex nature of spillovers from whites to nonwhites documented in the literature. For instance, 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. As a specification check, 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. However, both stories suggest that more careful investigation of the mechanisms of peer influences particularly for black students is merited.

While Table 4 suggests that peer spillovers vary by race, Figure 3 describes 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, though the response to nonwhite peers appears to be more robust. 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, 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 lower tail of the achievement distribution.

6.3 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 corresponding to those shown in Figure 3. 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, though again only nonwhite peer achievement is statistically significant.

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 4, which compares the quantile derivatives of the portion of peers who are nonwhite on white and nonwhite achievement. 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. 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.

The table also considers the other exogenous peer effects, with the main conclusion being that endogenous effects appear to be considerably larger than exogenous effect. This 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 7.1 paints a clearer picture of the effect of desegregating peer groups taking into account both endogenous and exogenous effects.

7 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, 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 provide some context, 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. I consider an experiment of desegregating schools across the lines of two such neighboring school districts in North Carolina–Durham (home of Duke University) and Chapel Hill (home of University of North Carolina-Chapel Hill). As illustrated in Table 7, the contrast across these districts is very stark. 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. Students in Durham schools have average reading scores that are about 70% of a standard deviation lower than Chapel Hill students. The contrast further extends to indicators of socioeconomic status; 80% of Chapel Hill parents have at least a 4-year degree compared to only 37% of Durham parents.

Yet, Table 7 also provides suggestive evidence of considerable differences in within-district segregation. The average class for a white Durham student is only 51% nonwhite, while the average class for a nonwhite Durham student is 76% nonwhite. In contrast, the racial composition of the classroom does not appear to vary as much in Chapel Hill schools. The standard deviation in the percentage nonwhite peers in the average class in Durham is .22, compared to .11 for Chapel Hill. This within-district segregation is even more disconcerting taken in the context of recent trends. Evidence suggests that segregation in North Carolina's public schools has actually increased over the past decade, a marked contrast to the sharp declines in segregation over the previous decades.[48] The resegregation of public schools 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.[49]

At the same time that segregation is increasing in public schools, North Carolina policymakers advocate desegregated/mixed-ability classrooms as most conducive to narrowing the racial achievement gap, though there is little evidence on the magnitude of such an effect. With these observations in mind, Section 7.1 simulates achievement outcomes when students are randomly assigned to classrooms within the district and compares this to random assignment across districts. Because I hold the student population to be fixed in the experiment, this clearly abstracts away from important issues of residential sorting, proximity constraints and the potential to select out of public schools. To provide evidence on the returns to "better" peer groups in a context where general equilibrium sorting effects are unlikely to be important, I then estimate the achievement premium associated with moving a student from Durham to Chapel Hill in Section 7.2.

7.1 Desegregating Peer Groups

To begin to quantify the potential effects of desegregation, I perform an experiment where I randomly assign students to classrooms, thereby creating racially desegregated and mixed-ability peer groups. 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 races does not obscure the impact of altering peer groups. While various efficiency and equity measures are relevant, I begin by comparing the

[48]See Clotfelter et al. (2005).

[49]For instance, the Fourth Circuit Federal 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. See Chemerinsky (2003) for a more detailed discussion of the courts' role.

average achievement for whites and nonwhites, the gap between average white and nonwhite achievement and the overall average achievement under these alternative assignment policies.

Table 8 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. The 95% confidence intervals are reported in brackets. I use parameter estimates from the two-stage quantile regression to simulate achievement outcomes under the alternative assignment policies.[50]

In both districts, within-district desegregation creates the largest gains nonwhite students, around .04 on average, though whites also have marginal gains in achievement on average. However, the effect is only statistically significantly different for nonwhites in Durham, perhaps partially due to sample sizes. As a result, the achievement gap decreases a little, by about .02 in both districts, but it does not appear to be a statistically significant change. Not surprisingly, desegregation across districts produces gains for Durham students and marginal losses to Chapel Hill students on average, compared to within-district desegregation. Nonwhites in Durham gain .05 of a standard deviation relative to within-district desegregation, while nonwhites in Chapel Hill lose .01. Whites in Durham make comparable gains to nonwhites, .05, while whites in Chapel Hill make comparable losses to nonwhites, .01. Thus, the point estimates suggest that the gains (losses) to nonwhite achievement are offset by gains (losses) to white achievement for students in each of the districts with the result that desegregation across districts has minimal additional effects on the gap.

The confidence intervals further show that the losses to Chapel Hill students from the merger are not statistically significantly different from the observed groupings or desegregation within districts. Furthermore, the gains to nonwhite Durham students are not significantly different from the merger relative to the within district desegregation. The real gains for nonwhite Durham students seem to be coming from the within district segregation. The subgroup whose gains from the merger are most significant are the white Durham students. This is because the larger effects of peers are within race and lower-achieving white Durham students are benefitting from their placement with higher-achieving white Chapel Hill peers.

[50]In the simulations, it is necessary to assign a value of θ_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 θ_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.

We do not see offsetting losses for white Chapel Hill students because the marginal effect of peers is diminishing across quantiles for whites, and Chapel Hill students are more highly concentrated in these upper quantiles where the peer effects are not as important.

Figure 5 provides further evidence on the effects of the merger on the gap and achievement at different points of the achievement distribution. The benefits of the merger compared to the observed groupings are presented in the left-hand-side figure, while the benefits of the merger over within-district integration are presented on the right-hand-side. In both cases, the gains appear to be diminishing across percentiles of the achievement distribution for the population overall. Comparing to the observed groupings, the nonwhite students make the largest gains from the merger, ranging from .06 to .14, while whites gain between 0 and .08. As shown in Figure 6, the white-nonwhite achievement gap remains highest at the lowest percentiles of the achievement distribution and lower at the upper percentiles under each of the alternative groupings. Within district integration (relative to observed groupings) has ambiguous effects on the percentile achievement gap, actually increasing the gap at some percentiles. However, the merger narrows the gap at all percentiles compared to within-district integration, with some of the largest gains being for students in the middle percentiles

As emphasized above, an innovation of this paper is allowing for heterogeneity in peer effects by estimating quantile rather than mean regressions. To provide further evidence on the importance of this dimension for policy implications, I compare predictions from the 2SLS estimator, where the marginal effects are allowed to vary by race, to the predictions of the two-stage quantile regression estimator. As Figure 7 shows, the predictions of the 2SLS and two-stage quantile regression estimators for the achievement gap at different percentiles of the achievement distribution are quite different. Most remarkably, the 2SLS estimator severely overpredicts the achievement gap for the lower to middle percentiles. Furthermore, the 2SLS regression actually predicts increases in the achievement gap from integration at the lowest quantiles, whereas the two stage quantile regression predicts in most cases that the gap is actually narrowed for the lower quantiles, particularly from the integration across district lines. This contrast is particularly important given the important placed on improving outcomes for students in the lowest percentiles.

It is worth noting that the Durham public school district is much larger—2099 students versus only 667 in Chapel Hill. On the one hand, because of the large discrepancy in achievement, we would expect the benefits from the merger of these two districts to capture something of an upper bound on the benefits of integration, but as there are significantly fewer Chapel Hill students it is more difficult to make this case. Thus, I also conducted a much broader experiment where I randomly assigned students to any school in the state of North Carolina. The results were very similar to this localized example, in that the larger experiment only narrowed the average gap by about 6% of a standard deviation. This narrowing was not only fairly uniform across percentiles of the achievement distribution, but the gains were largely through improvements in average nonwhite achievement rather than losses in average white achievement, as suggested by the local experiment.

7.2 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 from random assignment to a classroom in a Chapel Hill public school and the effect on a Chapel Hill student from random assignment to a Durham school. Again, I focus on fifth graders in the 2002-03 academic year and assign to every student the average correlated effect and teacher fixed effect to isolate the peer effect. The results in Section 5 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 9 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 9 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 effects of moving a student from a higher-performing to a lower-performing peer group? Table 9 reveals that there are 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, without even accounting for other non-peer related differences in school quality.

8 Conclusion

To answer the central question posed in this paper, desegregating peer groups only appears to lead to marginal improvements in the achievement gap. 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, generally only spillovers from nonwhite peers are statistically significant. The lack of cross-racial spillovers limits the effect of desegregation. I perform an experiment of merging a predominantly white, higher-achieving school district with a neighboring lower-performing, racially-mixed school district, where one might expect the returns to be particularly large, and find only small 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. However, there do appear to be gains from removing a nonwhite student from a low-achieving, predominantly nonwhite peer group as illustrate in the partial equilibrium experiment, though again the benefit primarily arises from the removal of negative peer influences.

To place my results in context, previous studies attempting to link changes in black outcomes to desegregation have met with mixed success. The failure to account for contemporaneous peer effects may explain some of the mixed evidence in the literature, since ignoring endogenous effects when they exist will bias estimates of exogenous effects, or in the present context, peer racial composition or ability.[51] Furthermore, while I do find a negative correlation between the percentage nonwhite in the classroom and achievement for white and nonwhite students, the effects are very small, with a one standard deviation decrease in the percentage nonwhite (more than 20%) only raising achievement by about .02 of a standard deviation.

While the focus of this paper is on isolating one mechanism through which desegregation may help narrow the achievement gap–peers–it is worth emphasizing that a full assessment of the effect of desegregation would need to take other factors into account. In particular, I do not explore other potential effects of desegregation, such as the reallocation of resources and general equilibrium effects of residential sorting or selection out of public schools. 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.[52] I discuss the conditions needed to identify peer effects under

[51] See Moffitt (2001) for a careful discussion.

[52] Models of strategic interactions among agents are also important in the industrial organization literature, particularly in models of entry and auction models. For instance, see Ackerberg and Gowrisankaran

minimal functional form assumptions. In particular, I describe a set of informational assumptions and conditions on the quantile structural function that permit quantile instrumental variables techniques to be applied to the identification of peer effects. 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 percentiles of the achievement distribution, moving from the typical linear-in-means model to a more flexible model leads to a much richer understanding of the effect 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 tradeoffs inherent in alternative grouping strategies and in particular, the effect on the lower quantiles of the achievement distribution.

This paper can ultimately serve as a building block to answer the broader and more elusive question of what is the "optimal" grouping of students in classroom given different policy objectives. A useful, and straightforward, extension of the model would allow other moments of peer achievement, beyond the mean, to enter achievement. This would be particularly important for evaluating the effect of tracking and developing broader implications for optimal classroom assignment. Furthermore, when placed in the larger context of a general equilibrium model, the production function estimates could help provide a deeper understanding of the effects of alternative school choice mechanisms and competition.

(Forthcoming), Aguirregabiria and Mira (Forthcoming), Bresnahan and Reiss (1991), Ciliberto and Tamer (2004), Pakes et al. (2005), Pesendorfer and Schmidt-Dengler (2003), and Rysman (2004). See Athey and Haile (Forthcoming) for an overview of auction models.

References

Abadie, A., Angrist, J. and Imbens, G. (2002), 'Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings', *Econometrica* **70**(1), 91–117.

Ackerberg, D. A. and Gowrisankaran, G. (Forthcoming), 'Quantifying equilibrium network externalities in the ACH banking industry', *RAND Journal of Economics* .

Aguirregabiria, V. and Mira, P. (Forthcoming), 'Sequential estimation of dynamic discrete games', *Econometrica* .

Arnott, R. and Rowse, J. (1987), 'Peer group effects and educational attainment', *Journal of Public Economics* **32**, 287–305.

Athey, S. and Haile, P. A. (Forthcoming), Nonparametric approaches to auctions, *in* J. Heckman and E. Leamer, eds, 'Handbook of Econometrics', Vol. 6, Elsevier, Amsterdam.

Betts, J. R. and Zau, A. (2002), Peer groups and academic achievement: Panel evidence from administrative data. Working Paper.

Bishop, J. H., Bishop, M., Gelbwasser, L., Green, S. and Zuckerman, A. (2003), Nerds and freaks: A theory of student culture and norms.

Bresnahan, T. F. and Reiss, P. C. (1991), 'Entry and competition in concentrated markets', *The Journal of Political Economy* **99**(5), 977–1009.

Brock, W. A. and Durlauf, S. N. (2001*a*), 'Discrete choice with social interactions', *The Review of Economic Studies* **68**(2), 235–260.

Brock, W. A. and Durlauf, S. N. (2001*b*), Interactions-based models, *in* J. Heckman and E. Leamer, eds, 'Handbook of Econometrics', Vol. 5, Elsevier, Amsterdam, pp. 3297–3380.

Card, D. and Rothstein, J. (2005), Racial segregation and the black-white test score gap. Working Paper.

Chemerinsky, E. (2003), 'The segregation and resegregation of American public education: The courts' role', *North Carolina Law Review* **81**(4), 1597–1622.

Chernozhukov, V. and Hansen, C. (2005), 'An IV model of quantile treatment effects', *Econometrica* **73**(1), 245–262.

Chesher, A. (2003), 'Identification in nonseparable models', *Econometrica* **71**, 1405–1441.

Ciliberto, F. and Tamer, E. (2004), Sequential estimation of dynamic discrete games. Working Paper.

Clotfelter, C. T., Ladd, H. F. and Vigdor, J. L. (2005), Federal oversight, local control, and the specter of "resegregation" in southern schools. NBER Working Paper #11086.

Cook, M. D. and Evans, W. N. (2000), 'Families or schools? Explaining the convergence in white and black academic performance', *Journal of Labor Economics* **18**, 729–754.

Cooley, J. C. (2006), Alternative mechanisms of peer spillovers in the classroom: Interpreting exogenous peer effects. Working Paper.

Ferguson, R. F. (1998), Teachers' perceptions and expectations and the black-white test score gap, *in* C. Jencks and M. Phillips, eds, 'The Black-White Test Score Gap', The Brookings Institution, Washington, DC, chapter 8, pp. 273–317.

Figlio, D. N. and Page, M. E. (2002), 'School choice and the distributional effects of ability tracking: Does separation increase equality?', *Journal of Urban Economics* **51**, 497–514.

Fryer, Jr., R. G. and Torelli, P. (2005), An empirical analysis of 'acting white'. NBER Working Paper #11334.

Graham, B. S. (2004), Identifying social interactions through excess variance contrasts. Working Paper.

Guryan, J. (2004), 'Desegregation and black dropout rates', *American Economic Review* **94**(4), 919–943.

Hanushek, E. A., Kain, J. F., Markman, J. M. and Rivkin, S. G. (2003), 'Does peer ability affect student achievement?', *Journal of Econometrics* **18**(5), 527–544.

Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (2004), New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. Working Paper.

Honoré, B. E. and Hu, L. (2004), 'On the performance of some robust intrumental variables estimators', *Journal of Business and Economic statistics* **22**(1), 30–39.

Hoxby, C. M. and Weingarth, G. (2005), Taking race out of the equation: School reassignment and the structure of peer effects. Working Paper.

Imbens, G. W. and Newey, W. K. (2003), Identification and estimation of triangular simultaneous equations models without additivity. Working Paper.

Jencks, C. and Phillips, M., eds (1998), *The Black-White Test Score Gap*, The Brookings Institution, Washington, DC.

Koenker, R. (2004), 'Quantile regression for longitudinal data', *Journal of Multivariate Analysis* **91**(1), 74–89.

Lazear, E. P. (2001), 'Educational production', *Quarterly Journal of Economics* **116**(3), 777–803.

Levinsohn, J. and Petrin, A. (2003), 'Estimating production functions using inputs to control for unobservables', *Review of Economic Studies* **70**, 317–341.

Ma, L. and Koenker, R. (2004), Quantile regression methods for recursive structural equation models. Working Paper.

Manski, C. (1993), 'Identification of endogenous social effects: The reflection problem', *The Review of Economic Studies* **60**(3), 531–542.

Milgrom, P. and Shannon, C. (1994), 'Monotone comparative statics', *Econometrica* **62**(1), 157–180.

Moffitt, R. A. (2001), Policy interventions, low-level equilibria and social interactions, *in* S. N. Durlauf and H. P. Young, eds, 'Social Dynamics', Brookings Institution, Washington, DC, pp. 45–82.

Olley, G. S. and Pakes, A. (1996), 'The dynamics of productivity in the telecommunications equipment industry', *Econometrica* **64**(6), 1263–1297.

Pakes, A., Ostrovsky, M. and Berry, S. (2005), Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). Working Paper.

Pesendorfer, M. and Schmidt-Dengler, P. (2003), Identification and estimation of dynamic games. Working Paper.

Rivkin, S. G. (2000), 'School desegregation, academic attainment, earnings', *The Journal of Human Resources* **35**(2), 333–346.

Rysman, M. (2004), 'Competition between networks: A study of the market for yellow pages', *Review of Economic Studies* **71**, 483–512.

Simmons, T. (1999), 'Lower expectations limit black's potential', *Raleigh News and Observer* .

Slavin, R. E. (1990), 'Achievement effects of ability grouping in secondary schools: A best-evidence synthesis', *Review of Educational Research* **60**(3).

Sweeting, A. (2004), Coordination games, multiple equilibria and the timing of radio commercials. Working Paper.

Todd, P. E. and Wolpin, K. I. (2003), 'On the specification and estimation of the production function for cognitive achievement', *Economic Journal* **113**(485), F3–F33.

Vigdor, J. and Nechyba, T. (2004), Peer effects in North Carolina Public Schools. Working Paper.

A Appendix

A.1 Proofs

Proof of Theorem 3.2. I illustrate how the game in effort maps into a game in achievement. Given Assumption 3.2, ex ante expected achievement can proxy for effort. Denote the ex ante expected value of achievement ($\tilde{Y}_i$) as

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

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

$$\begin{aligned} e_1 &= \tilde{g}^{-1}(\tilde{Y}_1, e_2, ..., e_N; S) \\ &\vdots \\ e_N &= \tilde{g}^{-1}(\tilde{Y}_N, e_1, ..., e_{N-1}; S). \end{aligned}$$

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

$$\begin{aligned} e_{-i} &= (..., G(\tilde{Y}_{i-1}, \tilde{Y}_{-(i-1)}; S), G(\tilde{Y}_{i+1}, \tilde{Y}_{-(i+1)}; S), ...) \\ &\equiv G_{-i}(\tilde{Y}_i, \tilde{Y}_{-i}; S). \end{aligned}$$

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

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

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 $\tilde{Y}_{-i}$:

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

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

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

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

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

□

Proof of Theorem 4.1. Because all results hold conditional on the exogenous state variables $(X_i, \bar{X}_{-i}, K, P_i)$, dependence on these variables is suppressed. Following the proof in Imbens and Newey (2003):

$$\begin{aligned}
F_{\bar{Y}_{-i}^*|\bar{P}_{-i}}(\bar{Y}_{-i}^*|\bar{P}_{-i}) &\overset{(1)}{=} Pr(\bar{Y}_{-i}^* \leq \bar{y}_0|\bar{p}_0) \\
&\overset{(2)}{=} Pr(h(\bar{P}_{-i}, \mu) \leq \bar{y}_0|\bar{p}_0) \\
&\overset{(3)}{=} Pr(\mu \leq h^{-1}(\bar{p}_0, \bar{y}_0)|\bar{p}_0) \\
&\overset{(4)}{=} Pr(\mu \leq h^{-1}(\bar{p}_0, \bar{y}_0)) \\
&\overset{(5)}{=} F_\mu(h^{-1}(\bar{p}_0, \bar{y}_0)).
\end{aligned}$$

The first equality follows by definition; the second by the representation of peer achievement in (4.2). The third follows by Assumption 4.4, and the fourth by Assumption 4.2. Therefore, $\mu = h^{-1}(\bar{p}_0, \bar{y}_0)$ is identified by the joint distribution of $(\bar{Y}_{-i}^*, \bar{P}_{-i})$. □

Proof of Theorem 4.2. Following the proof in Imbens and Newey (2003, Corollary 6):

$$
\begin{aligned}
&F_{Y_i^*|\bar{Y}_{-i}^*,\mu}(Y_i^*|\bar{Y}_{-i}^*,\mu) \\
&\qquad = Pr(Y_i^* \leq y_0|\bar{y}_0,\mu_0) \\
&\qquad = Pr(q(\bar{Y}_{-i}^*,\mu,\theta_i) \leq y_0|\bar{y}_0,\mu_0) \\
&\qquad = Pr(\theta_i \leq q^{-1}(\bar{y}_0,\mu_0,y_0)|\bar{y}_0,\mu_0) \\
&\qquad = Pr(\theta_i \leq q^{-1}(\bar{y}_0,\mu_0,y_0)) \\
&\qquad = F_{\theta_i}(q^{-1}(\bar{y}_0,\mu_0,y_0)) = q^{-1}(\bar{y}_0,\mu_0,y_0).
\end{aligned}
$$

Since the inverse of the structural function is identified, the function itself is also identified on the joint support of $(\bar{Y}_{-i}^*,\mu,\theta_i)$. □

A.2 Tables

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

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

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

	White		Nonwhite	
	Mean	Std. Dev.	Mean	Std. Dev.
Reading score (standardized)	0.3642	0.9021	-0.3038	0.8853
Class size	22.84	3.43	21.90	3.64
Male	0.5044	0.5000	0.4882	0.4999
Parent HS/some post-sec.	0.5852	0.4927	0.7580	0.4283
Parent 4-year degree+	0.3664	0.4818	0.1514	0.3585
Free/reduced price lunch	0.2431	0.4290	0.7029	0.4570
Free reading hours per week	0.9436	0.6281	0.8344	0.6170
TV hours per week	2.214	1.497	2.974	1.852
Characteristics of Classroom Peer Groups				
Avg. peer reading score	0.2049	0.4286	-0.0444	0.4588
Avg. white peer reading score	0.3490	0.4435	0.2450	0.5322
Avg. nonwhite peer reading score	-0.2395	0.5955	-0.3324	0.4646
% nonwhite	0.2384	0.2099	0.5244	0.2643
% parents HS degree	0.6174	0.2169	0.6857	0.1959
% parents 4-year degree	0.3168	0.2377	0.2465	0.2081
% FRP lunch	0.3410	0.2084	0.5191	0.2473
% without peers of other race	0.1468	0.3539	0.0637	0.2442
N	587,483		288,693	

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 3: Comparison of Contemporaneous to Lagged Achievement
(Dependent variable: standardized reading score)

	OLS Contemporaneous	OLS Twice-Lagged	1st Stage Contemporaneous	2nd Stage Contemporaneous
Avg. peer reading	-0.0715***	-0.1454***		0.4794***
	[0.0088]	[0.0081]		[0.0511]
Accountability	0.0033	0.1353***	-0.0200***	-0.0101***
	[0.0039]	[0.0200]	[0.0053]	[0.0024]
Low reading	-1.1135***	-1.1472***	-0.0096***	-1.1077***
	[0.0024]	[0.0044]	[0.0006]	[0.0037]
Accountable*low reading	0.1081***	0.1414***	0.0043***	0.1038***
	[0.0041]	[0.0055]	[0.0008]	[0.0038]
% low reading	-0.2234***	-0.3098***	-1.1831***	0.3977***
	[0.0143]	[0.0174]	[0.0110]	[0.0571]
Accountable*% low reading			0.1733***	
			[0.0144]	
Nonwhite	-0.2642***	-0.2535***	-0.0039***	-0.2620***
	[0.0020]	[0.0028]	[0.0005]	[0.0024]
Male	-0.0632***	-0.0520***	-0.0026***	-0.0617***
	[0.0015]	[0.0021]	[0.0006]	[0.0020]
Parent HS degree	0.2139***	0.2159***	-0.0030***	0.2156***
	[0.0033]	[0.0046]	[0.0008]	[0.0029]
Parent 4-year+	0.5672***	0.5591***	-0.0012	0.5680***
	[0.0037]	[0.0052]	[0.0009]	[0.0033]
Free reading hours	0.2266***	0.2512***	0.0071***	0.2227***
	[0.0035]	[0.0051]	[0.0009]	[0.0048]
% nonwhite	-0.0939***	-0.1002***	-0.3520***	0.1000***
	[0.0108]	[0.0163]	[0.0096]	[0.0187]
% male	-0.0559***	-0.0599***	-0.1245***	0.0123
	[0.0122]	[0.0172]	[0.0113]	[0.0096]
% HS degree	-0.0555***	0.0169	0.2021***	-0.1654***
	[0.0170]	[0.0242]	[0.0157]	[0.0137]
% 4-year degree	0.0109	0.1478***	0.6110***	-0.3240***
	[0.0185]	[0.0269]	[0.0167]	[0.0326]
Avg. free reading hrs.	0.1479***	0.0858***	0.3493***	-0.0431**
	[0.0187]	[0.0277]	[0.0171]	[0.0201]
N	876,176	421,787	876,176	876,176
R^2	0.5356	0.5541	0.9042	0.4364

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

Table 4: Heterogeneous Reference Groups
(Dependent variable: standardized reading score)

	2SLS		Median	
	White	Nonwhite	White	Nonwhite
Avg. white peer reading	0.3957***	0.1973	0.3867***	0.0857
	[0.0901]	[0.1360]	[0.0909]	[0.1629]
Avg. nonwhite peer reading	0.0157	0.1585	0.0182	0.2729*
	[0.0449]	[0.1577]	[0.0536]	[0.1462]
Accountability	-0.0026	0.0049	-0.0150***	-0.0184**
	[0.0037]	[0.0085]	[0.0042]	[0.0089]
Low reading	-1.1832***	-0.9991***	-1.1525***	-0.9788***
	[0.0037]	[0.0036]	[0.0049]	[0.0054]
Accountable*Low reading	0.1251***	0.0606***	0.1542***	0.1076***
	[0.0056]	[0.0062]	[0.0067]	[0.0074]
% white low reading	0.3673***	0.1822	0.3510***	0.0474
	[0.1055]	[0.1550]	[0.1078]	[0.1841]
% NW low reading	-0.0037	0.1032	-0.0037	0.2205
	[0.0439]	[0.1517]	[0.0526]	[0.1349]
% nonwhite	-0.0646***	-0.1050***	-0.0851***	-0.0769***
	[0.0105]	[0.0209]	[0.0120]	[0.0194]
Male	-0.0495***	-0.0881***	-0.0429***	-0.0860***
	[0.0016]	[0.0025]	[0.0021]	[0.0039]
Parent HS degree	0.2636***	0.1531***	0.2633***	0.1497***
	[0.0043]	[0.0046]	[0.0055]	[0.0074]
Parent 4-year +	0.6087***	0.4174***	0.6218***	0.4107***
	[0.0046]	[0.0083]	[0.0063]	[0.0138]
Free reading hours	0.2745***	0.0739***	0.2973***	0.0592***
	[0.0043]	[0.0057]	[0.0064]	[0.0065]
% male	-0.0039	0.0026	-0.0141	0.0092
	[0.0143]	[0.0293]	[0.0180]	[0.0343]
% HS degree	-0.1392***	-0.1209***	-0.1488***	-0.1705***
	[0.0240]	[0.0431]	[0.0247]	[0.0290]
% 4-year degree	-0.2654***	-0.2023**	-0.2727***	-0.2185***
	[0.0576]	[0.0915]	[0.0566]	[0.0749]
Avg. free reading hrs.	-0.0232	0.0390	-0.0263	0.0495
	[0.0425]	[0.0600]	[0.0429]	[0.0599]
No peers of other race	0.0008	0.0576	0.0009	0.0214
	[0.0092]	[0.0487]	[0.0111]	[0.0492]
N	587,484	288,693	587,483	288,693
R^2	0.4895	0.4941		

*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.

Table 5: Average Marginal Effects of Peers for Whites
(Dependent variable: standardized reading score)

	Mean	.1 Quantile	Median	.9 Quantile
Avg. white reading	0.1841***	0.3220***	0.1715***	0.0971
	[0.0493]	[0.0574]	[0.0403]	[0.0642]
Avg. nonwhite reading	0.0032	-0.0216	0.0101	-0.0056
	[0.0335]	[0.0450]	[0.0298]	[0.0430]
% white low performing	0.0586***	0.1156***	0.0526***	0.0238
	[0.0198]	[0.0227]	[0.0162]	[0.0256]
% NW low performing	-0.0044	-0.0188	-0.0012	-0.0070
	[0.0184]	[0.0249]	[0.0164]	[0.0237]
% nonwhite	-0.0179***	-0.0127***	-0.0179***	-0.0215***
	[0.0046]	[0.0065]	[0.0025]	[0.0051]
% male	-0.0002	0.0020	-0.0012	-0.0015
	[0.0019]	[0.0021]	[0.0016]	[0.0024]
% parents with HS degree	-0.0310***	-0.0343***	-0.0323***	-0.0311***
	[0.0058]	[0.0074]	[0.0053]	[0.0077]
% parents with 4-year degree	-0.0660***	-0.0973***	-0.0648***	-0.0468**
	[0.0160]	[0.0194]	[0.0135]	[0.0211]
Avg. free reading	-0.0022	-0.0092**	-0.0019	0.0031
	[0.0039]	[0.0048]	[0.0031]	[0.0049]

*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 two-stage quantile regression regression broken out by subgroup, i.e., that depicted in Figures 3 and 4. Marginal effects are averaged over quantiles for the first column.

Table 6: Average Marginal Effects of Peers for Nonwhites
(Dependent variable: standardized reading score)

	Mean	.1 Quantile	Median	.9 Quantile
Avg. white reading	0.1095	0.1589*	0.0444	0.1413
	[0.0819]	[0.0958]	[0.0845]	[0.0930]
Avg. nonwhite reading	0.1056*	0.0777	0.1265*	0.1285**
	[0.0625]	[0.0842]	[0.0677]	[0.0651]
% white low performing	0.0426	0.0687	0.0103	0.0564
	[0.0385]	[0.0453]	[0.0399]	[0.0436]
% NW low performing	0.0420	0.0345	0.0508	0.0504
	[0.0293]	[0.0398]	[0.0311]	[0.0307]
% nonwhite	-0.0194**	-0.0331***	-0.0203***	-0.0084
	[0.0089]	[0.0127]	[0.0051]	[0.0092]
% male	0.0014	0.0011	0.0009	0.0030
	[0.0033]	[0.0044]	[0.0032]	[0.0034]
% parents with HS degree	-0.0352***	-0.0433***	-0.0334***	-0.0299***
	[0.0059]	[0.0079]	[0.0057]	[0.0059]
% parents with 4-year degree	-0.0583***	-0.0786***	-0.0455**	-0.0584***
	[0.0173]	[0.0217]	[0.0156]	[0.0188]
Avg. free reading	0.0012	-0.0014	0.0036	-0.0045
	[0.0046]	[0.0056]	[0.0044]	[0.0049]

*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 two-stage quantile regression regression broken out by subgroup, i.e., that depicted in Figures 3 and 4. Marginal effects are averaged over quantiles for the first column.

Table 7: Avg. Characteristics of Durham and Chapel Hill Public Schools (Grades 4 and 5, Academic Year 2002-03)

	Durham	Chapel Hill
Reading	0.3164	1.0393
Avg. white reading	0.8700	1.2870
Avg. nonwhite reading	0.0432	0.1585
% Free/reduced pricelunch	0.4140	0.1476
% Nonwhite	0.6696	0.2195
% NW in avg. white class	0.5090	0.2222
% NW in avg. nonwhite class	0.7563	0.2497
% Parents with HS degree	0.6116	0.1862
% Parents with 4-year degree	0.3606	0.8014
N	4143	1294

Table 8: Merger of Durham and Chapel Hill Districts (5th graders 2002-03, N=2,766)

	Nonwhite	White	Gap	Average
Durham Students				
Observed	0.0623	0.9358	0.8736	0.3390
	[0.0444, 0.0782]	[0.9261, 0.9429]	[0.8525, 0.8908]	[0.3269, 0.3497]
Deseg. within district	0.1064	0.9607	0.8543	0.3771
	[0.0881, 0.1739]	[0.9265, 0.9873]	[0.7840, 0.8756]	[0.3566, 0.4222]
Deseg. across district	0.1603	1.0059	0.8456	0.4282
	[0.1118, 0.2520]	[0.9719, 1.0454]	[0.7552, 0.8999]	[0.3884, 0.4939]
Chapel Hill Students				
Observed	0.1909	1.2968	1.1059	1.0597
	[0.1913, 0.2257]	[1.2833, 1.3126]	[1.0728, 1.1091]	[1.0525, 1.0778]
Deseg. within district	0.2358	1.3189	1.0830	1.0867
	[0.2110, 0.2868]	[1.3011, 1.3506]	[1.0303, 1.1169]	[1.0709, 1.1176]
Deseg. across district	0.2291	1.2914	1.0624	1.0637
	[0.1666, 0.3531]	[1.2510, 1.3476]	[0.9659, 1.1428]	[1.0265, 1.1203]
Overall				
Observed	0.0739	1.0949	0.9296	0.5128
	[0.0585, 0.0912]	[1.0842, 1.1053]	[1.0002, 1.0382]	[0.5054, 0.5231]
Deseg. within district	0.1181	1.1185	0.9094	0.5482
	[0.0998, 0.1838]	[1.0938, 1.1449]	[0.9310, 1.0254]	[0.5288, 0.5862]
Deseg. across district	0.1666	1.1317	0.8978	0.5815
	[0.1188, 0.2612]	[1.0948, 1.1784]	[0.8835, 1.0322]	[0.5446, 0.6482]

95% confidence interval reported in brackets.

Table 9: Durham to Chapel Hill

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

A.3 Figures

Figure 1: Student Accountability in Charlotte-Mecklenburg Schools

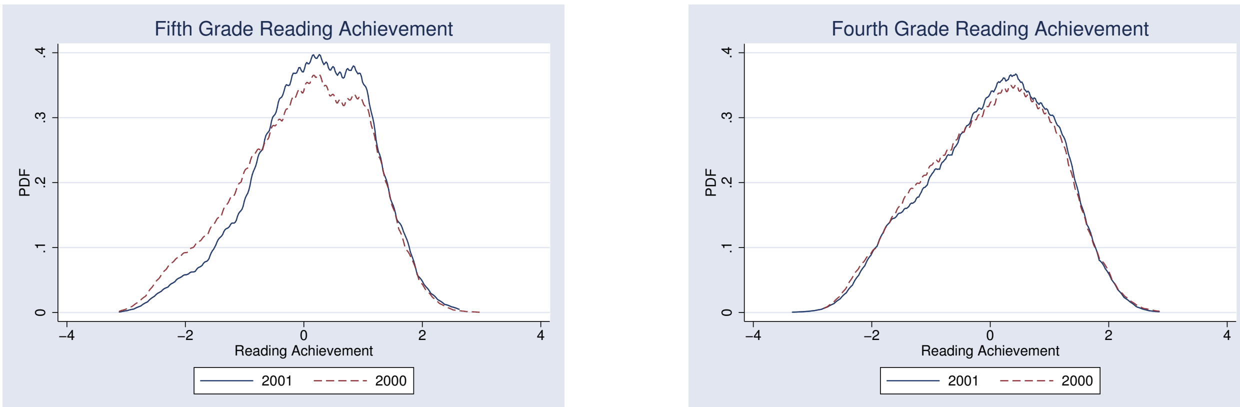
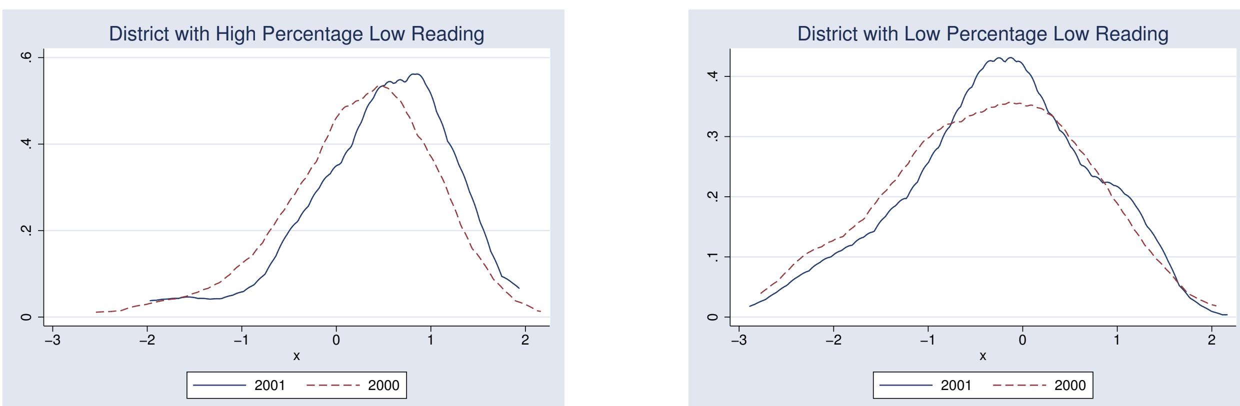


Figure 2: Comparison of Two Districts with Different Concentrations of Low Achievers



Low achiever is defined as a student who performs below the 30^{th} percentile based on prior test scores. In the high percentage district 25% of students were in danger of failing as compared to only 14% in the low percentage district.

Figure 3: Effect of Average Peer Achievement: Two-stage quantile regression

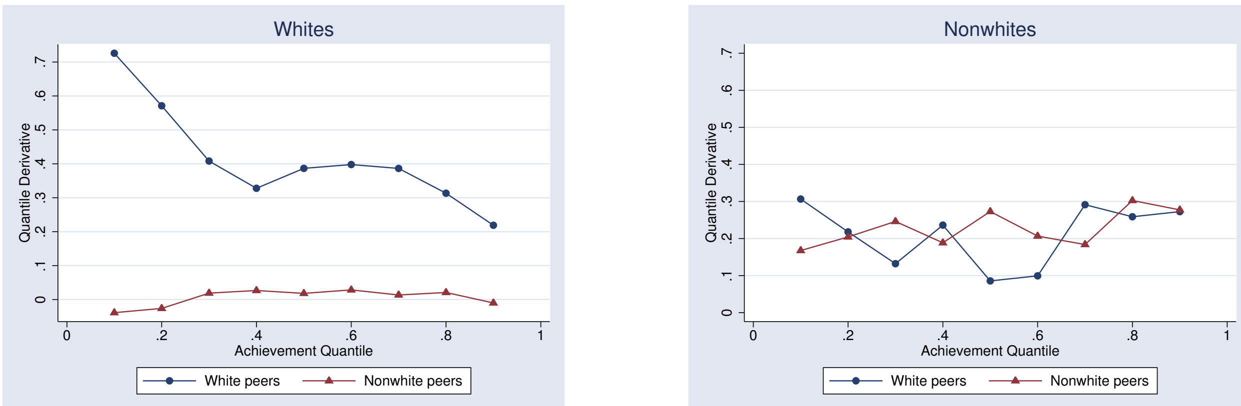


Figure 4: Exogenous Peer Group Composition Effects: % Nonwhite

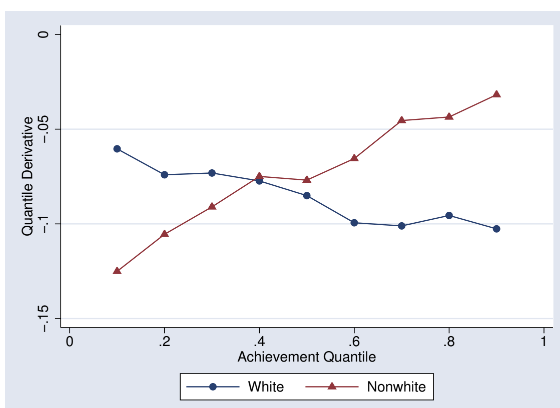


Figure 5: Achievement Gains from Merger of Durham and Chapel Hill

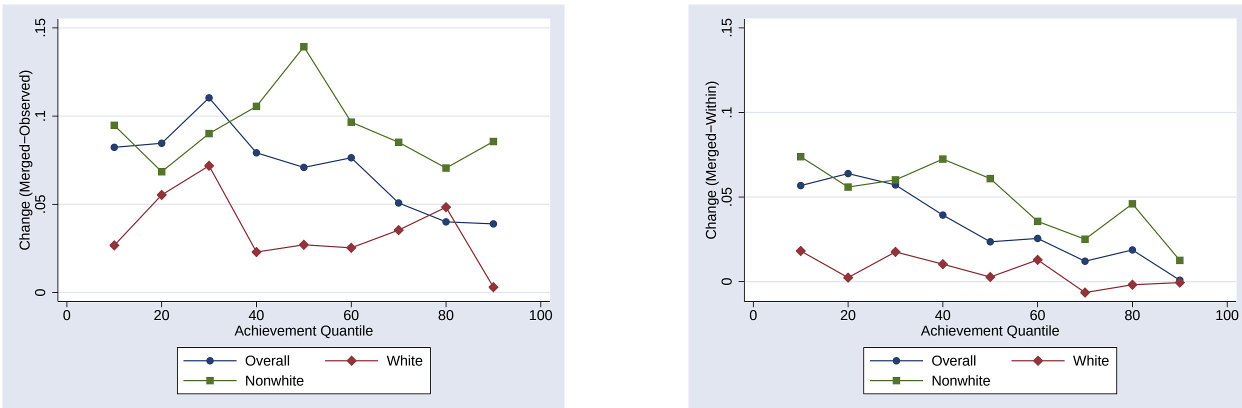


Figure 6: Achievement Gap–Two Stage Quantile Regression

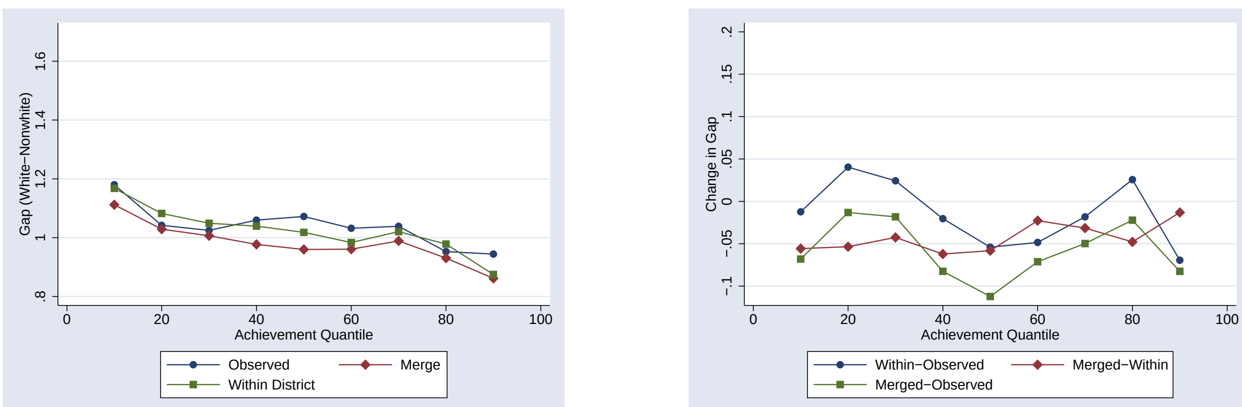


Figure 7: Achievement Gap–2SLS