WCER Working Paper No. 2010-3
February 2010

Desegregation and the Achievement Gap: Do Diverse Peers Help?

Jane Cooley
Department of Economics/
Wisconsin Center for Education Research
University of Wisconsin–Madison
jcooley@ssc.wisc.edu

WCER
Wisconsin Center for Education Research
School of Education • University of Wisconsin–Madison • http://www.wcer.wisc.edu/

WCER working papers are available on the Internet at http://www.wcer.wisc.edu/publications/workingPapers/index.php. Recommended citation:

> Cooley, J. (2010). *Desegregation and the achievement gap: Do diverse peers help?* (WCER Working Paper No. 2010-3). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/workingPapers/papers.php

The research reported in this paper was supported by the Spencer Foundation, the University of Wisconsin Graduate School, and the Institute for Research on Poverty, and by the Wisconsin Center for Education Research, School of Education, University of Wisconsin–Madison. Any opinions, findings, or conclusions expressed in this paper are those of the author and do not necessarily reflect the views of the funding agencies, WCER, or cooperating institutions.

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

Jane Cooley[†]

February 4, 2010

Abstract

Understanding peer effects is critical to evaluating the effect of public school segregation on the achievement gap. This paper develops a new approach to identifying the effect of peer behavior on achievement, using a framework that integrates previously unexplored types of heterogeneity in peer spillovers. Applying the strategy to North Carolina public elementary school students, I find peer effects exist primarily within race-based reference groups, and the magnitude diminishes across the percentiles of the achievement distribution. While on average desegregating peer groups only narrows the achievement gap marginally, this masks important distributional effects, particularly gains for lower-achievers.

Keywords: racial achievement gap; peer effects; desegregation
JEL: I20, I21, J15

*I thank Pat Bajari, Han Hong, Tom Nechyba, Peter Arcidiacono, Arie Beresteanu, Charlie Clotfelter, Steven Durlauf, Stephen Ryan, Karl Scholz, Chris Taber, Chris Timmins, Jake Vigdor and participants at numerous seminars and conferences for their helpful comments. I also thank Caleb White and Jeff Traczynski for research assistance. I am grateful to the Spencer Foundation, University of Wisconsin Graduate School and Institute for Research on Poverty for financial support and the North Carolina Education Research Data Center for providing the data. 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 designed 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 gap remain a puzzle, the persistent segregation of public schools is frequently considered to be a contributing factor. Despite mixed evidence in the literature,[1] policy makers often advocate desegregation as a means of narrowing the gap.[2] The narrowing of the gap could occur through several channels, such as the redistribution of resources or the creation of "better" peer groups. Distinguishing between these channels may be important for policy, particularly given recent high profile Supreme Court cases that have deemed race-based assignment policies unconstitutional.[3] I focus on the latter channel, isolating the effect of racially diverse peer groups on the achievement of white and nonwhite students.

A central contribution of this paper is the identification and estimation of behavioral spillovers of peers (as proxied by peer achievement). This channel has received little attention in the peer effects literature on achievement production.[4] In exploring behavioral spillovers, I incorporate in a production function framework mechanisms highlighted by research as important determinants of student motivation and academic success.[5]

In contrast, previous studies that explore the effect of racial diversity on achievement focus on the separation of an effect of predetermined peer characteristics (such as race and prior achievement) from unobserved shared inputs.[6] The primary challenge for these studies is that students are not randomly assigned to peer groups, so that it is difficult to determine

[1]For instance, see Card and Rothstein (2007), Cook and Evans (2000), Guryan (2004), Hanushek et al. (2009), Hoxby and Weingarth (2005), Jackson (2009) and Rivkin (2000). Vigdor and Ludwig (2008) provide an overview.

[2]For instance, in the recent 2008 Supreme Court cases regarding the constitutionality of race-conscious assignment policies, the Supreme Court took into account evidence on the effect racial diversity on the achievement gap (Linn and Welner, 2007).

[3]See 2007 Supreme Court cases for school integration policies in Seattle, WA and Louisville, KY.

[4]Instead, the literature has focused on spillovers from lagged peer achievement or predetermined characteristics of students, i.e., see Hanushek et al. (2003), Arcidiacono et al. (2009), among others. Cooley (2009) describes why the contrast may be important. See Graham (2008) and Giorgi et al. (2007) for some notable exceptions.

[5]For instance, see Jencks and Mayer (1990), Ogbu (2003), Lazear (2001) and Bishop et al. (2003). The broader social interactions literature has also noted that peers play a role in determining behaviors that may be related to achievement. Cipollone and Rosolia (2007), Gaviria and Raphael (2001), Nakajima (2004), Krauth (2005) find evidence that peers affect the decision to drop out of high school and other behaviors, such as alcohol consumption, smoking, or drug use.

[6]See Hanushek et al. (2009) and Hoxby and Weingarth (2005), among others.

whether observed correlation in achievement derives through peers or through unobserved shared inputs, part of the identification problem highlighted by Manski (1993). However, even under random assignment, the simultaneity problem, whereby an individual student influences his peers' actions and vice versa (and thereby creating *social multiplier* effects), remains. In the achievement context, these social multiplier are capture in spillovers from contemporaneous peer achievement (or *endogenous effects*).

The existing norm in the literature is that reduced form estimates of the effect of predetermined peer characteristics are sufficient for determining the effects of regrouping.[7] However, Cooley (2009) demonstrates that knowledge of the endogenous effect is often necessary if one wishes to infer the effect of regrouping, in this case creating racially diverse classrooms. Intuitively, this follows because of existing matching between students and unobserved resources in the data. For instance, nonwhite students experience lower unobserved teacher quality on average than whites.[8] An ideal counterfactual for determining the effect of racially diverse classrooms might be to compare achievement under the observed groupings to predicted achievement in a racially integrated setting with the teacher quality fixed at the same level for all students in both settings. The key problem is that teacher quality is not observed independently of peer composition in the reduced form model because the effect of teacher quality varies based on the peer composition. I can separate the two effects in my setting by estimating the endogenous effect of peers (and therefore the social multipliers associated with different teacher quality). This allows me to capture the true effect of desegregation.[9]

I begin by constructing a theoretical model of student achievement where students make behavioral choices. Peers can affect these choices either through peer pressure (i.e., with costs associated with deviating from group norms or alternatively discouragement effects) or directly through an achievement production process (i.e., detracting from or enhancing the learning environment). This framework results in an achievement best response function, with observed achievement deriving from student optimizing behavior. This best response is similar in form to the achievement production function with peer spillovers.[10] By situating peer effects in the context of optimizing behavior, the model clarifies the conditions that need to hold to address the simultaneity problem: namely I need to find an exogenous shifter that affects peer achievement through students' choices of "effort" and that does not directly

[7] See Hanushek et al. (2003), Ammermueller and Pischke (2006), among many others.

[8] See, for instance, Clotfelter et al. (2006).

[9] Graham et al. (2008) also provide a unique strategy for directly estimating the effect of desegregation, though in a setting where unobserved teacher quality is plausibly randomly assigned across students.

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

affect a student's own achievement.[11]

The introduction of *student accountability* policies in North Carolina public schools, 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 (and consequently achievement) of those who perceive themselves to be in danger of failing under the new policy. Furthermore, because student accountability only applied to 5^{th} graders, 4^{th} grade classrooms act as a control group. Intuitively, I identify peer spillovers by contrasting classrooms with different portions of students that are *held accountable* to classrooms of similar composition where students are not held accountable. The central identifying assumption is that the percentage of students held accountable is independent of the unobserved group effect. Importantly, if teachers respond to the shift in student effort rather than the policy itself, this does not violate the identifying assumptions. I present evidence that suggests that the alternative hypothesis of a direct teacher response to the policy is unlikely particularly given pre-existing school accountability policies that applied directly to teachers. That said, if anything my estimates of the effect of peer achievement would be biased *downward* if student accountability shifted resources toward low achievers.

My second contribution comes from exploring new sources of heterogeneity in peer spillovers, which are important for determining the effect of racial diversity. In particular, I build a framework that allows peer effects to vary both across the percentiles of the achievement distribution and by race with the potential for students to form race-based reference groups within the classroom.[12] In the absence of race-specific effects creating racially diverse classrooms could help narrow the gap if lower-achieving nonwhite students benefit from being grouped with higher-achieving white peers. Incorporating heterogeneity in this dimension is important as the linear-in-means model, which has received the most attention in the literature, limits equity and efficiency implications of alternative assignment policies in that no matter how students are assigned to peer groups, average achievement remains the same.[13]

[11] The peer effects literature that exploits information about social networks to identify behavioral peer effects (Bramoulle, Djebbari and Fortin, 2009; Giorgi, Pellizzari and Redaelli, 2007; Laschever, 2008) is closest in spirit to this strategy.

[12] I take the classroom to be the peer group of interest both because students in elementary grades spend the majority of the school day with the same classroom peers and classroom assignment is directly manipulable by policy makers.

[13] 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) and convincingly in empirical evidence from Hoxby and Weingarth (2005).

The complication introduced by allowing peer spillovers to vary by the percentiles of the conditional 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 a reasonable informational structure for the model, while still yielding a tractable estimator. I make use of the techniques developed in the recent literature on endogenous quantile instrumental variable models.[14] 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. Hoxby and Weingarth (2005) and Hanushek et al. (2003), among others, also incorporate nonlinearities by allowing the peer effect to vary based on the prior achievement distribution. Unlike the prior studies, however, I capture how responses to peers vary across the percentiles of the *conditional* achievement distribution (that is, after conditioning on observable characteristics). The quantile estimator has a natural interpretation (in terms of unobserved ability) and permits additional flexibility, providing an even richer picture of the benefits and costs of alternative grouping strategies.

The second important source of heterogeneity is in the race dimension. If students respond more to peers of the same race, the potential benefits of desegregation could be muted. For instance, lower-achieving nonwhites might not benefit from higher-achieving white peers. Prior research suggests that this type of heterogeneity may be important. For instance, some studies, such as Fordham and Ogbu (1986), Fryer (2009) and Ogbu (2003), find that whites and blacks value achievement differently. I bring a new perspective to this question by estimating the extent to which students form different race-based reference groups within the classroom.

A central finding of this paper is that peer spillovers are stronger within race than across races. The positive within-race spillovers diminish across the percentiles of the achievement distribution, so that lower-achieving students benefit relatively more than higher achievers from increases in average peer achievement. The spillovers from peer achievement are much larger in magnitude than prior studies that use lagged measures of peer achievement.[15] This is not surprising given that effort spillovers captured by contemporaneous peer achievement entail social multiplier effects, whereas the spillovers deriving through predetermined characteristics as proxied by prior peer achievement do not.

In terms of a direct effect of racial composition, I find evidence consistent with prior

[14] See Abadie et al. (2002), Chernozhukov and Hansen (2005), Chesher (2003), Honoré and Hu (2004), Imbens and Newey (2003), and Ma and Koenker (2006).

[15] For instance, see Hanushek et al. (2009) and Vigdor and Nechyba (2007).

research, such as Hanushek et al. (2009), that achievement is lower in higher percentage nonwhite peer groups. However, the spillovers from contemporaneous achievement of peers of the same race, an aspect not captured in prior studies, play a central role in determining the achievement benefits associated with creating racially diverse classrooms. To attempt to quantify the effect of desegregation, I simulate the consequences of a classroom integration policy for achievement using my peer effect estimates. I find that on average the effects of desegregation are small. However, this masks important distributional effects, most notably the narrowing of the gap at the lower percentiles of the achievement distribution. In contrast, simulations using reduced form estimates of the peer effect (which do not capture the endogenous peer effect) severely misstate the effect of desegregation.

2 Data

I use administrative data for North Carolina public school students from the academic years 1996-07 to 2002-03.[16] I focus on reading test scores.[17] The range of test scores varies considerably across grades and years, as does the cutoff for *achievement level 3*, the level designated "consistent mastery" and the cutoff for passing the exam. Suppose y_{igt} denotes the raw test score for student i in grade g at time t. I normalize scores separately by grade using 1997 scores as a benchmark, with comparisons based on the deviation from the cutoff for achievement level 3 ($y_{gt}^{(3)}$). Formally, The standardized score, Y_{igt}, is constructed as follows:

$$Y_{igt} = \frac{(y_{igt} - y_{gt}^{(3)}) - \frac{1}{N}\sum_i (y_{ig,97} - y_{g,97}^{(3)})}{SD_g(y_{ig,97} - y_{g,97}^{(3)})},$$

where $SD_{g,97}(y_{igt} - y_{g,97}^{(3)})$ denotes the standard deviation for a given grade in 1997.

A unique feature of these data is that each student record is linked to a teacher identification number.[18] This permits the identification of classroom peer groups for grades where student instruction takes place primarily within self-contained classrooms. Thus, I restrict

[16] 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.

[17] Evidence suggests that schools have a larger effect on mathematics achievement, so these results may understate the overall role of peers. (See Rivkin et al. (2005), among others.) However, as I discuss in Section 4, the policy variation used for identification, the introduction of student accountability, coincides with a rescaling of the math test.

[18] In some cases the data center was unable to reliably identify the teacher; these cases are dropped from the analysis.

the analysis to elementary students in grades 3 through 5, where the teacher ID can reliably identify the classroom peer group.[19] 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.

Students remain in the data as long as they attend North Carolina public schools. 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, and parental education. I define *nonwhite* students to be black or Hispanic or American Indian, as these primarily comprise the traditionally disadvantaged racial subgroups in North Carolina; all other students are *white*.[20]

Data on parental education are collected differently across schools. In some cases, particularly in elementary school, the teacher provides a best guess of parental education. To correct for potential measurement error, I assume that parental education is fixed over the period and choose the most frequent report.[21] 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).

In order to estimate race-specific spillovers, I exclude classrooms that do not have at least two students of each race, to at least allow the potential that students can respond to peers of the opposite and same race. 14% of white observations are dropped, as compared to only 7% of nonwhite. However, average achievement is comparable in the restricted sample.[22]

Table 1 reveals well-documented disparities in the background characteristics and achievement of white and nonwhite students in the restricted sample. On average, whites have higher achievement than nonwhites, .48 compared to -.21. They also have better-educated parents. 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 32% of the peers of whites are

[19] 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.

[20] When there are discrepancies in the student's reported race over time, I take the most frequently reported value.

[21] I also try using data from grades 6 to 8, when available, under the assumption that middle schoolers are better able to report parental education. The results are not sensitive to the different specifications.

[22] The comparison sample can be found in Appendix Table 7.

nonwhite, compared to 49% 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 and 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 (prior) peer achievement and peer characteristics along with the typical individual, teacher and resource inputs. The lack of a theoretical basis for the source of these spillovers can obscure the identification challenges and make estimates difficult to interpret and apply to policy questions of interest. In particular, it is not clear whether prior or contemporaneous peer achievement belongs in the achievement production function and this matters for identification (as I show below).

Given this, I construct a model of peer spillovers in the classroom 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. The optimizing framework permits me to incorporate insight from theoretical models of social interactions and evidence about sociological and psychological determinants of student motivation into the achievement production context. 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.

I define a peer group to be a classroom of students in a particular time period.[23] 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, $e_{-i} = (e_1, ..., e_{i-1}, e_{i+1}, ...e_N)$.

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

Furthermore, i's achievement depends on predetermined 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, peers may affect an individual's achievement through their innate characteristics (or *contextual/exogenous effects*), which enter through S_i. Contextual 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.[24]

Finally, a student cannot perfectly predict his achievement on an exam, even after choosing his own effort and observing the effort of his peers. For instance, the student may not fully know his own ability. This is particularly likely for elementary schools students, given their limited experience taking standardized exams. Furthermore, ability is relative, and while a fourth grader may have some knowledge of his standing in the classroom, it would be difficult for him to know how his ability compares to that of students in other schools. Also, unpredictable random factors, such as how well he slept the night before, may affect a student's performance on a given test day. These types of unobservables are captured by θ_i, which can be thought of as an ex post random shock to achievement.[25] 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 affected by predetermined variables S_i. For instance, a student with highly educated parents may face higher expectations regarding academic performance and thereby derive greater utility from high achievement

[24] Lazear (2001) presents such a model where the classroom learning environment is treated as a congestible public good. Figlio (2007) and Kinsler (2006) find empirical evidence of these negative externalities.

[25] This contrasts with the assumption generally made in social interactions models that an individual knows the unobservable at the time of choosing his action. I choose this assumption in part because it seems realistic in this setting where the action is not the outcome being estimated in the data, but also because it maps into the simple two-step estimation strategy pursued in this paper.

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.[26]

The utility function (like the production function) permits both contextual and endogenous peer effects. Peer characteristics may enter through 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.[27] This type of peer spillover has received a great deal of attention in the broader social interactions literature.[28]

The vector of characteristics $S = (S_1, ..., S_N)$ is common knowledge to all students in the classroom, while (θ_i, θ_{-i}) are observed ex post. Students possess a common prior on θ, $f(\theta_i|S)$.[29] 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:

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

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, $e^* \equiv (e_1^*, ..., e_N^*)$, involves everyone playing their best responses.[30]

If effort were observable, the natural object of interest would be the best response to peer effort. As this is not the case, assuming that the achievement production function

[26]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.

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

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

[29]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.

[30]The existence of an equilibrium follows from Brouwer's fixed point theorem, given that $e_i^*(e_{-i}; S)$ is a continuous mapping and on the bounded space $[\underline{e}, \overline{e}]^N$.

is monotonically increasing in effort ensures that the game in effort maps into a game in achievement that is observable in the data. Denote the corresponding achievement equilibrium as $(Y_1^*, ..., Y_N^*)$. Given monotonicity of achievement in effort, such an achievement equilibrium can be described as

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

where $\tilde{Y}_i = \int_\Theta g(e_i, e_{-i}; S_i, \theta_i) f(\theta_i|S) d\theta_i$.[31]

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, peer characteristics and classroom inputs (S_i, S_{-i}) and unobservables (θ_i).

4 Identification

A growing body of research considers the difficulties associated with identifying peer effects.[32] The pioneering work of Manski (1993) presents two negative identification results for the linear-in-means model. First, the effect of peer characteristics (*contextual peer effects*) cannot be separated from the effect of peer behaviors (*endogenous peer effects*). Second, *social effects* (reduced form combination of contextual and endogenous) cannot be separated from unobserved group effects, or *correlated effects*. The literature on peer effects in education typically simplifies the problem by focusing on contextual peer effects, while minimizing the importance of endogenous peer effects in the context of achievement production.[33] In contrast, the model presented in Section 3 illustrates a potentially important role for endogenous peer effects in achievement production deriving through student effort. I discuss in Section 7 why identification of endogenous effects is likely to be important for understanding the effects of regrouping.

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

[31]See Appendix A.3 for details.

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

[33]For example, Ammermueller and Pischke (2006), Hanushek et al. (2003), Betts and Zau (2004), and Vigdor and Nechyba (2007), among many others study the effect of observed peer characteristics: race, income and ability as measured by lagged achievement. A point of confusion is that lagged peer achievement is often referred to as an "endogenous" effect. Here, I use the term to refer to the spillovers from contemporaneous peer achievement.

Durlauf (2001*a*) in the discrete choice setting. While such an argument could be applied by assuming that the peer effects here also take a nonlinear form, the second problem of distinguishing unobserved group effects from social effects remains. Intuitively, the problem is that correlation in student achievement 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.[34]

The reflection problem can be recast as a simultaneity problem. Individual and peer achievement are simultaneously determined in equilibrium and share unobserved common inputs. An exclusion restriction that shifts the achievement of the individual independently of his peers can be exploited to identify a causal effect of peers. While this method has been applied in other areas of the social interactions literature,[35] it has not been applied to the achievement production context for at least two reasons. First, as discussed above, it has not been apparent why simultaneity is important in the achievement context. Second, it is not clear where exclusion restrictions would come from in the achievement context.

In Section 4.1, I describe a potential source of an instrument in the the equilibrium model of achievement and the conditions this instrument must meet to be valid. These assumptions also guarantee identification in the linear-in-means framework. Section 4.2 discusses the particular instrumental variable used in this study. Section 4.3 discusses the implications of nonrandom assignment to peer groups, the aspect of the identification problem that has received the most attention in the literature.

4.1 Simultaneity in Achievement

To discuss identification, it is useful to distinguish between the different components of the characteristics that are observable to the student. 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. P_i captures an education policy that affects i's utility from achievement and is discussed further 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

[34] Graham (2008) 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 rather than fixed.

[35] See Cipollone and Rosolia (2007) for a recent example.

unobserved to the econometrician (i.e., a *correlated effect*). Intuitively, μ can be thought of as the unobserved part of classroom productivity, such as unobserved teacher quality.

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),[36] 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 than is possible in a linear model. 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. 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 characteristics.

The central problem in identifying the structural function (4.1) is that Y_i^* and $\bar{Y}^*_{-i}$ are simultaneously determined in equilibrium. Furthermore, $\bar{Y}^*_{-i}$ is a function of μ, which is unobserved to the researcher. I solve the identification problem using an 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 additional assumptions needed to identify peer spillovers in the context of the equilibrium model of classroom achievement production described in Section 3.

(A1) There exists a variable P_i that affects i's utility from effort (3.2), but does not directly affect achievement production (3.1).

(A2) Conditional on $(X_i, \bar{X}_{-i}, K, \mu)$, θ_i is independent of $(P_i,\ \bar{P}_{-i})$.

[36] 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.

(A3) Conditional on $(X_i, \bar{X}_{-i}, K)$, μ, θ_i are jointly independent of $\bar{P}_{-i}$.

(A4) With probability one, $h(X_i, \bar{X}_{-i}, P_i, \bar{P}_{-i}, K, \mu)$ is strictly monotonic in μ.

(A5) Conditional on $(X_i, \bar{X}_{-i}, K)$, θ_i is independent of μ.

(A1) and (A2) 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 (A1) 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.

North Carolina's student accountability policies, which were enacted for fifth graders in the 2000/01 academic year, provide a potential shifter. They require that fifth graders perform above a certain level on standardized End of Grade (EOG) exams (achievement level 3) in order to be automatically promoted to the next grade. To have any identifying power, the policy must also have different effects on students within the same peer group. I expect students who performed below (or close to) achievement level 3 in the year before the standards were put in place 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, being confident that they would meet the cut-off for passing even with minimal effort.[37]

The independence of θ_i and $P_i, \bar{P}_{-i}$ (A2) 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 low-achieving students who are in danger of being retained under the policy (fifth graders beginning in 2000/01) draw from the same distribution of θ as similarly low-achieving students in similar peer groups for whom the policy does not apply (fifth graders before 2000/01 and fourth graders in all years).

Given (A1) and (A2), the system of response functions for students $i = 1, ..., N$ in a given

[37] While student accountability applies both to math and reading, I focus on reading scores in part because, as mentioned above, a new math test was introduced in 2001 at the same time as accountability.

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 predetermined variables that are common knowledge to all students in the peer group, including μ, which is unobservable to the researcher. 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 (A3). 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, (A3) 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 to these new standards (for instance, by shifting resources to low achievers) in a way that is proportional to the percentage of low achievers.[38] One reason I expect that teacher incentives are less affected is because student accountability

[38]Note that, as I discuss below, this is not a problem if the teacher response is to student effort rather than the policy per se, as the former can be interpreted as a peer effect.

was preceded by school accountability in 1996/97. Hence, teachers already had fairly strong incentives to focus on low achievers. I discuss this assumption further and provide more details on student accountability 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 (A4). 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 $\mathcal{U}(0,1)$. Then, given (A3) and (A4), μ can be recovered from the first-stage regression as shown in Imbens and Newey (2003, Theorem 1) as $F_{\bar{Y}^*_{-i}|X_i,\bar{X}_{-i},P_i,\bar{P}_{-i}}(\bar{Y}^*_{-i}|X_i,\bar{X}_{-i},P_i,\bar{P}_{-i}) = \mu$.[39]

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, (A5), 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 characteristic μ that is unobserved to the researcher but observed to the students is independent of the individual shock, which is fairly straightforward given that θ_i is realized ex post.

Under (A5), 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 characteristics $(X_i, \bar{X}_{-i}, K, P_i)$ and the common component μ. Given (A2), (A4), and (A5), $q(\bar{Y}^*_{-i}, X_i, \bar{X}_{-i}, P_i, K, \mu, \theta_i)$ is then identified on the joint support of $(\bar{Y}^*_{-i}, X_i, \bar{X}_{-i}, P_i, K, \mu, \theta_i)$.[40] Intuitively, conditioning on the unobserved group effect μ controls for the endogeneity of peer achievement, thus identifying the structural function.

4.2 Student Accountability

In this section, I provide a little more background on North Carolina's student accountability policy and discuss the conditions for it to be a valid instrument. As mentioned above, 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 in all years effectively act

[39] See Appendix A.3 for details.

[40] See Appendix A.3 for details. The proof of this result follows from Imbens and Newey (2003, Corollary 6).

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 fail in a given year. I find that the retention rate for fifth graders increased by 50% after student accountability policies were enacted (from 0.010 to 0.015). Over the same period, retention of fourth graders only increased by 7% (from 0.015 to 0.016). The relatively small increase in retention (particularly taking into account the percentage not meeting the standard, as many as 24% in a given year) can be explained because students who do not meet the standard are not automatically retained but instead are required to take summer school or receive extra tutoring. For the purposes of satisfying assumption (A1), what is important is that the threat of retention and the alternatives of summer school or additional tutoring all serve to potentially motivate students to work harder in the classroom (and any additional inputs occur outside the classroom, as discussed further below). The additional help may even better equip them to succeed.

As mentioned before, an 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.[41] 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

[41] 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.

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 (A3) may not hold if teachers or administrators redistribute resources to low achievers as a result of student accountability in a way that is proportional to the number of low achievers. This does not rule out that teachers responded to student accountability. In estimation, I control for a direct effect by including a quantile-specific shift in the achievement of students after student accountability, which would capture shifts like the teacher targeting material towards low achievers and away from high achievers.[42] Because it is quantile-specific, it would also capture the potential for a change in curriculum that results from student accountability to have differential effects across the percentiles of the conditional achievement distribution.

While I am not aware of any studies on student accountability policies in themselves, in part because these policies generally do not exist in isolation of school accountability, previous studies on school accountability show that teachers are very responsive and find evidence that achievement of marginal and/or lower-achieving students increases as a result.[43] One reason I think that this is less likely to be a concern in my setting is that teachers and schools already had strong incentives to shift attention to low achievers well before the introduction of student accountability. Under the School Based Management and Accountability Program of 1996, bonuses for schools and teachers were awarded based on growth scores and the criteria that not too many students perform below achievement level 3 on the standardized EOG exams.

That said, the student accountability policy included several direct effects on student achievement, such as the provision of extra tutoring and summer school. For the purposes of identification, it is sufficient that the direct effect of any change in inputs to students below the threshold occurs outside the classroom. In estimation, I include school by year fixed effects, which would control for any changes in allocations of funds at the school level that coincide with student accountability. The policy should also not directly affect the teacher's allocation of effort within the classroom, at least not in proportion to the percentage of each achievement type.

However, this does not prohibit an *indirect* response of teachers to the policy. In other

[42] 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.

[43] See, for instance, Jacob (2005), Neal and Schanzenbach (forthcoming), Reback (2008).

words, the instrument is still valid if the teacher changes allocation of effort across students in response to changes in student effort (which may have occurred as a result of the policy). Part of the difficulty in supporting this assumption is the lack of a model that endogenizes the response of teachers to student effort and correspondingly the composition of the classroom. This is beyond the scope of the current paper. If a teacher changes how she teaches in response to a different composition of students or a different relative level of effort of students in the classroom (i.e., the teacher may just spend more time with students who are more engaged), I attribute this to a peer effect. The reason this is intuitively appealing is that the estimates would then be applicable to determining the effects of alternative grouping strategies, particularly desegregation in this paper. In other words, the effect of the regrouping results from peer effects deriving both through peer effort and characteristics as well as changes in teacher effort in response to the classroom composition.

I discuss further the potential for direct teacher responses to the policy in the context of my results in Section 6.1. Importantly, if a shift as a result of student accountability occurs in the same way as predicted by studies on school accountability (such as in Jacob (2005), Neal and Schanzenbach (forthcoming), Reback (2008)), this would suggest that my estimates of the effects of peers on achievement are actually biased *downward*.

4.3 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. To control for time-varying selection into schools, I include school-by-year fixed effects in the form of a location-specific shift (described formally in equation (5.2)), i.e., permitting the fixed effects to have different effects across the percentiles of the conditional achievement distribution. The resulting identifying variation derives from plausibly exogenous cross-cohort variation in composition, a strategy pursued by Hanushek et al. (2003), Lavy and Schlosser (2007), among others. However, unlike these studies which consider grade-level peer groups, the focus on class-level peer groups may raise additional concern about non-random assignment to classrooms within schools.

Policies for classroom placement differ widely across schools. 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 higher-achieving children. By selecting better teachers, they effectively select into better peer groups. This might lead one to mistakenly conclude that positive peer spillovers exist, when the positive correlation in outcomes arises instead from selection. Furthermore, school administrators may not assign students randomly to classrooms, but may instead employ some sort of ability tracking.

The instrumental variable strategy pursued in this paper obtains consistent estimates of the endogenous peer effect, as long as students are not reassigned to classrooms as a result of student accountability. In this case, the pre-accountability fifth grade classrooms and fourth grade classrooms of similar composition act as controls for any existing matching between teachers and students. Using a simple difference-in-difference strategy, I find support for this, i.e., that the concentration of low-performing students in a given classroom did not change within a school after student accountability was introduced for fifth graders relative to the composition for fourth graders in the same years.

The primary remaining challenge resulting from non-random assignment to classrooms within a given school/year is the identification of contextual peer effects. In Section 6.1, I provide evidence that estimates of contextual effects are unlikely to be biased by nonrandom assignment to classrooms.

5 Estimation

Estimation of the quantile structural function, the best response of students to peer achievement, proceeds in the two steps described in detail in Sections 5.1 and 5.2. First, I recover the residual from equation (4.2), the first stage regression predicting the ex ante expected value of peer achievement. This residual captures the unobserved group effect or classroom productivity. I then estimate the quantile structural function defined in equation (4.1), controlling for the endogeneity of peer achievement by conditioning on the first stage residual. Note that if the second stage were linear-in-means, the control function approach would be equivalent to the two-stage least squares estimator, where the fitted value rather than the residual from the first stage is plugged into the second stage. I pursue the control function approach because it is consistent with the informational assumptions of the model, i.e., where students observe something about the classroom (teacher) that is unobserved to the researcher when choosing effort.

5.1 First Stage

The first stage of the estimator is analogous to the first stage in 2SLS. Suppose time is indexed $t = 1, ..., T$ and classrooms $c = 1, ..., C$. 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 achievement of classroom peers of a given race k is approximated as

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

where dependence of the parameters on the each race subgroup k, k' is suppressed.

The covariates X_{it} include the sex of the student, parental education, indicators for students who performed below the cutoff for passing (achievement level 1 or 2) in the prior year and students who performed at achievement level 3 in the prior year.[44] Achievement level 4 designates "superior mastery" of the material and is the excluded category.

P_{it} is the utility shifter; it indicates students for whom student accountability policies are binding, i.e., 5^{th} graders in 2001 or later who performed below achievement level 3 or at level 3 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 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, and the percentage of nonwhite students in the classroom, and the percentage of peers who are below passing and those at achievement level 3. I also include interactions of the percentage of peers white/nonwhite who were below passing in the prior year and those at achievement level 3 with 5^{th} grade and post-2001 among the $\bar{X}_{-ict}$. This allows for a different effect of the composition of low-achievers and marginal students before and after student accountability and for the possibility that the composition of low-achievers has a different effect in 5^{th} grade independent of student accountability. Thus, the identifying variation for the endogenous peer effect comes from comparing 4^{th} and 5^{th} grade classrooms with similar compositions of low-achievers pre- and post-2001, when student accountability was introduced for 5^{th} graders.

Other than school by year fixed effects, $SchYr_{it}$, classroom level-inputs K_{ct} include whether a teacher has an advanced degree (beyond a bachelors), a quadratic in teacher

[44] I group achievement levels 1 and 2 because preliminary regressions suggested that the students at these two achievement levels responded similarly to accountability.

experience, an indicator for years/grades when student accountability policies are in place (i.e., 5^{th} grade in 2001 and beyond) and a dummy for 5^{th} grade.

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 relatively small.

I estimate the two first-stage regressions for white and nonwhite peer achievement separately for students of each race.[45] 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 school by year fixed effects, $Sc\hat{h}Yr_{it}$.

5.2 Quantile Structural Function

In the second stage, I estimate the structural function (4.1), which describes a student's achievement as a function of peer characteristics and peer achievement at different points of the conditional achievement distribution. 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 peer characteristics.[46] 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. While certainly there are advantages to considering how responses vary by observed predetermined student characteristics, the quantile approach is appealing because it offers considerable flexibility, can be estimated for a large number of quantiles,[47] and is not sensitive to outliers.

[45] 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.

[46] See Hanushek et al. (2009), Hanushek et al. (2003) and Hoxby and Weingarth (2005).

[47] This is emphasized in Chernozhukov and Hansen (2005).

While it is feasible to estimate the quantile structural function without assuming a parametric form,[48] I assume a parametric approximation for the system of equations because of the large number of covariates. Therefore, I approximate (4.1) as

$$Y^*_{ict} = \beta_0 + \beta_1 \bar{Y}^W_{-ict} + \beta_2 \bar{Y}^{NW}_{-ict} + X_{it}\beta_3 + \bar{X}_{-ict}\beta_4 + \beta_5 P_{it} \\ + K_{ct}\beta_6 + \beta_7 Sc\hat{h}Yr^W_{it} + \beta_8 Sc\hat{h}Yr^{NW}_{it} + \beta_9 \hat{\mu}^W_{ct} + \beta_{10}\hat{\mu}^{NW}_{ct} + u_{ict}, \quad (5.2)$$

where dependence of the parameters on the quantile ($\beta(\theta_i)$) and race is suppressed to simplify notation. School by year fixed effects are allowed to vary by race, capturing the fact that the effectiveness of the school may vary across races.[49] The $\hat{\mu}_{ct}$'s capture the unobserved classroom productivity, that simultaneously affects the achievement of an individual and his peers. These also enter achievement in a flexible way, with the marginal effect permitted to vary both by race and quantile.[50]

It is worth noting that this approximation of the achievement best-response predicts a unique equilibrium. While the mains results focus on the effects of the mean of peer achievement (as is most closely related to the peer effects literature), I also consider variants of the estimator where other moments (rather than the mean) of the peer achievement distribution are included in the achievement best response. For each subgroup and a given quantile $\theta_{it} = \tau$, I estimate $\vec{\hat{\beta}}(\tau)$ using a quantile regression that minimizes the sum of the weighted absolute value of residuals.

6 Results

The first results I discuss below consider whether students respond differently to peers of different races, focusing on the mean and median case of the estimator described in the previous section. Then, I proceed to consider distributional effects, i.e., how the marginal effect of peer achievement varies across percentiles of the conditional achievement distribution.

The first stage results are presented in Table 2. Classes with a larger percentage of peers who are below the threshold for passing in the prior year or just above witness a larger shift in average peer achievement when student accountability policies are introduced. The shifts are

[48] See Imbens and Newey (2003) for a discussion of the fully nonparametric estimator.

[49] This would also help control for potential discrimination at the school level.

[50] 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 response to $\bar{Y}_{-ict}$ not $\hat{\bar{Y}}_{-ict}$.

largest in classrooms with more peers below the threshold. For instance, the percent of white peers below the threshold for passing with student accountability shifts average white peer achievement by 0.25 of a standard deviation for whites and 0.16 for nonwhites. The percent of nonwhite students below the threshold for passing shifts nonwhite peer achievement by 0.10 and 0.15 for whites and nonwhites, respectively, with student accountability. Having more students just above the threshold for passing also shifts average peer achievement. Not only does the instrument satisfy conditions on joint significance so that it is not a weak instrument, it further passes the test of overidentifying restrictions in the simple two stage least squares case.[51]

The shifts associated with student accountability in the first stage peer achievement regression are mirrored at the individual level in the second stage results. Table 3 presents estimates using a mean regression in the second stage and the median case of the two-stage quantile regression described in Section 5. Each column corresponds to a separate regression for students of a given race. Student accountability has about twice as large an effect on students below the threshold for passing as those just above for the mean case (0.16 relative to 0.07 for whites and 0.10 relative to 0.05 for nonwhites).[52] For the median case, the relative effect of student accountability on students below the threshold is even larger, 0.20 related to 0.06 for whites and 0.16 relative to 0.02 for nonwhites.

Both the 2SLS and median two-stage quantile estimators predict that white students receive positive achievement spillovers from their white peers of 0.50, but spillovers from their nonwhite peers are much smaller in magnitude, -0.002 and -0.05, and not statistically significantly different from 0. Similarly, nonwhite students receive large spillovers from their nonwhite peers, 0.64 for the mean and 0.54 for the median. Spillovers from their white peers are smaller in magnitude, 0.16 and 0.17 for the mean and median case, and not statistically significantly different from 0. Thus, it appears that in terms of achievement, white students derive spillovers almost entirely from their white peers and similarly for nonwhites. In the model in Section 3 I posited that peer achievement spillovers might derive through some combination of direct spillovers from peer effort in achievement production and/or indirect spillovers in utility, a conformity type effects. While I do not attempt to distinguish between

[51] Note that to test this I use a 2SLS regression (the mean version of this estimator), where I restrict the school-by-year fixed effects to be the same. Allowing for the two types of fixed effects from the separate first stage regressions (the strategy pursued throughout) has very little effect on parameter estimates.

[52] A previous version permits a separate effect for students at achievement levels 1 and 2 to see whether there was evidence that those at achievement level 1 became discouraged as a result of the new policy or simply did not try harder because there was no hope of passing. However, the shift was not statistically significantly different for the two types.

the two mechanisms, the finding that spillovers derive primarily through same-race peers may be more consistent with the conformity mechanism.

An effect of desegregation may derive not only through peer achievement but also through a shift in peer characteristics. Table 3 shows that a higher percentage of nonwhite students negatively affects white and nonwhite achievement, though the effect for nonwhites is about twice as large (-0.09 compared to -0.05 in the mean case, -0.11 compared to -.06 for the median case). This finding is consistent with prior results in the literature, such as Hanushek et al. (2009), Vigdor and Nechyba (2007), among others. It is worth noting that as no income controls are included, the effect of the higher concentration of nonwhites may also be proxying for an income effect. In contrast to prior research, I do not find that peer parental education has much of a direct effect on white achievement, though it has considerably large negative effects on nonwhite achievement. A higher percentage nonwhite peers who are low-achievers (achievement levels 1,2 or 3) helps the performance of nonwhite students, and similarly a higher percentage of white peers who are low-achievers helps white students. These contextual effects are apparently counterintuitive, but as discussed briefly in Appendix A.4 and expanded in Cooley (2009), the model predicts that the sign of contextual peer effects is actually ambiguous after conditioning on peer achievement. Intuitively, this follows when spillovers derive through unobservable characteristics. For instance, after conditioning on peer achievement, a higher level of peer parental education would predict a lower level of peer effort.

Estimates of the effect of individual characteristics are not included in the table, but are consistent with intuition and prior research. Estimates of the effect of measurable teacher quality show that white students benefit from teacher experience at a diminishing rate, but that teacher experience does not have a statistically significant effect on nonwhites. In contrast, having a teacher with an advanced degree has no effect on whites, while there is weak evidence that it hurts the performance of nonwhites.

As discussed above, heterogeneity in peer effects across the achievement distribution may also play an important role in determining an effect of racial desegregation, particularly given that nonwhites are more heavily concentrated in the lower tails of the achievement distribution. While Table 3 presents estimates for the median student, 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 findings at other percentiles of the achievement distribution also support the median finding of a lack of cross-racial spillovers. The spillovers from peers of the same race is largest for the students at the lower quantiles and roughly diminishes across quantiles. The positive effect of white peers on whites diminishes from a high of close to 1 to a low of .3 for students in the upper quantiles and rises slightly for the highest quantile to 0.6. The positive effect of nonwhite peers on nonwhites diminishes from a high of close to 1.2 to a low of .45 for students in the middle and rises slightly for students in the upper quantiles up to 0.6.

To provide some insight into magnitudes, Tables 4 and 5 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 effect of a one standard deviation increase in white peer achievement is 0.22 for whites and 0.07 for nonwhites on average, while the marginal of a one standard deviation increase in nonwhite peer achievement is 0.01 for whites and 0.28 for nonwhites. Overall, the effects of white peers on whites is smaller in magnitude than the effect of nonwhite peers on nonwhites. This table further emphasizes that the effect of same-race peers for students at the median or above is about half the magnitude as the effect of same-race peers for the lowest-achieving students.

The effects of the achievement of same race-peers are larger in magnitude than previous estimates in the literature using lagged peer achievement, suggesting that failure to consider contemporaneous spillovers may severely understate the effect of peers, particularly for the lowest quantiles of students. In fact, these estimates are comparable in magnitude to some of the more important determinants of student achievement found in the literature, such as teacher quality and class size. For instance, Rivkin et al. (2005) report that a one standard deviation increase in teacher quality leads to approximately 0.095 of a standard deviation increase in reading. Using findings from Project Star, they report that this change is comparable to a reduction in class size of 10 students in fourth grade and 13 in fifth grade. The positive effect of same-race peers at the median and above is slightly double the magnitude of this effect, while the effect same-race peers on the lowest achieving students is 0.40 for whites and 0.49 for nonwhites are 4 to 5 times the magnitude.

Tables 4 and 5 also reveal that increasing the percentage of nonwhite peers has a small negative effect on nonwhites of -0.02 at the median and upper quantiles. The effect is not statistically significantly different from 0 for nonwhites at the lowest quantiles of for whites at any of the quantiles. In all cases, the effect of a 1 standard deviation increase in percentage nonwhite is much smaller in magnitude than a 1 standard deviation increase in average peer achievement of the same race. The effect of racial composition is further illustrated in Figure 4, which compares the quantile derivatives of the percentage of peers who are nonwhite on white and nonwhite achievement. The negative effect of percentage nonwhite on nonwhites and whites is roughly diminishing across quantiles, with the lowest effects for nonwhites in the middle of the distribution.

One potential interpretation of the stronger within-race spillovers is that students are simply responding more to peers who are more similar in other dimensions, such as ability. This could be reflected in the above regressions because, as shown in Table 1, average nonwhite achievement is much lower than average white achievement. Thus, the above results could just be reflecting the fact that lower-achieving students, as proxied by nonwhite, respond more to lower-achieving peers and vice versa for whites. To test this, I estimate the effect of different quantiles of overall classroom peer achievement on students at different quantiles of the achievement distribution.[53] If the above intuition holds, we would expect to find that the lowest quantile of the achievement distribution responds most to lower quantiles of peer achievement, the median to the median of peer achievement, and the upper quantile to the upper quantile of peer achievement. This is not the case. I find that the lower-achieving students always benefit more (relative to students at higher quantiles) to the 25^{th}, median and 75^{th} percentiles of classroom peers. While students at the median benefit most from increases in median achievement, students at the upper quantile are not affected by increases in the upper quantile of peer achievement but only by the median.

To test this further and make sure that the within-race spillovers are not an artifact of focusing on mean peer achievement of each race, I also estimate a specification that replaces the mean peer achievement of each race with the 25^{th}, median and 75^{th} percentiles respectively, allowing the marginal effects to vary by race (comparable to the estimation in Section 5, but using different moments of the peer achievement distribution). I find that spillovers in this specification remain entirely within race, with spillovers from students from the opposite race being insignificantly different from 0. This specification also does not show a pattern that suggests that students respond most to increases in the achievement of students of the same "ability" even within race. For instance, the lower-achieving students benefit most from increases in the 75^{th} percentile of the achievement distribution of peers of the same race. Thus, these alternative specifications suggest that the within-race spillovers cannot be attributed to "ability" similarities as measured through achievement.

[53] It is worth noting that the instruments, percentage of students below the threshold and just above interacted with student accountability, are significant predictors of the different quantiles of peer achievement, even the 75^{th} percentile.

Evidence in Fryer (2009) and elsewhere suggests that peer effects for nonwhites varies based on the percentage of nonwhite students in the classroom. Thus, I further test whether results vary based on the racial composition of the classroom. In particular, I estimate the mean regression in Table 3 for classrooms that have more than the mean percentage nonwhite (approximately 0.37) and classrooms with less than the mean. Not surprisingly, splitting the sample creates noisier estimates. However, I still do not find any evidence of cross-racial spillovers in classes with larger percentages or smaller percentages of nonwhite students.

6.1 Sensitivity Analysis

Instrumental Variable. As discussed in Section 4.2, an alternative interpretation of my results is that the instrument, student accountability policies, is really capturing a response of teachers and not a response of students and therefore is not a valid instrument. Importantly, this would suggest that my peer effect estimates are actually biased *downward* if teachers respond in ways predicted by studies on school accountability.[54] Recall that because I am controlling for a direct effect of student accountability, the identifying assumption is violated when the shift in teacher effort is in proportion to the percentage of low achievement and is a direct response to the policy (rather than a indirect response to the change in student effort as a response to the policy).

It is difficult to reconcile my findings of peer spillovers primarily within racial groups with the alternative that student accountability causes teachers to shift attention to low achievers in proportion to the percentage of low achievers in the class. If the teacher decided to teach more to the lower end of the distribution to ensure that these students were not retained, this effect would more likely be shared by all students in the classroom regardless of race.

That said, I also consider evidence of direct responses of teachers to the policies. For instance, if teachers are under additional pressure to respond to student accountability, we might observe increased turnover of fifth grade teachers relative to fourth grade teachers. I do not find evidence of this. Another way that teachers may respond to the increased pressure is by switching to teaching fourth graders who are not held to the student accountability standards. I do not find evidence of this. As mentioned above, I also check for principal responses in how students are allocated to classrooms to see whether the low-types are reshuffled in fifth grade relative to fourth grade. Again, this does not appear to be the case.

Given that teachers and schools are forward looking, if the response derives through teachers rather than students we might expect these types of resource shifts to occur at

[54] See Jacob (2005), Neal and Schanzenbach (forthcoming), and Reback (2008).

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. Furthermore, if teachers are forward looking and shift attention in fourth grade and fifth grade in the same way, this actually helps my identification strategy, as I control for a different effect of percentage low-achievers in 2001 and later (when student accountability applies to 5^{th} graders). Recall that the identifying variation comes from the additional effect of percentage low-achievers in 5^{th} grade after accountability is in place, when students are officially facing the threat of failure.

If the effect of student accountability derives primarily through a teacher reallocation of effort among students, low-achievers in smaller classrooms might have larger increases in achievement because the teacher has more time to allocate to them on a per student basis. Performing a simple difference-in-difference estimation of the effect of student accountability across large and small classrooms, this does not appear to be the case. In fact, if anything the achievement for students below the cutoff for passing actually increases more after student accountability for students in *larger* classrooms.

I test for potential direct responses of students by considering whether student's leisure time allocation changes after student accountability. Using the same difference-in-difference type strategy (where I compare fifth graders after student accountability using as controls fourth graders and fifth graders prior to accountability), I find that time spent reading for fun does not change but hours watching television decreases for the students below the threshold who are subject to accountability. While not conclusive, this finding is consistent with the hypothesis that students are reallocating effort toward activities conducive to achievement after student accountability policies are in place.

Contextual Effects. Though school-by-year fixed effects control for arguably the most salient form of selection into peer groups, nonrandom assignment to classrooms within schools is a remaining potential source of bias in contextual peer effects.[55] To consider potential endogeneity of contextual effects, I reestimate Table 3's mean regression using teacher and year fixed effects (rather than school-by-year fixed effects). I find that results are qualitatively similar, though standard errors are much larger. In particular, the point estimates of the contextual effects do not appear to be statistically different, though in some cases they lose their statistical significance. I choose not to pursue the strategy of using teacher fixed effects, out of concern that it is not a sufficiently long panel of teachers (a fact that would explain

[55] Recall that the instrumental variable strategy alleviates this concern for endogenous effects.

the large standard errors). Furthermore, Jackson (2009) finds evidence that teachers move in response to changes in student composition of schools, suggesting that school by year fixed effects may be a better control for selection into schools (and peer groups) than a time invariant teacher fixed effect.

To further explore potential bias in my estimates of the contextual effects due to nonrandom assignment to classrooms within schools, I recover a subset of school-years where the students appear to be randomly assigned across classrooms based on observable characteristics.[56] Formally, I calculate a joint test of whether the classroom composition is significantly different from the school-grade composition in terms of observable characteristics—percentage male, nonwhite, parental education, and prior achievement level. I designate schools as apparently randomly assigning students to classrooms if the p-value is greater than .1 or schools have only 1 classroom per grade. This is about 72% of the schools in my sample.[57]

This subset of schools is remarkably similar in terms of observables to the main sample, with the biggest difference that the main sample has 26% of parents with a 4-year degree, and the apparent random assignment schools only have 24%. Apparent random assignment schools have a slightly lower percentage nonwhite, 36.7% compared to 37.5% in the main sample. This is consistent with the intuition that policies for classroom assignment vary widely across schools and suggests that, at least in terms of observables, schools that do not randomly assign students across classrooms are on average not very different than schools that do. Estimates of the peer effects in the mean and median regression in Table 3 on the apparent random assignment sample of schools are both qualitatively and quantitatively similar to the estimates on the main sample.[58]

The above tests provide supportive evidence that school-by-year fixed effects along with the instrumental variables strategy are adequately controlling for nonrandom assignment to peer groups and that selection to classrooms within schools is not biasing estimates of contextual effects. It is worth noting that Clotfelter et al. (2003) also find that the majority of segregation in North Carolina public schools in elementary grades comes from across-school segregation and find very little evidence of within school segregation in these early grades.

[56] Vigdor and Nechyba (2004) use a similar intuition, as do Lavy and Schlosser (2007) in their balancing tests.

[57] See Appendix Table 8 for comparison of schools.

[58] Comparison tables available upon request.

Value-added. To help control for unobservable student characteristics and prior inputs, studies often estimate the peer effects using a value-added specification, i.e., controlling for a student's prior achievement.[59] I do not pursue this strategy in the quantile model because I want to estimate how peer spillovers vary across percentiles of the conditional achievement distribution, a case in which the residual could arguably be interpreted as unobserved ability. It is less clear how to interpret the residual in the value-added model and correspondingly heterogeneity in spillovers across the percentiles of the conditional value-added distribution. That said, when I control for lagged achievement in the linear-in-means case I find that estimates of the peer effects do not change.

7 Desegregating Peer Groups

While the estimates above show that peer effects are quite important determinants of student achievement, it remains difficult to infer an actual effect of desegregation directly from the parameter estimates. Intuitively, my findings suggest that the lack of cross-racial spillovers would certainly limit the benefits of desegregation. Nonwhites would experience gains in achievement if grouped with higher achieving nonwhite peers on average, and whites would experience losses if grouped with lower-achieving white peers on average. Furthermore, given that lower-achieving students benefit relatively more than higher-achieving students, we might expect efficiency gains in terms of increases in average achievement, to the extent that desegregation also creates more mixed-ability classrooms. The magnitude of the gains are difficult to determine because of the social multipliers created. Furthermore, even though much smaller in magnitude as demonstrated in Tables 4 and 5, the characteristics of the racially-diverse peers would also affect student achievement. The finding that a higher concentration of nonwhites is negatively correlated with achievement suggests that creating more diverse peer groups would raise nonwhite achievement while having little effect on white achievement. Disparities in parental education and prior achievement levels also enter into the overall effect of any student reassignment policy.

Often peer effect studies attempt to infer an effect of desegregation from the reduced-form effects of percentage nonwhite holding other observable peer characteristics fixed. These strategies do not take into account the joint distribution of student characteristics. In other words, it would be impossible to only change the racial composition of classrooms and hold other peer characteristics, such as parental education, fixed when parental education is corre-

[59]See Hanushek et al. (2003) for a detailed discussion of this approach.

lated with race.[60] The challenge is similar in my context. However, in my model the potential effect of desegregation is even more complicated because it takes into account spillovers from peer achievement as well as peer characteristics. Thus, to quantify the total effect of desegregation, I use the parameter estimates from the best response functions estimated above to simulate the new equilibrium achievement that would result from the reassignment to racially diverse classrooms, thus taking into account the full joint distribution of student characteristics.

It is important to note that the experiment abstracts away from issues of residential sorting, proximity constraints and the potential to select out of public schools, arguably providing an upper bound on the benefits of desegregation. While this limitation is shared by other peer effect studies in the literature, but should be kept in mind when interpreting the findings.

As discussed above, a common assumption in the peer effects literature is that reduced form estimates (those that estimate the effect of predetermined peer characteristics on achievement or the *social effect* of peers) are sufficient to determine the effects of regrouping policies. However, Cooley (2009) points out that this may not be the case in many settings, like the current one. Though in principle a reduced form specification of the overall effect of percentage nonwhite students on achievement could be estimated quite flexibly, it would not be feasible to infer from this specification the effect of desegregating peer groups in observational data because of matching between teachers and students across schools.[61]

Intuitively, this follows because in the perfectly integrated system, holding resources fixed at some level means that nonwhite (white) students might receive higher (lower) resources on average than in the initial observed assignment. If this reallocation of resources creates social multiplier effects (that vary based on the composition of the classroom), it is not possible to separate an effect of racial integration from a resource effect without estimates of the social multiplier (or the endogenous effect).[62] Importantly, this follows even when the reduced form estimator obtains consistent estimates of the social effect of peers. The problem is that the residual or unobservable fixed effect also is a function of the peer group through the social multiplier (i.e., the fact that the effect of resources multiplies in different ways given different sets of students).

[60] Graham et al. (2008) also discuss this limitation in the peer effects literature, pointing out studies generally stop short of estimating parameters that are directly policy relevant.

[61] For evidence of matching in the present context, see, for instance, Clotfelter et al. (2006), who find that nonwhite students are generally more likely to be assigned novice teachers than white students.

[62] Appendix A.5 traces out a simple example of this. For a more detailed explanation, see Cooley (2009).

I make this intuition explicit in the experiment below by first estimating the new equilibrium achievement when resources are equalized across classrooms/schools taking into account the social multipliers created by the redistribution. I then estimate the equilibrium achievement when students are randomly assigned to peer groups (effectively integrating classroom), holding resources fixed at the average level. I use the parameter estimates from the two stage quantile regression procedure described in Section 5 and in Tables 4 and 5.[63] I finally compare the simulated results from this two stage experiment to a reduced form specification, which is also estimated quite flexibly with the quantile estimator.

In reality, school desegregation policies are almost exclusively restricted to be within district lines, a result of the Supreme Court's ruling in the 1974 case of Milliken v. Bradley that a federal court could not force integration across district lines. As a result, it is not uncommon to find wealthy, high-achieving, predominantly white school districts next to poorer, lower-achieving, racially-mixed school districts. Thus, I consider an experiment of desegregating schools across the lines of two such neighboring school districts in North Carolina–Durham (a lower-achieving, predominately nonwhite school district and home of Duke University) and Chapel Hill (a higher-achieving, predominately white school district and home of University of North Carolina-Chapel Hill). I focus on 5^{th} graders in 2001-02 and restrict the sample to classrooms with at least one peer of the other race, the estimation sample.

As illustrated in Table 6, 5^{th} graders in Chapel Hill schools have average reading scores that are about 55% of a standard deviation higher than Durham 5^{th} graders. Durham public schools have a much larger minority population—56% of fifth graders are nonwhite, as compared to only 21% in Chapel Hill. But, Durham is also more segregated on average. For instance, the average class for a white Durham student is 45% nonwhite, while the average class for a nonwhite Durham student is 64% nonwhite. This is also reflected in other disparities in peer characteristics across the average white and nonwhite classrooms in Durham. In Chapel Hill, whites and nonwhite students are in classes that are 22% nonwhite on average. However, this is still evidence of some small disparities in teacher quality across the average white and nonwhite Chapel Hill classrooms.

Figures 5 to 7 show the gains (or losses) in achievement relative to observed predicted achievement for the overall student population, nonwhites and whites from (1) equalizing resources and (2) desegregation (random assignment to peer groups) across districts after

[63] In the simulations, it is necessary to assign a value of θ_i to each student. I treat θ_i as a random shock and assign students randomly to quantiles of the conditional achievement distribution.

equalizing resources using my two stage quantile estimator. The left-hand-side figures show changes in achievement for Durham students and the right-hand-side for Chapel Hill. Individual gains are plotted for different percentiles of the unconditional observed achievement distributions (for instance, the gains in achievement for students at the 10^{th} percentile of Durham's observed achievement distribution), with percentiles on the x-axis and standard deviation changes in achievement for the experiment relative to the observed achievement on the y-axis.

Figure 5 shows that equalizing resources has little effect on students in Durham and produces small gains for students at the lowest percentiles of the achievement distribution in Chapel Hill. Figure 6 shows that the gains in Chapel Hill are driven primarily by improvements for lower-achieving nonwhite students. Recall from Table 5 that observed teacher quality has little effect on nonwhite students. The gain for nonwhites in Chapel Hill is being driven by the increase in the unobservable school attribute (the school fixed effects), which are notably lower for nonwhites in Chapel Hill (0.02 compared to 0.07 for Durham students as shown in Table 6).

Not surprisingly, desegregation across districts produces gains for Durham students at all but the higher percentiles, as seen in Figure 5. The effects of desegregation are driven almost entirely by the diverse peer groups and not resource equalization. Figures 6 and 7 show that the gains at the lowest percentile are being driven both by large improvements to the lowest-achieving nonwhite and white Durham students, with respective gains of 0.9 and 1 standard deviation. The gains for nonwhite Durham students diminish across the quantiles, with the students above the median experiencing losses. The highest achieving nonwhite students experience losses of about -0.35 of a standard deviation. White students, on the other hand, experience small gains of as much as 0.05 at all but the highest percentile of the achievement distribution, where the losses are about -0.22 of a standard deviation. The losses for the highest-achieving Durham students may potentially be explained in part by existing sorting in the Durham schools.

Chapel Hill students on the other hand are generally worse off from the merger at all but the lowest percentiles. It is worth noting further, that the gains at the lower percentiles for nonwhite are driven almost entirely by the resource equalization rather than the Durham peers. Nonwhite students at the median experience losses of about -0.5 and higher-achieving whites experience losses of about -0.3. About a third of this loss is driven by the equalization of resources, most likely the lower teacher experience which is particularly meaningful for the higher-achieving white students (as seen in Table 4).

Translating these estimates into effects on the achievement gap, Figure 8 first shows the gap between whites and nonwhites at the different percentiles of the achievement distribution under the two different assignment policies.[64] Overall, the gap is larger for lower-achieving Chapel Hill students than for Durham students, whereas the gap at the upper quantiles is comparable. Figure 9 computes the changes in the gap at the different percentiles resulting from the experiments. Desegregation leads to a decrease in the gap for the lowest-achieving Durham students of about -0.1 and an increase for the highest-achieving of about 0.05. Across most percentiles, the gap narrows for Chapel Hill students, by -0.15 of a standard deviation for the lowest achievers and by as much as -0.25 for the highest achievers.

As seen in the previous figures, the effects at the lower percentiles are driven by gains to nonwhites, (and whites in the case of Durham). In contrast, the effect at the upper percentiles is driven by losses to both whites and nonwhites. In the case of Chapel Hill, the loss to whites dominates the loss to nonwhites, whereas the opposite is true in Durham. For Chapel Hill students at the middle of the achievement distribution the results are mixed, but mostly support some narrowing of the gap. These changes are driven mostly by gains to nonwhite students, though a couple percentiles are an exception where there are losses to nonwhites and the gap widens. Given these disparate gains and losses across the percentiles and districts, the overall achievement gap only narrows by 0.02 of a standard deviation on average from desegregation.

As emphasized above, an important contribution of this paper is estimating the endogenous, rather than the reduced form effect of peers. I compare the above estimates to predictions from a reduced form quantile estimator, which includes the same controls as in the 2SQR approach but without average peer achievement. Contrasting Figures 9 with the reduced form results in Figure 10, we can begin to quantify the importance of isolating the social multiplier effect for regrouping analyses.

The predictions of my approach and the reduced form approach are very different, particularly for Chapel Hill students. The reduced form approach predicts that the gap actually increases for almost all percentiles of the achievement distribution in Chapel Hill, in contrast to the opposite predictions from my estimator. For Durham, the reduced form approach predicts a larger narrowing in the gap at the lower percentiles of the initial achievement distribution and particularly for students in the middle of the distribution. In almost all cases,

[64] Note that the x-axis differs here from the previous graphs, in that the gap at the 10^{th} percentile is calculated as the difference in the 10^{th} percentile of the white and nonwhite achievement distributions in the different settings. The previous graphs compared gains of students at different points of the observed achievement distribution.

a larger percentage of the change resulting from the merger is attributed to a change in resources for the reduced form approach rather than to peers. Thus, the estimators attribute very different weights to the benefits of resource equalization relative to diverse peers. For a particularly stark contrast, my estimator predicts that all of the improvement for Durham students from desegregation at the lowest percentile is driven by the new peers whereas the reduced form estimator predicts that the change is all driven by resource equalization.

Whether the narrowing in the gap from desegregation results from resource redistribution or peer reallocation may be a key question for policy. This example illustrate how obtaining estimates of the endogenous effect of peers may be central for accurately decomposing the effects of regrouping policies. However, even if policy makers only care about the overall policy effect of desegregation, this comparison also make clears that the magnitudes and even sign of the policy effect are markedly different when one fails to take into account the social multipliers created by desegregation.

8 Conclusion

This paper presents new evidence on the effect peer achievement on a student's achievement. My findings suggest that, in ignoring behavioral spillovers deriving through contemporaneous achievement, prior studies have severely understated the effect of peers.[65] I also find that students appear to form race-based reference groups within classrooms. White students conform only to white peer achievement and nonwhites to nonwhite peers. This type of heterogeneity has not been explored before in the literature and has important implications for understanding the effect of desegregation.

The effects of the achievement of same-race peers are comparable in magnitude to some of the more important determinants of student achievement found in the literature, such as teacher quality and class size. For instance, Rivkin et al. (2005) report that a one standard deviation increase in teacher quality leads to approximately 0.095 of a standard deviation increase in reading. Using findings from Project Star, they report that this change is comparable to a reduction in class size of 10 students in fourth grade and 13 in fifth grade. The positive effect of same-race peers at the median and above is slightly double the magnitude of this effect, while the effect of same-race peers on the lowest achieving students are 4 to 5 times the magnitude.

[65]This is also supported in Graham (2008) who finds large social effects of peers using excess variance contrasts.

To answer the central question posed in this paper, I find that desegregating peer groups would lead to only a small narrowing of the achievement gap of 0.02 of a standard deviation on average. However, an important contribution of this paper is to explore how the effect of peers varies across the percentiles of the achievement distribution and the implications for desegregation. I find that simply focusing on the mean effect of desegregation masks important distributional effects. For lower-achieving students the gap narrows under desegregation by about 10% of the initial gap. Importantly the narrowing at the lower percentiles is driven by improvements to nonwhite achievement rather than losses to white. There is also some evidence of narrowing in the achievement gap at the upper percentiles, but this derives through losses to achievement for both higher-achieving whites and nonwhites.

To place my results in context, previous studies on desegregation have found mixed evidence as to whether higher compositions of nonwhite students hurts nonwhite achievement. 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 contextual peer effects, or in the present context, peer racial composition.[66] I further argue that estimates of endogenous effects may often be necessary for determining the effects of regrouping. I show that the predictions from reduced form peer effect regressions both misstate the overall effect of desegregation on the achievement gap and confound effects of resource reallocation from desegregation with the effect of diverse peers.

My identification and estimation strategy brings new insight to the desegregation literature by allowing the effect of desegregation to operate both through heterogeneity in responses to peer achievement of different races (both by race and across the percentiles of the achievement distribution), while also permitting the percentage nonwhite to have a direct effect on achievement. While endogenous peer effects are much larger in magnitude than the contextual peer effect, I also find that having a higher percentage nonwhite in classrooms hurts nonwhite achievement and is particularly bad for nonwhite students in the middle of the achievement distribution. The effect on whites is not statistically significantly different from 0.

The focus of this paper is on isolating one mechanism, namely peers, through which desegregation may help narrow the achievement gap. While this is a potentially important component which the literature has had a hard time quantifying, 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 the general equilibrium effects of residential sorting

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

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.[67] I discuss the conditions needed to identify peer effects under 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.

To the best of my knowledge, my paper is also the first to apply a flexible nonparametric estimator to peer effects in education and more broadly to educational production functions. 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 richer understanding of the effect of peers and alternative grouping strategies.

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.

Ammermueller, A. and Pischke, J.-S. (2006), Peer effects in European primary schools: Evidence from PIRLS. NBER Working Paper.

Arcidiacono, P., Foster, G., Goodpaster, N. and Kinsler, J. (2009), Estimating spillovers in the classroom with panel data, with an application to the classroom. Working Paper.

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

Betts, J. R. and Zau, A. (2004), Peer groups and academic achievement: Panel evidence from administrative data. University of California, San Diego and Public Policy Institute of California and Public Policy Institute of California.

[67] Models of strategic interactions among agents are also important in the industrial organization literature, particularly in models of entry and auction models.

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

Bramoulle, Y., Djebbari, H. and Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* **150**(1), 41–55.

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. (2007), 'Racial segregation and the black-white test score gap', *Journal of Public Economics* **91**(11-12), 2158–2184.

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.

Cipollone, P. and Rosolia, A. (2007), 'Social interactions in high school: Lessons from an earthquake', *American Economic Review* **97**(3), 948–965.

Clotfelter, C. T., Ladd, H. F. and Vigdor, J. L. (2003), 'Segregation and resegregation in North Carolina's public school classrooms', *North Carolina Law Review* **81**(4), 1463–1512.

Clotfelter, C. T., Ladd, H. F. and Vigdor, J. L. (2006), 'Teacher-student matching and the assessment of teacher effectiveness', *Journal of Human Resources* **41**(4), 778–820.

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. (2009), Can achievement peer effect estimates inform policy? a view from inside the black box. Working Paper.

Figlio, D. N. (2007), 'Boys named sue: Disruptive children and their peers', *Education Finance and Policy* **2**(4), 376–394.

Fordham, S. and Ogbu, J. (1986), 'Black students' school success: Coping with the "burden of acting white"', *The Urban Review* **18**, 176–206.

Fryer, Jr., R. G. (2009), 'An empirical analysis of 'acting white'', *Journal of Public Economics* pp. –.

Gaviria, A. and Raphael, S. (2001), 'School-based peer effects and juvenile behavior', *The Review of Economics and Statistics* **83**(2), 257–268.

Giorgi, G. D., Pellizzari, M. and Redaelli, S. (2007), Be as careful of the books you read as of the company you keep: Evidence on peer effects in educational choices, IZA Discussion Papers 2833, Institute for the Study of Labor (IZA).

Graham, B. S. (2008), 'Identifying social interactions through conditional variance restrictions', *Econometrica* **76**(3), 643–660.

Graham, B. S., Imbens, G. and Ridder, G. (2008), Measuring the average outcome and inequality effects of segregation in the presence of social spillovers. 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 Applied Econometrics* **18**(5), 527–544.

Hanushek, E., Kain, J. and Rivkin, S. (2009), 'New evidence about brown v. board of education: The complex effects of school racial composition on achievement', *Journal of Labor Economics* **27**(3), 349–383.

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.

Jackson, C. K. (2009), 'Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation', *Journal of Labor Economics* **27**(2), 213–256.

Jacob, B. (2005), 'Accountability, incentives and behavior: Evidence from school reform in Chicago', *Journal of Public Economics* **89**(5-6), 761–796.

Jencks, C. and Mayer, S. E. (1990), The social consequences of growing up in a poor neighborhood, *in* J. Laurence E. Lynn and M. G. Geary, eds, 'Inner-City Poverty in the United States', National Academy Press, Washington, DC, pp. 111–186.

Kinsler, J. (2006), Suspending the right to an education or preserving it? a dynamic equilibrium model of student behavior, achievement, and suspension. Working Paper.

Krauth, B. V. (2005), 'Peer effects and selection effects on smoking among canadian youth', *The Canadian Journal of Economics / Revue canadienne d'Economique* **38**(3), 735–757.

Laschever, R. (2008), The doughboys network: Social interactions and the employment of World War I veterans. Working Paper.

Lavy, V. and Schlosser, A. (2007), Mechanisms and impacts of gender peer effects at school, NBER Working Papers 13292, National Bureau of Economic Research, Inc.

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

Linn, R. L. and Welner, K. G. (2007), Race-conscious policies for assigning students to schools: Social science research and the Supreme Court cases, Technical report, National Acadmy of Education, Committee on Social Science Research Evidence on Racial Diversity in Schools.

Ma, L. and Koenker, R. (2006), 'Quantile regression methods for recursive structural equation models', *Journal of Econometrics* **134**(2), 471–506.

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

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.

Nakajima, R. (2004), Measuring peer effects on youth smoking behavior. Working Paper.

Neal, D. A. and Schanzenbach, D. W. (forthcoming), 'Left behind by design: Proficiency counts and test-based accountability', *Review of Economics and Statistics* .

Ogbu, J. U. (2003), *Black Americans in an Affluent Suburb: A Study of Academic Disengagement*, Lawrence Erlbaum Associates, Mahwah, NJ.

Reback, R. (2008), ‘Teaching to the rating: School accountability and the distribution of student achievement’, *Journal of Public Economics* **92**(5-6), 1394–1415.

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

Rivkin, S. G., Hanushek, E. A. and Kain, J. F. (2005), ‘Teachers, schools, and academic achievement’, *Econometrica* **73**(2), 417–458.

Vigdor, J. and Ludwig, J. (2008), Segregation and the black-white test score gap, *in* K. Magnuson and J. Waldfogel, eds, ‘Steady Gains and Stalled Progress: Inequality and the Black-White Test Score Gap’, Russell Sage Foundation, New York.

Vigdor, J. and Nechyba, T. (2004), Peer effects in elementary school: Learning from ’apparent’ random assignment. Working Paper.

Vigdor, J. and Nechyba, T. (2007), Peer effects in North Carolina Public Schools, *in* P. Peterson and L. Woessmann, eds, ‘Schools and the Equal Opportunity Problem’, MIT Press.

A Appendix

A.1 Tables

Table 1: Summary Statistics by Race: Mixed-Race Classrooms

	White		Nonwhite	
	Mean	Std. Dev.	Mean	Std. Dev.
Reading score (standardized)	0.4849	0.8964	-0.2107	0.8826
Male	0.5054	0.5000	0.4888	0.4999
Parent HS/some post-sec.	0.6030	0.4893	0.7803	0.4141
Parent 4-year degree+	0.3493	0.4768	0.1069	0.3090
Characteristics of Classrooms				
Avg. peer reading	0.2755	0.3971	0.0998	0.4147
Avg. white peer reading	0.4755	0.4233	0.3790	0.4809
Avg. nonwhite peer reading	-0.1565	0.5041	-0.2254	0.4252
% white ach. level 1 or 2	0.1637	0.1403	0.1893	0.1796
% nonwhite ach. level 1 or 2	0.3609	0.2439	0.3853	0.2098
% nonwhite	0.3155	0.1887	0.4865	0.2226
% parent with HS degree	0.6423	0.2136	0.7022	0.1954
% parent with 4-year +	0.2823	0.2364	0.2212	0.2087
Class size	23.15	3.366	22.42	3.524
Teacher with adv. degree	0.2752	0.4466	0.2560	0.4364
Teacher experience	12.45	9.680	12.02	9.845
N	344,885		207,323	

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 1997/98 to 2001/02. Includes only classrooms with at least 2 students of each race.

Table 2: First Stage Regressions
(Dependent variable: average white/NW peer reading score)

	White		Nonwhite	
Dependent Variable	Avg. White	Avg. NW	Avg. White	Avg. NW
Accountable*% white ach. level 1 or 2	0.2452***	0.1290**	0.1618***	0.0272
	[0.0398]	[0.0640]	[0.0598]	[0.0447]
Accountable*% white ach. level 3	0.0972***	0.0358	0.1521***	-0.0262
	[0.0321]	[0.0531]	[0.0530]	[0.0400]
Accountable*% NW ach. level 1 or 2	0.0295	0.1014	-0.0084	0.1488***
	[0.0280]	[0.0618]	[0.0384]	[0.0396]
Accountable*% NW ach. level 3	0.005	0.0083	-0.0257	0.0023
	[0.0288]	[0.0619]	[0.0369]	[0.0400]

*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 1997-98 to 2001-02. Peer and individual characterstics, classroom inputs, school by year fixed effects and constant as described in equation (5.1) also included.

Table 3: Heterogeneous Reference Groups

Dependent Variable: Reading	Mean White	Mean Nonwhite	Median White	Median Nonwhite
Endogenous Peer Effects				
Avg. white peer reading	0.4990***	0.1630	0.4996**	0.1730
	[0.1410]	[0.1311]	[0.2097]	[0.2147]
Avg. nonwhite peer reading	-0.0019	0.6427***	-0.05	0.5422**
	[0.1449]	[0.1290]	[0.2261]	[0.2130]
Contextual Peer Effects				
% white ach. level 1 or 2	0.7962***	0.2729	0.7765**	0.2877
	[0.2324]	[0.2185]	[0.3480]	[0.3566]
% white ach. level 3	0.3078***	0.1255	0.2904*	0.1422
	[0.0995]	[0.0968]	[0.1520]	[0.1566]
% NW white ach. level 1 or 2	-0.0174	0.7798***	-0.0768	0.6547**
	[0.1885]	[0.1605]	[0.2977]	[0.2682]
% NW ach. level 3	-0.026	0.1753***	-0.0432	0.1249
	[0.0653]	[0.0486]	[0.1051]	[0.0832]
% nonwhite	-0.0457**	-0.0921***	-0.0577	-0.1115***
	[0.0204]	[0.0239]	[0.0374]	[0.0310]
% male	0.0082	0.0812***	0.0145	0.0559
	[0.0178]	[0.0296]	[0.0307]	[0.0458]
% parents HS Degree	-0.0685	-0.3029***	-0.0507	-0.2714***
	[0.0466]	[0.0507]	[0.0729]	[0.0781]
% parents 4-year degree	-0.1624**	-0.4691***	-0.1366	-0.4133**
	[0.0703]	[0.1129]	[0.1122]	[0.1714]
Policy Variables				
Accountability	-0.0242**	-0.0330*	-0.0210	-0.0290
	[0.0092]	[0.0164]	[0.0145]	[0.0225]
Achievement level 1 or 2	-1.6660***	-1.5988***	-1.6366***	-1.5905***
	[0.0042]	[0.0060]	[0.0057]	[0.0073]
Achievement level 3	-0.7748***	-0.7309***	-0.7502***	-0.7104***
	[0.0026]	[0.0050]	[0.0039]	[0.0057]
Accountable*Level 1 or 2	0.1604***	0.1023***	0.1959***	0.1550***
	[0.0082]	[0.0113]	[0.0112]	[0.0127]
Accountable*Level 3	0.0740***	0.0453***	0.0545***	0.0220*
	[0.0047]	[0.0091]	[0.0064]	[0.0122]
Teacher adv degree	-0.0008	-0.0030	-0.0009	-0.0078*
	[0.0024]	[0.0028]	[0.0038]	[0.0050]
Teacher experience	0.0270***	0.0104	0.0235**	0.0128
	[0.0068]	[0.0110]	[0.0117]	[0.0149]
Experience^2	-0.0066***	-0.0025	-0.0053*	-0.0026
	[0.0018]	[0.0028]	[0.0032]	[0.0036]
N	344,885	207,323	344,885	207,323
R^2	0.5951	0.5354		

*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. Dummy variables for male, parent with high school degree and parent with 4-year degree included and have expected sign. School by year fixed effects, grade fixed effects, class inputs and constant also included.

Table 4: Average Marginal Effects of Peers for Whites
(Dependent variable: standardized reading score)

	Mean	.1 Quantile	Median	.9 Quantile
Avg white peer reading	0.2171**	0.4012***	0.2114**	0.2495*
	[0.1032]	[0.1210]	[0.0888]	[0.1418]
Avg nonwhite peer reading	-0.0068	-0.0113	-0.0252	-0.1866
	[0.1593]	[0.1696]	[0.1140]	[0.2624]
% white ach. level 1 or 2	0.1151**	0.2226***	0.1089**	0.1281
	[0.0571]	[0.0671]	[0.0488]	[0.0789]
% white ach. level 3	0.0497*	0.1066***	0.0460*	0.0512
	[0.0274]	[0.0322]	[0.0241]	[0.0366]
% NW ach. level 1 or 2	-0.0074	-0.0026	-0.0187	-0.1275
	[0.1013]	[0.1076]	[0.0726]	[0.1662]
% NW ach. level 3	-0.0076	-0.0058	-0.0101	-0.0491
	[0.0344]	[0.0363]	[0.0246]	[0.0566]
% nonwhite	-0.0088	-0.0062	-0.0109	-0.0170
	[0.0091]	[0.0089]	[0.0071]	[0.0141]
% male	0.0008	0.0037	0.0012	-0.0046
	[0.0034]	[0.0038]	[0.0026]	[0.0055]
% parents HS degree	-0.0148	-0.0318	-0.0108	0.0107
	[0.0212]	[0.0221]	[0.0156]	[0.0359]
% parents 4-year degree	-0.0392	-0.1030***	-0.0323	0.0028
	[0.0321]	[0.0347]	[0.0265]	[0.0501]
Teacher adv. degree	-0.0002	0.0009	-0.0004	-0.0014
	[0.0022]	[0.0027]	[0.0017]	[0.0032]
Teacher experience	0.0250**	0.0076	0.0228**	0.0376**
	[0.0125]	[0.0133]	[0.0113]	[0.0187]
Teacher experience^2	-0.0176*	-0.0073	-0.0150*	-0.0261*
	[0.0097]	[0.0104]	[0.0090]	[0.0137]

*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 5: Average Marginal Effects of Peers for Nonwhites
(Dependent variable: standardized reading score)

	Mean	.1 Quantile	Median	.9 Quantile
Avg white peer reading	0.0681	0.0620	0.0832	-0.0274
	[0.1128]	[0.1356]	[0.1033]	[0.1369]
Avg nonwhite peer reading	0.2845***	0.4877***	0.2306**	0.2563*
	[0.1069]	[0.1488]	[0.0906]	[0.1445]
% white ach. level 1 or 2	0.0429	0.0438	0.0517	-0.0192
	[0.0692]	[0.0832]	[0.0640]	[0.0838]
% white ach. level 3	0.0215	0.0235	0.0272	-0.0123
	[0.0321]	[0.0394]	[0.0300]	[0.0383]
% NW ach. level 1 or 2	0.1716***	0.3069***	0.1373**	0.1449
	[0.0664]	[0.0915]	[0.0563]	[0.0898]
% NW ach. level 3	0.0360*	0.0785***	0.0241	0.0247
	[0.0188]	[0.0260]	[0.0160]	[0.0251]
% nonwhite	-0.0191**	-0.0092	-0.0248***	-0.0178*
	[0.0077]	[0.0126]	[0.0069]	[0.0093]
% male	0.0075*	0.0162**	0.0050	0.0028
	[0.0045]	[0.0065]	[0.0041]	[0.0059]
% parents HS degree	-0.0594***	-0.0963***	-0.0530***	-0.0473**
	[0.0166]	[0.0241]	[0.0153]	[0.0215]
% parents 4-year degree	-0.0980**	-0.1661***	-0.0863**	-0.0599
	[0.0378]	[0.0505]	[0.0358]	[0.0473]
Teacher adv. degree	-0.0013	0.0012	-0.0034	-0.0001
	[0.0024]	[0.0032]	[0.0022]	[0.0032]
Teacher experience	0.0078	-0.0173	0.0126	0.0182
	[0.0166]	[0.0225]	[0.0147]	[0.0214]
Teacher experience^2	-0.0041	0.0109	-0.0073	-0.0092
	[0.0120]	[0.0162]	[0.0101]	[0.0158]

*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: Avg. Characteristics of Durham and Chapel Hill Public Schools (Grade 5, Academic Year 2001-02)

	Durham		Chapel Hill	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Reading	0.4063	0.8695	0.9529	0.7917
Avg. White Reading	0.8270	0.7859	1.1577	0.6530
Avg. Nonwhite Reading	0.0765	0.7859	0.1975	0.8029
% Nonwhite	0.5606	0.4965	0.2133	0.4100
School-year FE (white)	0.0843	0.2095	0.0793	0.0611
school-year FE (NW)	0.0725	0.1473	0.0174	0.1834
Nonwhite classroom characteristics				
Avg. White Reading	0.6949	0.4323	1.1476	0.2037
Avg. Nonwhite Reading	0.0597	0.3268	0.1791	0.4018
% Nonwhite	0.6401	0.1977	0.2244	0.0595
% Parent with HS Degree	0.5989	0.2042	0.1986	0.0673
% Parent with 4-year Degree	0.3553	0.2172	0.7919	0.0660
Teacher adv. Degree	0.2393	0.4269	0.3774	0.4870
Teacher experience	11.00	10.28	15.23	10.44
White classroom characteristics				
Avg. White Reading	0.8073	0.3785	1.1426	0.1892
Avg. Nonwhite Reading	0.1380	0.4355	0.1831	0.4395
% Nonwhite	0.4544	0.2332	0.2179	0.0895
% Parent with HS Degree	0.4990	0.1909	0.1923	0.0820
% Parent with 4-year Degree	0.4694	0.1965	0.7989	0.0848
Teacher adv. Degree	0.2118	0.4089	0.5294	0.4998
Teacher experience	10.71	10.65	14.57	10.18
N	1461		497	

A.2 Figures

Figure 1: Student Accountability in Charlotte-Mecklenburg Schools

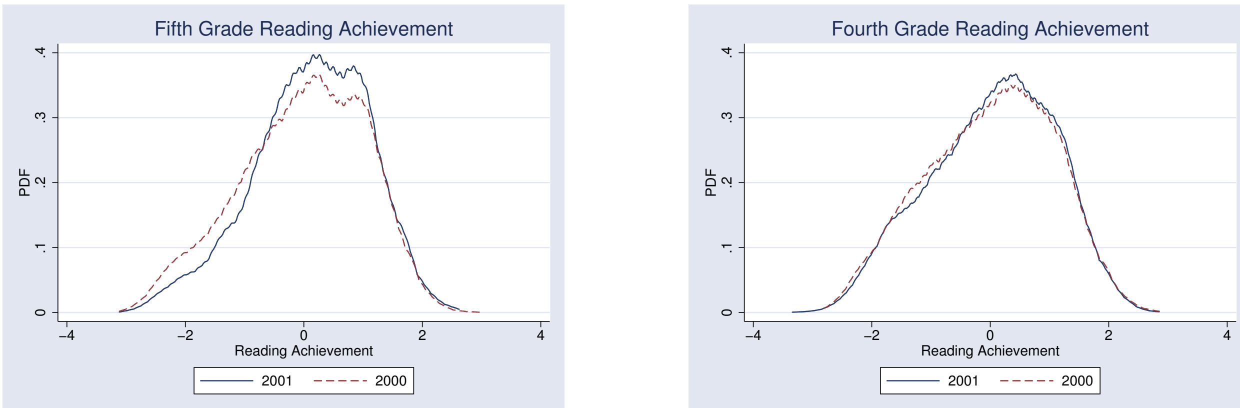
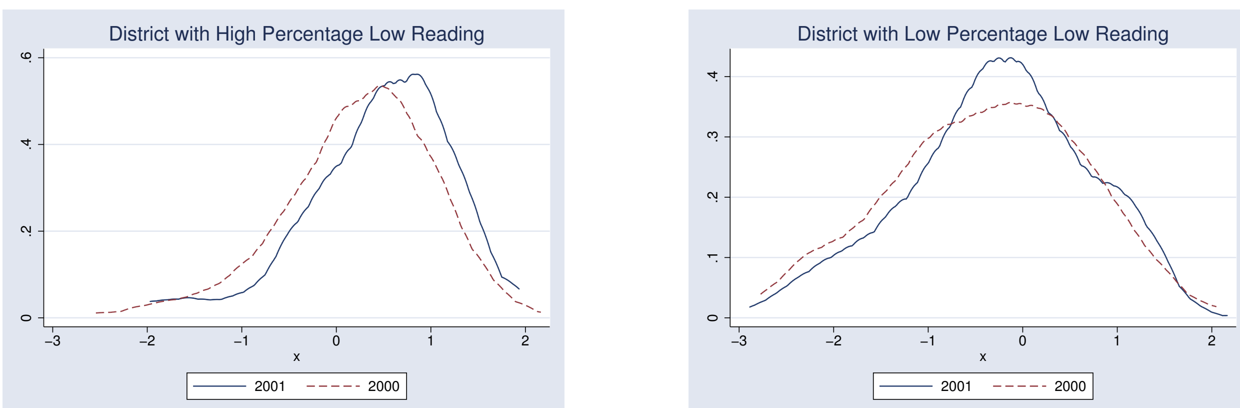


Figure 2: Comparison of Two Districts with Different Concentrations of Low Achievers



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

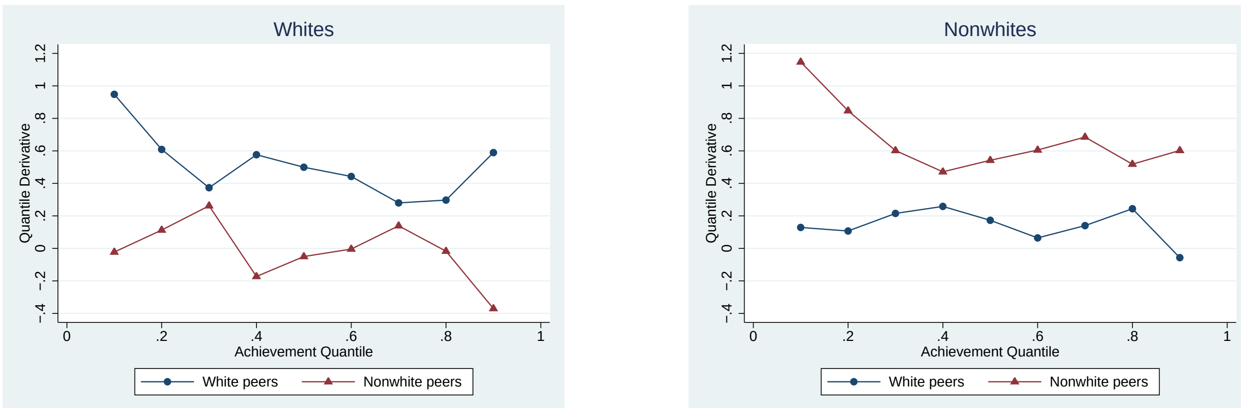


Figure 4: Contextual Peer Group Composition Effects: % Nonwhite

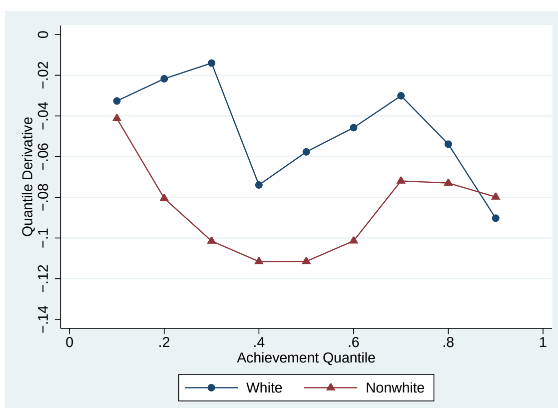


Figure 5: Overall Achievement Gains from Durham and Chapel Hill Experiment

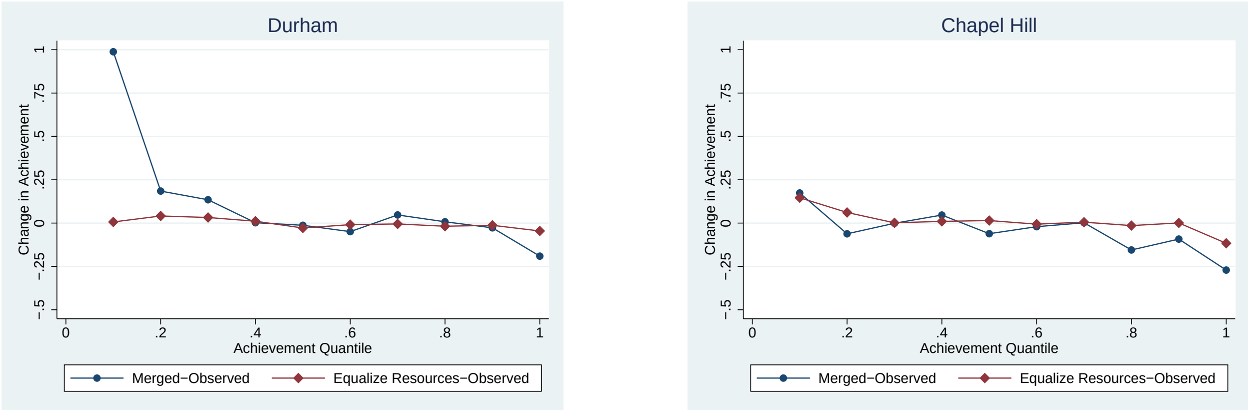


Figure 6: Nonwhite Achievement Gains from Durham and Chapel Hill Experiment

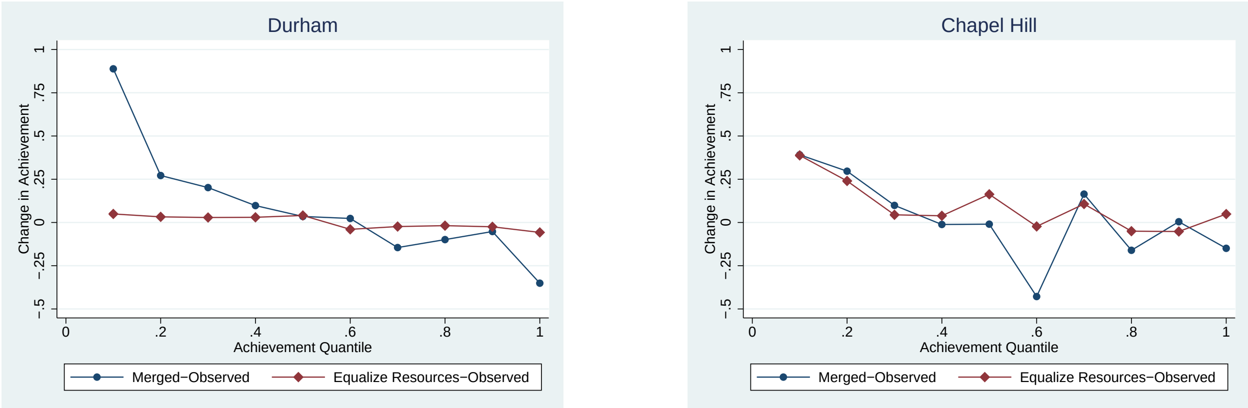


Figure 7: White Achievement Gains from Durham and Chapel Hill Experiment

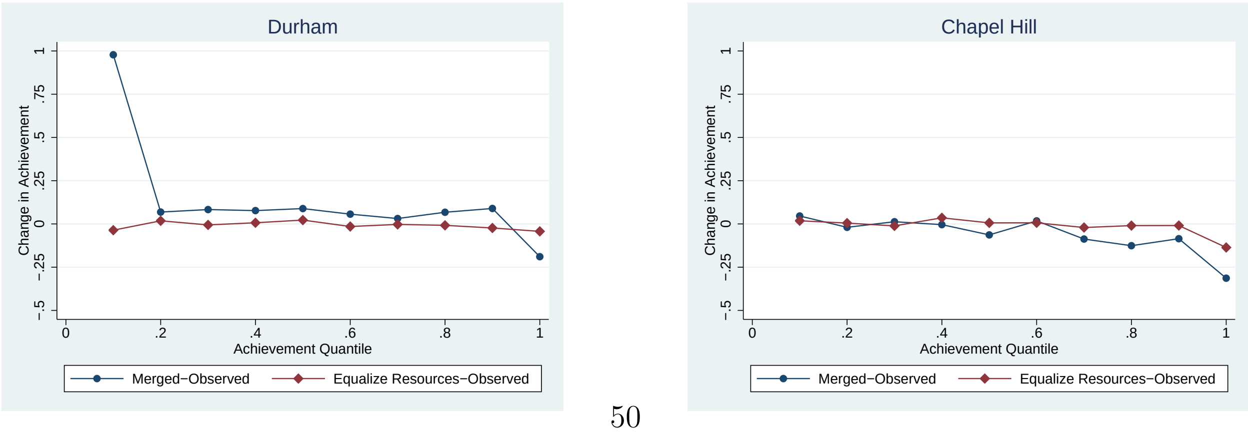


Figure 8: White-Nonwhite Achievement Gap from Experiment

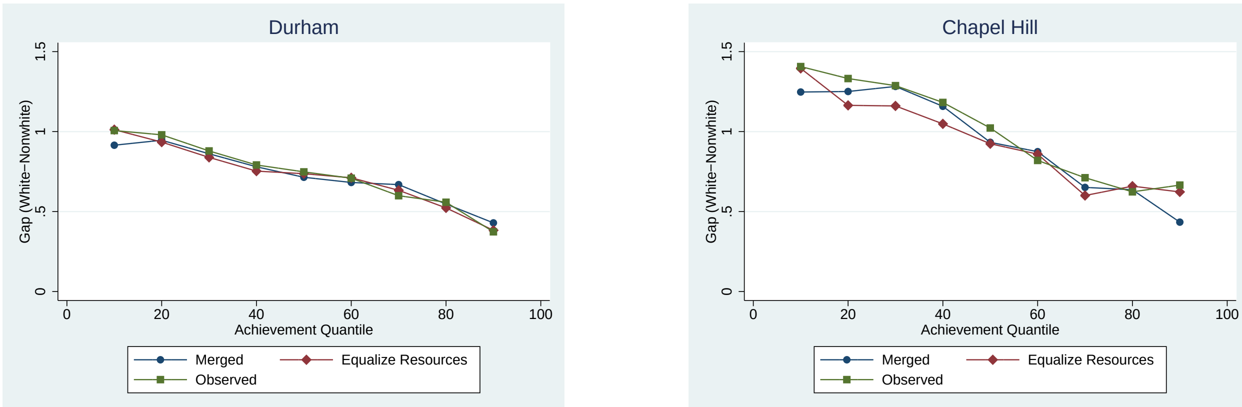


Figure 9: Changes in Achievement Gap from Experiment

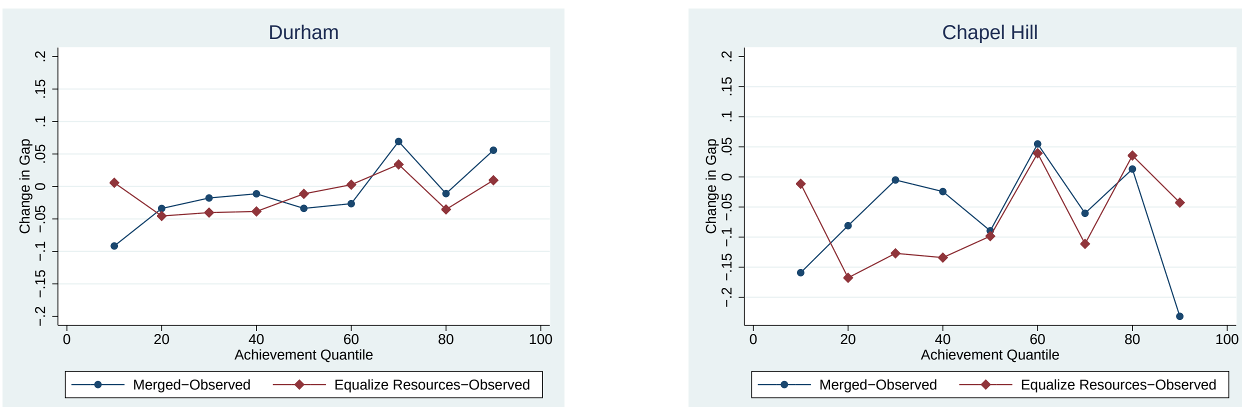
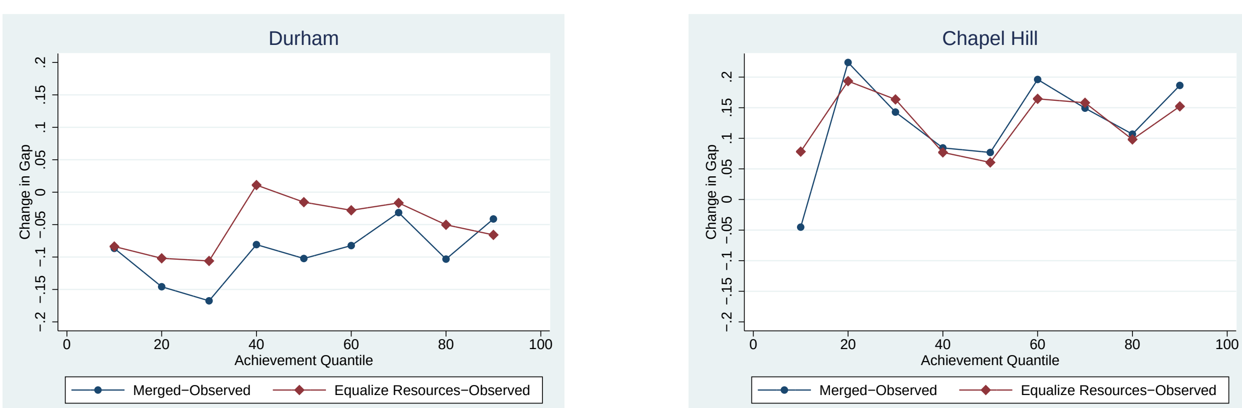


Figure 10: Changes in Achievement Gap from Experiment:Reduced Form



A.3 Proofs

Mapping from Effort to Achievement Equilibrium. I illustrate how the game in effort maps into a game in achievement. Given achievement in monotonically increasing in effort, 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 predetermined 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:

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

$\square$

Proof of Identification of μ. Because all results hold conditional on the exogenous characteristics $(X_i, \bar{X}_{-i}, K, P_i)$, 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 (A4), and the fourth by (A2). Therefore, $\mu = h^{-1}(\bar{p}_0, \bar{y}_0)$ is identified by the joint distribution of $(\bar{Y}_{-i}^*, \bar{P}_{-i})$. $\square$

Proof of Identification of QSF. 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 $(\bar{Y}^*_{-i}, \mu, \theta_i)$. $\square$

A.4 Interpretation of Contextual Effects

Cooley (2009) provides a detailed explanation of the interpretation of contextual effects using a simple linear-in-means context, but a condensed intuition is provided here to aid in interpreting the above results. First, consider contextual 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 characteristics that enter his achievement directly. This is the way that contextual 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 instance, higher peer parental education may substitute for a student's own effort. 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. Finally, when there are spillovers from peer effort, conditional on a given level of peer achievement, a higher level of peer parental education suggests a lower level of peer effort. Given these three countervailing effects, the sign of $\hat{\beta}_4$ is indeterminate. A similar conclusion holds for classroom productivity, μ, suggesting that the assumption of an upward bias from unobserved correlated effects may not hold.

A.5 Regrouping and Role of Social Multipliers

This example is taken from Cooley (2009) to illustrate why reduced form estimates may fail to capture the effect of regrouping. Consider a simple setting with two classrooms, $g \in \{c, d\}$, with two students each. Initially the allocation g_0 is such that students $\{1, 2\}$ are in c and $\{3, 4\}$ in d. The characteristics X_i are understood to include a dummy variable for whether the student is white or nonwhite. Consider a simplified linear version of equation 5.2 where there are heterogeneous responses to peer achievement across races, i.e., the statistical model is modified so that

$$
\begin{aligned}
Y_{1c} &= X_1\gamma_{xn} + X_2\tilde{\gamma}_{xn} + Y_{2c}\tilde{\gamma}_{yn} + \mu_c\gamma_\mu + \xi_{1c}, \\
Y_{2c} &= X_2\gamma_{xn} + X_1\tilde{\gamma}_{xn} + Y_{1c}\tilde{\gamma}_{yn} + \mu_c\gamma_\mu + \xi_{2c}, \\
Y_{3d} &= X_3\gamma_{xw} + X_4\tilde{\gamma}_{xw} + Y_{4d}\tilde{\gamma}_{yw} + \mu_d\gamma_\mu + \xi_{3d}, \\
Y_{4d} &= X_4\gamma_{xw} + X_3\tilde{\gamma}_{xw} + Y_{3d}\tilde{\gamma}_{yw} + \mu_d\gamma_\mu + \xi_{4d},
\end{aligned}
$$

where w denotes white and n nonwhite. Solving for peer achievement in terms of covariates and plugging back into the structural equation, the associated reduced form equations are then

$$
\begin{aligned}
Y_{1c} &= X_1\pi_{xn} + X_2\tilde{\pi}_{\bar{x}n} + \mu_c\pi_{\mu n} + \zeta_{1c}, \\
Y_{2c} &= X_2\pi_{xn} + X_1\tilde{\pi}_{\bar{x}n} + \mu_c\pi_{\mu n} + \zeta_{2c}, \\
Y_{3d} &= X_3\pi_{xw} + X_4\tilde{\pi}_{\bar{x}w} + \mu_d\pi_{\mu w} + \zeta_{3d}, \\
Y_{4d} &= X_4\pi_{xw} + X_3\tilde{\pi}_{\bar{x}w} + \mu_d\pi_{\mu w} + \zeta_{4d},
\end{aligned}
$$

where $\pi_{xr} \equiv \frac{\gamma_{xr}+\tilde{\gamma}_{yr}\tilde{\gamma}_{xr}}{1-\tilde{\gamma}_{yr}^2}$, $\tilde{\pi}_{xr} \equiv \frac{\tilde{\gamma}_{xr}+\tilde{\gamma}_{yr}\gamma_{yr}}{1-\tilde{\gamma}_{yr}^2}$, $\pi_{\mu r} = \frac{1+\tilde{\gamma}_{yr}}{1-\tilde{\gamma}_{yr}^2}$, and $r \in \{n, w\}$. If students are reassigned to create heterogeneous classes, the reduced form would be

$$
\begin{aligned}
Y_{1c} &= X_1\pi_{xnw} + X_3\tilde{\pi}_{xnw} + \mu_c\pi_{\mu nw} + \zeta_{1c}, \\
Y_{2d} &= X_2\pi_{xnw} + X_4\tilde{\pi}_{xnw} + \mu_d\pi_{\mu nw} + \zeta_{2d}, \\
Y_{3c} &= X_3\pi_{xwn} + X_1\tilde{\pi}_{xwn} + \mu_c\pi_{\mu wn} + \zeta_{3c}, \\
Y_{4d} &= X_4\pi_{xwn} + X_2\tilde{\pi}_{xwn} + \mu_d\pi_{\mu wn} + \zeta_{4d},
\end{aligned}
$$

with $\pi_{xwn} \equiv \frac{\gamma_{xw}+\tilde{\gamma}_{yw}\tilde{\gamma}_{xn}}{1-\tilde{\gamma}_{yw}\tilde{\gamma}_{yn}}$, $\tilde{\pi}_{xwn} \equiv \frac{\tilde{\gamma}_{xw}+\tilde{\gamma}_{yw}\gamma_{xn}}{1-\tilde{\gamma}_{yw}\tilde{\gamma}_{yn}}$, $\pi_{\mu wn} = \frac{1+\tilde{\gamma}_{yw}}{1-\tilde{\gamma}_{yn}\tilde{\gamma}_{yw}}$.

The change in the expected achievement for nonwhites is

$$
\begin{aligned}
E(Y_1 + Y_2|g_1) - E(Y_1 + Y_2|g_0) = {} & (X_1 + X_2)\pi_{xnw} + (X_3 + X_4)\tilde{\pi}_{xnw} \\
& - (X_1 + X_2)(\pi_{xn} + \tilde{\pi}_{xn}) + E((\mu_d + \mu_c)\pi_{\mu nw} - 2\mu_c\pi_{\mu n}|\vec{X}).
\end{aligned}
$$

This suggests at least two challenges associated with using the reduced form to estimate effects of regrouping. The first is a support assumption. Given heterogeneous responses to peers, it may not be possible to infer the contextual peer effect of racially mixed classrooms if we only observe homogenous classrooms, i.e., $\pi_{xwn}, \tilde{\pi}_{xwn}, \pi_{nw}, \tilde{\pi}_{xnw}$. Given a sufficiently rich support, however, and a sufficiently flexible estimator, this problem may be mitigated.

Assume that this is the case, and we observe both racially mixed and segregated classrooms. There remains a second problem that derives from the matching of teachers to students, and the fact that unobserved teacher quality in this example may be correlated with the racial composition of the classroom observed in the data. To recover the expected change in average nonwhite achievement, besides the estimated contextual effects it is also necessary to approximate $E((\mu_d + \mu_c)\pi_{\mu nw} - 2\mu_c\pi_{\mu n}|\vec{X})$ using the residuals from the reduced form equations above. Suppose estimates of $\mu_c^* \equiv E(\mu_c\pi_{\mu n}|\vec{X})$ and $\mu_d^* \equiv E(\mu_d\pi_{\mu w}|\vec{X})$ are obtained in the homogenous setting. Similarly for the racially mixed setting, I obtain estimates of $\mu_{cnw}^* \equiv 1/2\mu_c(\pi_{\mu nw} + \pi_{\mu wn})$ and $\mu_{dnw}^* \equiv 1/2\mu_d(\pi_{\mu nw} + \pi_{\mu wn})$. Even though I have effectively assumed that the distribution of teacher quality in the racially mixed setting is the same as in the homogeneous setting, I cannot recover the effect of the reassignment to teachers implicit in the reallocation from the reduced form parameters.

While this illustration is for a simple linear setting, the basic intuition could be extended to the more general framework considered in this paper.

A.6 Supplemental Tables

Table 7: Summary Statistics by Race

	White		Nonwhite	
	Mean	Std. Dev.	Mean	Std. Dev.
Reading score (standardized)	0.4723	0.8975	-0.2195	0.8827
Male	0.5056	0.5000	0.4891	0.4999
Parent HS/some post-sec.	0.6191	0.4856	0.7858	0.4103
Parent 4-year degree+	0.3241	0.4680	0.1030	0.3040
Characteristics of Classroom				
Avg. peer reading	0.3127	0.3963	0.0607	0.4356
Avg. white peer reading	0.4633	0.4126	0.3453	0.5031
Avg. nonwhite peer reading	-0.1230	0.5399	-0.2338	0.4377
% white ach. level 1 or 2	0.1676	0.1367	0.1793	0.1976
% nonwhite ach. level 1 or 2	0.3033	0.2917	0.3896	0.2167
% nonwhite	0.2395	0.2104	0.5311	0.2671
% parent with HS degree	0.6471	0.2130	0.7156	0.1948
% parent with 4-year +	0.2745	0.2366	0.2061	0.2070
Class size	23.03	3.507	22.13	3.719
No peers of other race	0.1455	0.3526	0.0674	0.2507
Teacher with adv. degree	0.2827	0.4503	0.2536	0.4351
Teacher experience	12.74	9.688	12.07	9.878
N	623,986		321,997	

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 1997/98 to 2001/02.

Table 8: Summary Statistics by Apparent Random Assignment†

	Estimation Sample		Random Assignment†	
	Mean	Std. Dev.	Mean	Std. Dev.
Reading score (standardized)	0.2238	0.9527	0.2105	0.9525
Male	0.4992	0.5000	0.5004	0.5000
Parent HS/some post-sec.	0.6696	0.4704	0.6830	0.4653
Parent 4-year degree+	0.2583	0.4377	0.2415	0.4280
Characteristics of Classroom				
Avg. peer reading	0.2095	0.4127	0.1961	0.4078
Avg. white peer reading	0.4393	0.4482	0.4157	0.4400
Avg. nonwhite peer reading	-0.1824	0.4772	-0.1925	0.4821
% white ach. Level 1 or 2	0.1733	0.1567	0.1795	0.1562
% nonwhite ach. Level 1 or 2	0.3700	0.2320	0.3722	0.2343
% nonwhite	0.3797	0.2184	0.3717	0.2177
% parent with HS degree	0.6648	0.2090	0.6779	0.1998
% parent with 4-year +	0.2594	0.2283	0.2428	0.2168
Class size	22.87	3.444	22.82	3.409
Teacher with adv. Degree	0.2680	0.4429	0.2723	0.4451
Teacher experience	12.29	9.745	12.30	9.740
	552208		396553	
N	623,986		321,997	

† Apparent random assignment schools are those that for a given school year had a p-value¿.1 for the joint test that the difference between classroom and school characteristics is significantly different from 0.
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 1997/98 to 2001/02. Includes only classrooms with at least 2 students of each race.