Quantitative Economics 4 (2013), 85–124 1759-7331/20130085

# Identifying peer achievement spillovers: Implications for desegregation and the achievement gap

Jane Cooley Fruehwirth
University of Cambridge and Christ's College

This paper develops a new approach to identifying peer achievement spillovers in the context of an equilibrium model of student effort choices. By focusing on the effect of contemporaneous peer achievement, this framework integrates previously unexplored types of heterogeneity in peer spillovers in the achievement production context. Applying the strategy to North Carolina public elementary school students, I find peer achievement spillovers exist primarily within race-based reference groups, and the magnitude of these spillovers diminishes across the percentiles of the achievement distribution. Simulations highlight the importance of peer achievement spillovers for determining the distributional effects of desegregation relative to flexible reduced-form specifications that focus entirely on predetermined peer characteristics.

Keywords. Peer achievement spillovers, endogenous peer effects, desegregation.

JEL classification. I20, I21, J15.

## 1. Introduction

Understanding the role of peers in achievement production is important for informing how allocations of students to classrooms and sorting across schools might affect student achievement. To date, the literature has provided a wealth of insight into how predetermined characteristics of peers, such as prior achievement or racial composition of the classroom, affect student outcomes.[1] Less is known about spillovers of peers deriving through contemporaneous peer achievement. These types of spillovers are unique in that they capture potentially time-varying behaviors, such as whether peers work hard or misbehave in class. Unlike spillovers from peer characteristics, achievement spillovers have the potential to generate *social multiplier* effects, where one student's behavior affects the behavior of her peer, which in turn affects the student's behavior,

Jane Cooley Fruehwirth: jc738@cam.ac.uk

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. This paper previously circulated as "Desegregation and the achievement gap: Do diverse peers help?"

[1] See, for instance, Gibbons and Telhaj (2008) and Sacerdote (2011) for recent overviews.


DOI: 10.3982/QE93

thus multiplying. As a result, small changes in inputs can lead to large changes in equilibrium achievement when social multipliers are present. The key contribution of this paper is to explore the role of contemporaneous peer achievement spillovers in achievement production.

The lack of evidence on contemporaneous peer achievement spillovers stems in part from the difficult identification challenge, coined the *reflection problem* in Manski (1993).[2] Because achievement is simultaneously determined within a group of students, it is difficult to separate an effect of peer behavior from the direct effect of a student's own behavior. I write down an equilibrium model of peer achievement that motivates potential sources of exclusions that may be used to identify the peer achievement spillovers. I then study peer achievement spillovers using longitudinal administrative data from North Carolina public elementary school fourth and fifth graders.

The model motivates my choice of an exclusion restriction—a student accountability policy. This policy was introduced in North Carolina public schools and requires that a student achieve above a certain level to be automatically promoted to the next grade. Intuitively, students in danger of scoring below the threshold (based on prior achievement) might be induced to work harder, whereas students well above the threshold would not. Thus, a student in a class with a higher percentage of peers affected by the policy would see a larger shift in average peer achievement as a result of the new policy. The identification strategy has the flavor of a difference-in-difference strategy. Peer effects are identified by comparing classrooms with similar compositions of low achievers (those potentially affected by the policy) pre- and postaccountability. Because only fifth graders are affected by the policy, fourth grade classrooms act as a further control group to eliminate any other changes that may have coincided with student accountability.

A remaining identification challenge that has received considerable attention in the literature is that students are nonrandomly assigned to classrooms. With selection into classrooms, it is difficult to separate an effect of peers from unobserved correlated effects, such as the students' abilities or the quality of the teacher. The instrumental variable strategy eliminates the problem of nonrandom assignment as a confounder of peer achievement spillovers, as long as students are not reassigned in response to the policy. I provide support for this assumption by showing that observable compositions of peer groups do not appear to change in response to student accountability.

I apply my strategy of estimating peer achievement spillovers to provide new insight into the effect of racially diverse peers on achievement. This is a timely policy question in the United States, particularly given the movement away from policies that explicitly integrate schools by race (Chemerinsky (2003)).

Addressing the effect of racially diverse peers requires capturing potential heterogeneity in peer spillovers by race and ability. While previous studies have examined the

[2]A burgeoning literature explores how to exploit networks to identify these contemporaneous peer achievement spillovers, such as Bramoulle, Djebbari, and Fortin (2009) and Calvo-Armengol, Patacchini, and Zenou (2009), among others. While a very promising area of research, they focus on friendship spillovers rather than the question at hand, classroom peer effect spillovers, which are arguably quite different in their underlying mechanisms and policy implications. The article by Giorgi, Pellizzari, and Redaelli (2010) is most similar in spirit to the present paper, but exploits overlapping peer groups in college, which cannot be applied in the elementary school setting where classrooms are self-contained.

potential consequences of segregation by contrasting the percentage black or Hispanic in the classroom, the focus on achievement spillovers highlights new channels whereby racially diverse peers may affect achievement. For instance, Fordham and Ogbu (1986), Fryer and Torelli (2010), and others suggested that students may place different weights on the opinion of peers from different races. Incorporating this insight into the achievement production context, I consider whether students form different race-based reference groups in the classroom, conforming to the achievement of peers of the same race. If this is the case, creating racially diverse peer groups may not generate social multipliers across races.

Furthermore, diverse peers may lead to efficiency gains if low-achieving students are more responsive to peers then high-achievers, as minority students are more heavily concentrated in the lower tails of the achievement distribution. Thus, I capture how responses to peer vary across the achievement distribution by applying a quantile treatment effect approach, exploiting insight developed in Imbens and Newey (2009) and Chernozhukov and Hansen (2005) for estimating quantile treatment effects with endogenous regressors.

I find 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.[3] This may not be surprising given that effort spillovers captured by contemporaneous peer achievement entail social multiplier effects, which have not been previously estimated.

While the heterogeneity in peer achievement spillovers by race and percentiles of the achievement distribution provide insight into a potential effect of racially diverse peers, the total effect remains difficult to quantify given the different channels of influence. Thus, I use my parameter estimates to simulate the effect on student achievement of creating racially diverse classrooms. To do this, I also need estimates of the spillovers from predetermined peer characteristics. Estimates of the effect of predetermined characteristics may be biased by selection into classrooms. I use school-by-year fixed effects to address selection, thus exploiting plausibly random cohort variation to identify these remaining peer effects as in studies by Lavy and Schlosser (2011), Hanushek, Kain, Markman, and Rivkin (2003), and Hoxby (2000). Overall, the spillovers from peer characteristics are swamped in magnitude by the spillovers from contemporaneous peer achievement. The simulations reveal that the effect of racially diverse peers varies considerably across the percentiles of the achievement distribution. The simulations might best be interpreted as an upper bound on the effect of racially diverse peers, given remaining concerns about bias in the estimates of peer characteristics and equilibrium responses that would change the composition of students attending public schools.

[3] For instance, see Hanushek, Kain, and Rivkin (2009) and Vigdor and Nechyba (2007).

## 2. Model

The literature on peer spillovers in education production 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. While the literature posits different sources of these peer spillovers, my paper is the first to set forth a theoretical model that would lead to such a production function. The model is useful primarily for motivating potential sources of exclusion restrictions by which peer achievement spillovers can be identified. Secondarily, it also motivates the inclusion of contemporaneous peer achievement in the production function.

I recast students as optimizing agents whose decisions are influenced by their peers. 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 achievement production function with peer spillovers that has traditionally been estimated in the literature.

In what follows, I take peer groups as given and define a peer group to be a classroom of students in a particular time period. A burgeoning literature instead examines the important question of the effect of social networks on student outcomes.[4] The classroom is arguably an important peer group measure, as it is something that can be (and is) manipulated by policy makers, parents, teachers, and principals, partially as a way to improve student outcomes. I discuss the implications of selection into peer groups in Section 4.

Let $i = 1, \dots, N$ index students in a given peer group. Because the focus is on interactions within a particular peer group, I suppress time and classroom subscripts for the moment. Achievement $Y_i \in \mathbb{R}$ is a standardized test score. Let $\mathbf{X}_i$ denote characteristics of a student $i$ such as race, sex, parental education, and known ability. $\mathbf{X}_{-i} = (\mathbf{X}_1, \dots, \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, \dots, \mathbf{X}_N)$ captures the characteristics of $i$'s peers. Besides the composition of the peer group, classrooms are differentiated by characteristics $\tilde{\mathbf{K}}$, which may include classroom resources, teacher quality, or overall classroom productivity.

The achievement production function is

$$Y_i = g(e_i, \mathbf{e}_{-i}; \mathbf{X}_i, \mathbf{X}_{-i}, \tilde{\mathbf{K}}, \theta_i). \tag{2.1}$$

The choice variable of a student $i$ is *effort*, which is chosen on the compact set $e_i \in [\underline{e}, \overline{e}]$. It is defined broadly to include different behavioral 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, $\mathbf{e}_{-i} = (e_1, \dots, e_{i-1}, e_{i+1}, \dots, e_N)$. Furthermore, $i$'s achievement depends on predeter-

[4] See Bramoulle, Djebbari, and Fortin (2009) and Calvo-Armengol, Patacchini, and Zenou (2009), among others.

mined variables, which may include individual and peer characteristics as well as classroom inputs such as teacher quality ($\mathbf{X}_i, \mathbf{X}_{-i}, \tilde{\mathbf{K}}$).

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 $\mathbf{X}_{-i}$. Second, peers may affect achievement through their 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, as in Lazear (2001).[5]

Finally, a student cannot perfectly predict his achievement on an exam, even after choosing his own effort and observing the effort of his peers. This could be for several reasons. A student may not fully know his own ability, given limited experience taking standardized exams or because ability is relative, so it is difficult to know how ability compares across schools. Unpredictable random factors, such as a good night's sleep, may also affect a student's performance on a given test day. These types of unobservables are captured by $\theta_i$. They are allowed to be correlated within peer groups to capture common shocks, such as construction outside the classroom on exam day.

A student's utility is defined as

$$U_i = u\big(Y_i, c_i(e_i, \mathbf{e}_{-i}); \mathbf{X}_i, \mathbf{X}_{-i}, \tilde{\mathbf{K}}, \mathbf{P}_i\big). \tag{2.2}$$

Students derive utility from achievement and disutility from effort. The costs to exerting effort are captured by the term $c_i(e_i, \mathbf{e}_{-i})$ 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 $\mathbf{X}_i$, $\mathbf{X}_{-i}$, $\tilde{\mathbf{K}}$. For instance, a student with highly educated parents may face higher expectations regarding academic performance and thereby derive greater utility from high achievement relative to an otherwise similar student with less educated parents. Equivalently, one could think of these variables as affecting the cost of effort; "good" teachers make effort less costly. $\mathbf{P}_i$ is a variable that affects a student's utility from achievement, but not achievement directly. In the example below, it is an education policy that imposes achievement standards for promotion to the next grade. It is discussed extensively in Section 4.

The utility function (like the production function) permits both contextual and endogenous peer effects. Peer characteristics may enter through $\mathbf{X}_{-i}$. Furthermore, the costs of effort include a "social component," which captures an alternative source of the endogenous peer effect, $\mathbf{e}_{-i}$. Intuitively, peer pressure imposes psychic costs to deviations from the behavioral norm, leading students to seek to conform to the behavior of peers. This type of peer spillover has received a great deal of attention in the broader social interactions literature (Brock and Durlauf (2001b)) and to a lesser extent in the education literature (Bishop, Bishop, Gelbwasser, Green, and Zuckerman (2003)).[6]

[5] Figlio (2007) found empirical evidence of negative externalities from disruptive behavior.

[6] An alternative model may have the utility from achievement depend on the achievement of peers, so that students care more about whether they perform better than others rather than how hard they work relative to others. The implications are similar.

The vector of characteristics $(\mathbf{X}, \tilde{\mathbf{K}}) = (\mathbf{X}_1, \ldots, \mathbf{X}_N, \tilde{\mathbf{K}})$ is common knowledge to all students in the classroom, while $(\theta_i, \boldsymbol{\theta}_{-i})$ are observed ex post. Students possess a common prior on $\theta$, $f(\theta_i|\mathbf{X}, \tilde{\mathbf{K}})$.[7] Suppose $\theta_i$ is defined on the set $\Theta$. Then the expected utility for a given level of effort, $(e_i, \mathbf{e}_{-i})$, is denoted as

$$\tilde{U}_i(e_i, \mathbf{e}_{-i}; \mathbf{X}, \tilde{\mathbf{K}}, \mathbf{P}_i) \equiv \int_{\Theta} U_i(e_i, \mathbf{e}_{-i}; \mathbf{X}, \tilde{\mathbf{K}}, \mathbf{P}_i, \theta_i) f(\theta_i|\mathbf{X}, \tilde{\mathbf{K}})\, d\theta_i.$$

A student chooses effort to maximize his expected utility conditional on his information set. Let the superscript asterisk (*) denote a utility-maximizing choice. The best response $e_i^*(\mathbf{e}_{-i}; \mathbf{X}, \tilde{\mathbf{K}}, \mathbf{P}_i)$ of a student $i$ to a given vector of peer effort is then

$$e_i^*(\mathbf{e}_{-i}; \mathbf{X}, \tilde{\mathbf{K}}, \mathbf{P}_i) \in \arg\max_{e_i} \tilde{U}_i(e_i, \mathbf{e}_{-i}; \mathbf{X}, \tilde{\mathbf{K}}, \mathbf{P}_i). \tag{2.3}$$

A pure strategy Nash equilibrium to the game, $\mathbf{e}^* \equiv (e_1^*, \ldots, e_N^*)$, involves everyone playing their best responses.

The existence of an equilibrium follows from Brouwer's fixed point theorem, given that $e_i^*(\mathbf{e}_{-i}; \mathbf{X}, \tilde{\mathbf{K}}, \mathbf{P}_i)$ is a continuous mapping on the bounded space $[\underline{e}, \overline{e}]^N$. If the cost of effort is (weakly) diminishing in peer effort and peer effort is a (weakly) complementary input to achievement production, $i$'s effort would be (weakly) increasing in peer effort. In this case, there may be multiple equilibria. This is particularly likely when peer spillovers have a strong influence on effort or achievement relative to other inputs, as discussed in Brock and Durlauf (2001a). In the application, I follow much of the literature in assuming that there is only one equilibrium. To address multiple equilibria directly would require specifying an equilibrium selection rule, which is beyond the scope of the present work, though previous studies have shown the potential for multiplicity to aid in identification in some contexts (e.g., Sweeting (2009)).[8] However, the use of rich data helps mitigate the problem, as a richer set of covariates is more likely to predict a unique equilibrium.

If effort were observable, the natural object of interest would be the best response to peer effort. As effort is not observable, assuming that the achievement production function 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 $(Y_1^*, \ldots, Y_N^*)$. Given monotonicity of achievement in effort, such an achievement equilibrium can be described as

$$Y_i^* = q\big(\tilde{\mathbf{Y}}_{-i}^*, \mathbf{X}_i, \mathbf{X}_{-i}, \tilde{\mathbf{K}}, \mathbf{P}_i, \theta_i\big), \tag{2.4}$$

where $\tilde{Y}_i = \int_{\Theta} g(e_i, \mathbf{e}_{-i}; \mathbf{X}, \tilde{\mathbf{K}}, \theta_i) f(\theta_i|\mathbf{X}, \tilde{\mathbf{K}})\, d\theta_i$.[9]

[7]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. It also maps into the simple two-step estimation strategy pursued in this paper.

[8]Recent work by Bisin, Moro, and Topa (2011) suggests a promising alternative way forward that may not rely on specifying a selection rule.

[9]For details, see Appendix A.1, available in a supplementary file on the journal website, http://qeconomics.org/supp/93/supplement.pdf.

Equation (2.4) 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 ($\mathbf{X}$, $\tilde{\mathbf{K}}$, $\mathbf{P}_i$) and unobservables ($\theta_i$).

## 3. Data

I use administrative data for North Carolina public school students from the academic years 1996–1997 to 2001–2002. I focus on reading test scores.[10] The range of test scores varies considerably across grades and years, as does *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

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

where $\mathrm{SD}_g(y_{ig,97} - y_{g,97}^{(3)})$ denotes the standard deviation for a given grade in 1997.[11]

A unique feature of these data is that each student record is linked to a teacher identification (ID) number.[12] This permits the identification of classroom peer groups for grades where student instruction takes place primarily within self-contained classrooms. Thus, I restrict the analysis to elementary students in grades 3–5, where the teacher ID can reliably identify the classroom peer group. I drop the bottom and top percentiles of class sizes to eliminate outliers, though results are robust to their inclusion. 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 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, Hispanic, or American Indian, as these primarily comprise the traditionally disadvantaged racial subgroups in North Carolina; all other students are *white*.[13]

[10] Ideally, I would use math scores as well, as evidence suggests that schools have a larger effect on mathematics achievement (e.g., Rivkin, Hanushek, and Kain (2005)). However, the student accountability policy used for identification coincides with a rescaling of the math test. Even after adjusting for the rescaling, there is a large 1-year spike in math achievement that year. This appears to be more of a data anomaly than real changes associated with student accountability.

[11] The test scores are vertically scaled, so that test scores are meant to be comparable across years. By benchmarking them to a single year, I maintain that comparability and am able to detect changes in mean achievement in response to student accountability, which was introduced in 2001.

[12] In some cases, the data center was unable to reliably identify the teacher; these cases are dropped from the analysis (about 12% of sample).

[13] When there are discrepancies in the student's reported race over time (only 0.5% of the sample), I take the most frequently reported value.

TABLE 1. Summary statistics by race: mixed-race classrooms.[a]

| | 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 | |

[a]Author's calculations using North Carolina Education Research Data Center, end of grade exams. The sample is restricted to grades 4 and 5 and academic years 1997–1998 to 2001–2002. Only classrooms with at least two students of each race are included. All means are statistically significantly different at the 95% confidence level across races.

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.[14] 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 4-year degree (this includes those who received 2-year degrees or obtained some post-secondary vocational training), and (3) those with at least a 4-year degree (this includes those with graduate and professional degrees).

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.[15]

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: 0.48 compared to −0.21. They also

[14]About 30% of the sample changes parental education, though it is not clear whether this is from parents acquiring education, a different parent used for measurement, or mismeasurement of education. I also try using data from grades 6–8, when available, under the assumption that middle schoolers are better able to report parental education. The results are not sensitive to the different specifications.

[15]For the comparison sample, see Appendix Table A.1 in the Supplement.

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 nonwhite, compared to 49% for nonwhites. Furthermore, by all traditional measures, whites are in much "better" peer groups than nonwhites.

## 4. Identification

A growing body of research considers the difficulties associated with identifying peer effects (e.g., Brock and Durlauf (2001b)). The linear-in-means model is the workhorse of the literature, and provides a useful starting point to illustrate these identification problems. The linear-in-means version of the best-response equation (2.4) is

$$Y_i^* = \bar{Y}_{-i}^* \beta_1 + \mathbf{X}_i \boldsymbol{\beta}_2 + \bar{\mathbf{X}}_{-i} \boldsymbol{\beta}_3 + \mathbf{P}_i \boldsymbol{\beta}_4 + \mathbf{K} \boldsymbol{\beta}_5 + \mu + \theta_i, \tag{4.1}$$

where $\bar{Y}_{-i}^*$ captures expected average peer achievement and, similarly, $\bar{\mathbf{X}}_{-i}$ captures average peer characteristics. I further distinguish between classroom characteristics that are observable ($\mathbf{K}$) and unobservable to the econometrician ($\mu$), that is, $\tilde{\mathbf{K}} \equiv (\mathbf{K}, \mu)$. Contextual effects in this model are captured by $\boldsymbol{\beta}_3$ and the endogenous peer effect is captured by $\beta_1$.

The key identification challenge I address in this paper is that $i$'s achievement and peer achievement ($\bar{Y}_{-i}$) are simultaneously determined. Thus, without further assumptions, we cannot separately identify the effect of $i$ on his peers from the effect of $i$'s peers on $i$. Brock and Durlauf (2001a) showed how this nonidentification resulting from simultaneity is a unique feature of the linear-in-means model and does not hold in nonlinear models, such as the binary choice model in their context. However, an even more difficult challenge that arises both in the linear-in-means and the more general specifications is that students share $\mu$. Thus, average peer achievement is correlated with unobserved classroom productivity.

The literature often replaces contemporaneous peer achievement with lagged peer achievement. This has the advantage of eliminating the simultaneity problem, while still capturing a persistent unobservable characteristic of the peer group, such as unobservable ability, that might affect $i$'s achievement (e.g., Hanushek et al. (2003)). The key remaining challenge is then that the predetermined characteristics may not be independent of $\mu$ because of selection into peer groups. I discuss this in Section 4.2, as it is also relevant to my setting for the identification of the spillovers from peer characteristics ($\boldsymbol{\beta}_3$).

A key difference in equation (4.1) from models considered in Manski (1993) and elsewhere in the literature is the existence of $\mathbf{P}_i$ whose peer counterpart $\bar{\mathbf{P}}_{-i}$ does not appear in the model. Assuming that $\boldsymbol{\beta}_4 \neq 0$, $\bar{\mathbf{P}}_{-i}$ provides a potential instrument to identify the endogenous peer effect even in the "worst case scenario" of the linear-in-means setting, given that $E((\mu + \theta_i)\bar{\mathbf{P}}_{-i} | \mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{P}_i, \mathbf{K}) = 0$. As highlighted in Moffitt (2001) and elsewhere, the literature has not proceeded with these types of exclusion restrictions, as in the achievement context it is particularly difficult to define how these exclusions may arise.

The equilibrium model in Section 2 puts some structure on the problem. The model imposes two assumptions:

A1. There exists a variable $P_i$ that affects $i$'s utility from effort, equation (2.2), but does not directly affect achievement production, equation (2.1).

A2. Conditional on $(\mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{K}, \mu)$, $\theta_i$ is independent of $(\mathbf{P}_i, \bar{\mathbf{P}}_{-i})$.

Together assumptions A1 and A2 ensure that there is no direct effect of $\mathbf{P}_j$ on the equilibrium achievement of $i$ for any peer $j \neq i$. The condition in A1 that $\mathbf{P}_i$ cannot enter $i$'s achievement production directly is necessary because of the direct spillovers from effort in achievement production. Intuitively, if $\mathbf{P}_j$ 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.[16]

North Carolina's student accountability policies, which were enacted for fifth graders in the 2000–2001 academic year, provide a potential exclusion. They require that fifth graders perform at the level of sufficient mastery or above (achievement level 3 or above) on standardized end of grade (EOG) exams so as to be automatically promoted to the next grade. This imposes an additional cost to poor performance that could induce students to work harder.

To have any identifying power, the effect of the policy must also differ across 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 because they face increased cost of scoring below achievement level 3. On the other hand, high achievers can effectively disregard the new standards, being fairly confident that they would score above achievement level 3 even with minimal effort.

Figure 1 illustrates how the distribution of fifth grade achievement varies before and after student accountability, using the example of North Carolina's largest school district. 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 student accountability policy. In contrast, the right-hand side figure illustrates little discernable shift over the same years in the distribution of achievement for fourth graders, who were not subject to student accountability.

I find that the retention rate for fifth graders did not increase much over this period, from 0.010 to 0.015. Over the same period, retention of fourth graders increased even less, 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

[16] This can be seen more clearly in Appendix A.1 in the Supplement, which describes the mapping of the effort best response into an achievement best response.

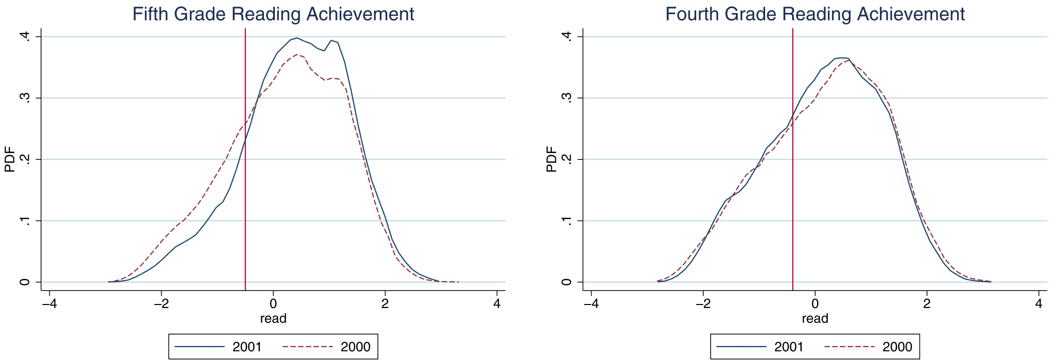


FIGURE 1. Density of reading achievement in 2000 and 2001 for the largest district. The density is calculated using an Epanechnikov kernel "optimal" bandwidth, minimizing the mean integrated squared errors based on a Gaussian distribution. The vertical line indicates the approximate cutoff for passing (achievement level 3) in each grade in 2001. The Kolmogorov–Smirnov test rejects equality of distributions between 2001 and 2000 for fifth grade at the 95% level, but not for fourth grade.

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. The additional tutoring would certainly affect their achievement directly, but under the policy it occurs after the classroom equilibrium achievement is realized.

Given that the policy has a differential effect on low achievers, classrooms with a larger percentage of low achievers would witness a larger shift in average peer achievement under the policy than classrooms with fewer low achievers. Data on fifth graders prior to the implementation of student accountability helps control for any innate differences across classrooms of different compositions (such as teacher quality). Furthermore, because the policy does not apply to fourth graders, they provide a useful control group for any other concurrent changes in policies that might have affected the distribution of achievement similarly across the two grades.

The independence of $\theta_i$ and $\mathbf{P}_i$, $\bar{\mathbf{P}}_{-i}$ imposed under A2 ensures that $\mathbf{P}_{-i}$ does not enter $i$'s expected utility through the distribution of $\theta$. Otherwise, $\mathbf{P}_{-i}$ 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 2000–2001) draw from the same distribution of $\theta$ as similarly low-achieving students in similar peer groups for whom the policy does not apply (fifth graders before 2000–2001 and fourth graders in all years).

The remaining concern is conditional mean independence of $\bar{\mathbf{P}}_{-i}$ and $\mu$. Even with the difference-in-difference type strategy employed here, this may not hold if teachers or administrators redistribute resources disproportionately to low achievers in fifth grade after student accountability policies are enacted. I am not aware of any studies on student accountability policies from which to draw to support this assumption, in part because these student accountability policies generally do not exist in isolation from school accountability. Previous studies show that teachers are very responsive to school accountability, which may also be a cause for concern in this setting. For instance, Jacob (2005), Neal and Schanzenbach (2010), and Reback (2008) found evidence that achievement of marginal and/or lower-achieving students increases as a result. One reason this may be less of a concern in the present 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.

I discuss further the potential for direct teacher responses to the policy in the context of my results in Section 6.1. If student accountability does shift teacher effort toward lower/middle achievement (such as in Jacob (2005), Neal and Schanzenbach (2010), and Reback (2008)), this would suggest that, if anything, my estimates of the effects of peers on achievement are actually biased *downward*.

Importantly, the instrument is still valid if the teacher changes her allocation of effort across students in response to changes in student effort that may have occurred as a result of the policy. For instance, the teacher may just spend more time with students who are more engaged. This definition of a peer effect is useful, as estimates can then be applied to determine the effects of regrouping. An effect of regrouping stems from direct peer effects (deriving through peer effort and characteristics) as well as changes in teacher effort in response to the peer effort. Reduced-form models of peer effects implicitly make a similar assumption, that is, if teachers on average teach differently with more low-achieving students, then this is part of the estimated peer effect of having low-achieving students in the classroom.

### 4.1 *Nonlinear model*

I now show how the identifying assumptions for the linear-in-means model can be extended to a nonlinear context. As discussed above, allowing for nonlinearities in peer spillovers is particularly important for the question of the effects of desegregation, as nonwhite students are more heavily concentrated in the lower tails of the achievement distribution.

As in the above discussion, I simplify the achievement best-response function in equation (2.4) to depend on the expected average peer achievement and average peer characteristics, rather than the entire vector of expected peer achievement and peer characteristics,[17] that is,

$$Y_i^* = q\big(\bar{Y}_{-i}^*, \mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{P}_i, \mathbf{K}, \mu, \theta_i\big). \tag{4.2}$$

Because the function is allowed to be nonseparable in $\theta_i$, it permits a rich picture of the distributional achievement trade-offs associated with peers. I assume that $q(\cdot)$ is strictly

[17]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.

increasing in $\theta_i$, a property that is also satisfied by models that are additively separable in the residual. Since the structural function $q(\cdot)$ is only identified up to positive monotone transformations when the error is nonseparable, I follow the literature on quantile treatment effects in assuming that $\theta_i$ is independently and identically distributed $\mathcal{U}(0,1)$. Since $\theta_i$ is inherently without units, assuming a uniform distribution simply pins down $\theta$. In contrast, the additive model normalizes $\theta_i$ to have the same units as $Y_i$. By fixing $\theta_i = \tau$, equation (4.2) describes the dependence of the $\tau$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 $(Y_i^*, \bar{Y}_{-i}^*, \mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{P}_i, \mathbf{K})$ if there exists a unique $q(\cdot)$ that rationalizes $F(Y_i^*, \bar{Y}_{-i}^* | \mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{P}_i, \mathbf{K})$, the observed joint distribution of achievement and peer achievement conditional on exogenous characteristics.

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

$$\bar{Y}_{-i}^* = h(\mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{P}_i, \bar{\mathbf{P}}_{-i}, \mathbf{K}, \mu). \tag{4.3}$$

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

Equations (4.2) and (4.3) form a triangular system of equations. These equations are comparable to the second and first stages, respectively, of a two-stage least-squares regression for the linear-in-means setting. The following set of assumptions extends the identification argument to the nonlinear setting:

A3. Conditional on $(\mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{K})$, $\mu$ and $\theta_i$ are jointly independent of $\bar{\mathbf{P}}_{-i}$.

A4. With probability 1, $h(\mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{P}_i, \bar{\mathbf{P}}_{-i}, \mathbf{K}, \mu)$ is strictly monotonic in $\mu$.

A5. Conditional on $(\mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{K})$, $\theta_i$ is independent of $\mu$.

The requirement of full independence under assumption A3 is stronger than the mean independence required for the linear-in-means context, but is a necessary tradeoff for identification of the production function under weaker functional form assumptions. Assumption A4 requires that the reduced-form equation for average expected peer achievement (4.3) is strictly monotonic in the unobserved group effect. Note that this is automatically assumed in the linear-in-means model given additive separability of the residual. To fix a value for $\mu$, I assume that it is distributed $\mathcal{U}(0,1)$. Then, given A3 and A4, $\mu$ can be recovered from the first-stage regression as shown in Imbens and Newey (2009, Theorem 1) as $F_{\bar{Y}_{-i}^* | \mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{P}_i, \bar{\mathbf{P}}_{-i}}(\bar{Y}_{-i}^* | \mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{P}_i, \bar{\mathbf{P}}_{-i}) = \mu$.[18]

Given that $\mu$ can be recovered from (4.3), it remains to be shown that the structural function, $q(\cdot)$, is identified. This requires the additional assumption, A5, that $\mu$ is independent of the individual type, $\theta_i$. This assumption is intuitively appealing given that the characteristic $\mu$ is observed to the student, whereas $\theta_i$ is realized ex post.

[18] See Appendix A.1 for details.

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 $\tau$th quantile of achievement on peer achievement conditional on observed characteristics $(\mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{K}, \mathbf{P}_i)$ and the common component $\mu$. Given A3, A4, and A5, $q(\bar{Y}^*_{-i}, \mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{P}_i, \mathbf{K}, \mu, \theta_i)$ is then identified on the joint support of $(\bar{Y}^*_{-i}, \mathbf{X}_i, \bar{\mathbf{X}}_{-i}, \mathbf{P}_i, \mathbf{K}, \mu, \theta_i)$.[19] Intuitively, conditioning on the unobserved group effect $\mu$ controls for the endogeneity of peer achievement, thus identifying the structural function.

### 4.2 *Nonrandom assignment*

The remaining concern for identification is that peer groups are not randomly assigned. 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. I provide support for this assumption in Section 6.1.

Selection is still problematic, however, for the identification of contextual peer effects. For instance, higher income or better educated parents may be more likely to select better teachers. If part of these good teacher attributes are unobservable to the researcher ($\mu$), then we might erroneously conclude that students benefit from being grouped with higher socioeconomic status peers, when the benefit in fact comes from assignment to better teachers. A similar concern arises if parents also select classrooms based on peer characteristics (e.g., Epple and Romano (2010)).

Particularly for the case of elementary school, where students are generally less likely to be tracked by ability, the bulk of nonrandom assignment to peer groups arises from sorting across rather than within schools. For instance, Clotfelter, Ladd, and Vigdor (2003) showed that over the same time period of this study, only about one-fifth of racial segregation arose from within school segregation, with between school sorting accounting for the remaining four-fifths.

To control for time-varying selection into schools, I include school-by-year fixed effects in the form of a location-specific shift, permitting the fixed effects to have different effects across the percentiles of the conditional achievement distribution (described formally in equation (5.2)). Identifying variation derives from plausibly exogenous cross-cohort variation in peer composition, a strategy also pursued by Hoxby (2000), Hanushek et al. (2003), and Lavy and Schlosser (2011), 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.[20]

[19]See Appendix A.1 for details. The proof of this result follows from Imbens and Newey (2009, Corollary 6).

[20]In fact, choosing to focus instead on grade level peer groups would not necessarily eliminate the concern about nonrandom assignment to classrooms within schools. The grade-year outcome is still dependent on the students' peer groupings within schools.

In Section 6.1, I show that estimates of contextual effects do not appear to be biased by nonrandom assignment to classrooms after controlling for school-by-year fixed effects, by considering a sample of schools that appear to randomly assign students to classrooms based on observable characteristics.

## 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 (4.3), 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 (4.2), controlling for the endogeneity of peer achievement by conditioning on the first-stage residual. 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, where students observe something about the classroom that is unobserved to the researcher and hence respond to $\bar{Y}$ rather than the predicted value.

### 5.1 *First stage*

Suppose time is indexed $t = 1, \dots, T$ and classrooms are indexed $c = 1, \dots, C$. 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 $\mathrm{NW}_i$ be an indicator for a nonwhite student, and let the superscripts $k \in \{\mathrm{W}, \mathrm{NW}\}$ indicate white and nonwhite, respectively. Then $\bar{Y}^{\mathrm{NW}}_{-ict} = \frac{1}{\sum_j \mathrm{NW}_j - \mathrm{NW}_i}(\sum_j \mathrm{NW}_j Y^*_j - \mathrm{NW}_i Y^*_i)$ denotes the observed mean achievement of student $i$'s nonwhite classroom peers and similarly $\bar{Y}^{\mathrm{W}}_{-ict}$ for white peers.

The reduced-form equation for achievement of classroom peers of a given race $k$ is approximated as

$$\begin{aligned} \bar{Y}^k_{-ict} = {} & \alpha_0 + \mathbf{X}_{it}\boldsymbol{\alpha}_1 + \bar{\mathbf{X}}_{-ict}\boldsymbol{\alpha}_2 + \mathbf{P}_{it}\boldsymbol{\alpha}_3 + \bar{\mathbf{P}}_{-ict}\alpha_4 + \mathbf{K}_{ct}\boldsymbol{\alpha}_5 \\ & + \mathrm{SchYr}_{it} + \mu_{ct} + \delta_{ict}, \end{aligned} \tag{5.1}$$

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

The covariates $\mathbf{X}_{it}$ include the sex of the student, parental education, and 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. I group achievement levels 1 and 2 because preliminary regressions suggested that these students responded similarly to accountability. The excluded category is achievement level 4, which designates "superior mastery."

$\mathbf{P}_{it}$ indicates students for whom student accountability policies are "binding," that is, fifth graders in 2001 or later who performed below or at achievement level 3 in the prior year. The percentage of peers of each race who are held accountable are the instruments for peer achievement, that is, $\bar{\mathbf{P}}_{-ict} = \{\bar{\mathbf{P}}^{\mathrm{W}}_{-ict}, \bar{\mathbf{P}}^{\mathrm{NW}}_{-ict}\}$.

The mean characteristics of $i$'s peers are captured by $\bar{\mathbf{X}}_{-ict}$, the percentage of peers with college-educated or high-school-educated parents, 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 white–nonwhite peers who were below passing in the prior year and those at achievement level 3 with fifth grade and post-2001 among the $\bar{\mathbf{X}}_{-ict}$. 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 fifth grade independent of student accountability. Thus, the identifying variation for the endogenous peer effect comes from comparing fourth and fifth grade classrooms with similar compositions of low achievers pre- and post-2001.

Other than school-by-year fixed effects, $\mathrm{SchYr}_{it}$, classroom-level inputs $\mathbf{K}_{ct}$ include whether a teacher has an advanced degree (beyond a bachelors), a quadratic in teacher experience, an indicator for years/grades when student accountability policies are in place, and a dummy for fifth grade.

The remaining residual $\delta_{ict}$ is measurement error, that is, that the sample average of observed peer achievement is an approximation for ex ante expectations of average peer achievement ($\bar{Y}_{-ict} = \bar{Y}^*_{-ict} + \delta_{ict}$). Given that classes are sufficiently large, about 23 students on average, $\delta_{ict}$ should be relatively small.

I estimate the two first-stage regressions for white and nonwhite peer achievement separately for students of each race.[21] From these regressions, I recover four estimates of the correlated effect $\hat{\mu}_{ct} = \mu_{ct} + \delta_{ict}$ as the residual from ordinary least-squares estimates of (5.1) and four values of the predicted school-by-year fixed effects, $\widehat{\mathrm{SchYr}}_{it}$.

## 5.2 *Quantile structural function*

In the second stage, I estimate the structural function (4.2), which describes a student's achievement as a function of peer characteristics and peer achievement at different points of the conditional achievement distribution. Previous studies also recognize the importance of capturing these types of nonlinearities, but pursue 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 (e.g., Hanushek et al. (2003), Hoxby and Weingarth (2005), Sacerdote (2011)). 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. While there are advantages to considering how responses vary by observed predetermined student characteristics, the quantile approach

[21]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.

is appealing because it offers considerable flexibility, can be estimated for a large number of quantiles, and is not sensitive to outliers (e.g., Chernozhukov and Hansen (2005)).

While it is feasible to estimate the quantile structural function without assuming a parametric form,[22] I assume a parametric approximation for the system of equations because of the large number of covariates. Therefore, I approximate (4.2) as

$$\begin{aligned} Y^*_{ict} &= \beta_0 + \bar{Y}^{\mathrm{W}}_{-ict}\beta_1 + \bar{Y}^{\mathrm{NW}}_{-ict}\beta_2 + \mathbf{X}_{it}\boldsymbol{\beta}_3 + \bar{\mathbf{X}}_{-ict}\boldsymbol{\beta}_4 + \mathbf{P}_{it}\boldsymbol{\beta}_5 \\ &\quad + \mathbf{K}_{ct}\boldsymbol{\beta}_6 + \widehat{\mathrm{SchYr}}^{\mathrm{W}}_{it}\beta_7 + \widehat{\mathrm{SchYr}}^{\mathrm{NW}}_{it}\beta_8 + \hat{\mu}^{\mathrm{W}}_{ct}\beta_9 + \hat{\mu}^{\mathrm{NW}}_{ct}\beta_{10} + u_{ict}, \end{aligned} \tag{5.2}$$

where dependence of the parameters on the quantile ($\boldsymbol{\beta}(\theta_i)$) and race is suppressed to simplify notation. School-by-year fixed effects are allowed to vary by race, capturing that the effectiveness of the school may vary across races or potential discrimination at the school level. The $\hat{\mu}_{ct}$'s capture the unobserved classroom productivity. These also enter achievement in a flexible way, with the marginal effect permitted to vary both by race and by quantile.

This approximation of the achievement best response predicts a unique equilibrium. While I focus primarily on the effects of the mean of peer achievement to maintain a tighter connection to the existing peer effects literature, I also consider variants where other moments of the peer achievement distribution are included in the achievement best response. For each subgroup and a given quantile $\theta_{it} = \tau$, I estimate $\hat{\boldsymbol{\beta}}(\tau)$ using a quantile regression that minimizes the sum of the weighted absolute value of residuals.

## 6. Results

The first stage results are presented in Table 2. Classes with a larger percentage of peers who are below or just above the threshold for passing in the prior year witness a larger shift in average peer achievement when student accountability policies are introduced. 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 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 with accountability also shifts average peer achievement, but by a smaller magnitude. I test that the instruments are not weak and pass the test of overidentifying restrictions using the simplified case of the mean two-stage least-squares (2SLS) regression.[23]

The shifts associated with student accountability in the first-stage regressions are mirrored at the individual level in the second stage. 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 for both whites and nonwhites. Student accountability

[22] See Imbens and Newey (2009) for a discussion of the fully nonparametric estimator.

[23] To test this, I restrict the school-by-year fixed effects to be the same across races for tractability and use xtivreg2 (Schaffer (2005)) with robust standard errors. The $F$-statistic of 12.38 for whites and 31.84 for nonwhites satisfy conditions that the instruments are not weak. The instruments further pass the test of overidentifying restrictions with a $p$-value of 0.70 for whites and 0.33 for nonwhites.

TABLE 2. First-stage regressions (dependent variable: average white/nonwhite peer reading score).[a]

| | White | | Nonwhite | |
|---|---|---|---|---|
| Dependent Variable | Avg. White | Avg. Nonwhite | Avg. White | Avg. Nonwhite |
| 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 ∗ % nonwhite ach. level 1 or 2 | 0.0295 | 0.1014 | −0.0084 | 0.1488*** |
| | [0.0280] | [0.0618] | [0.0384] | [0.0396] |
| Accountable ∗ % nonwhite ach. level 3 | 0.0050 | 0.0083 | −0.0257 | 0.0023 |
| | [0.0288] | [0.0619] | [0.0369] | [0.0400] |
| % White ach. level 1 or 2 | −1.6392*** | −0.0005 | −1.6337*** | −0.0638***† |
| | [0.0207] | [0.0319] | [0.0334] | [0.0231] |
| % White reading ach. level 3 | −0.7478*** | −0.0612** | −0.6938***† | −0.0325 |
| | [0.0175] | [0.0272] | [0.0277] | [0.0213] |
| % Nonwhite ach. level 1 or 2 | −0.0034 | −1.2998*** | −0.0181 | −1.2262***† |
| | [0.0141] | [0.0294] | [0.0186] | [0.0194] |
| % Nonwhite ach. level 3 | −0.0302** | −0.4817*** | −0.0279 | −0.3787***† |
| | [0.0145] | [0.0303] | [0.0183] | [0.0198] |

[a] Standard errors are given in brackets, clustered at the peer group level. The sample is restricted to fourth and fifth graders for academic years 1997–1998 to 2001–2002. Peer characteristics, individual characteristics, classroom inputs, school-by-year fixed effects, grade fixed effects, and the constant as described in equation (5.1) are included but not shown in the table. *, significant at 10%; **, significant at 5%; ***, significant at 1%. † indicates parameters in nonwhite regression are statistically significantly different from white at the 10% level.

has about twice as large an effect on students below the threshold for passing as on those just above for the mean case (0.16 relative to 0.07 for whites and 0.10 relative to 0.05 for nonwhites).[24] For the median case, the relative effect of student accountability on students below the threshold is even larger: 0.20 relative to 0.06 for whites and 0.16 relative to 0.02 for nonwhites.

Both the 2SLS and the 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 peer achievement, white students derive spillovers almost entirely from their white peers and similarly for nonwhites, though the difference is only statistically significant for nonwhite peer achievement.[25] As I discuss further below, these estimates of

[24] A previous version permits a separate effect for students at achievement levels 1 and 2, but the estimated shift was not statistically significantly different for the two types.

[25] Grouping blacks and Hispanics together as nonwhite is less than ideal, given that they are likely to respond differently to peer pressure, particularly given different language barriers. Because Hispanics make

same-race spillovers are quite sizable in magnitude compared to prior estimates using lagged peer achievement in the literature. In Section 2, I posit 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 effect. While I do not attempt to distinguish between the two mechanisms, the finding that spillovers derive primarily through same-race peers may be more consistent with the conformity mechanism.

Turning to contextual peer effects, 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 −0.06 for the median case). This finding is consistent with prior results in the literature, such as Hanushek, Kain, and Rivkin (2009) and Vigdor and Nechyba (2007), among others. As no income controls are included, the effect of the higher concentration of nonwhites may also proxy 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 of 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.2 and expanded in Fruehwirth (2012), 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 and teacher quality are not included in the table, but are consistent with intuition and prior research.

Figure 2 describes how the marginal effect of average peer achievement varies across quantiles of the achievement distribution for whites and nonwhites. These findings also suggest 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 0.3 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 0.45 for students in the middle and rises slightly for students in the upper quantiles up to 0.6. That lower-achieving students are particularly highly influenced by their (same-race) peers is also supported by other papers investigating nonlinearities by prior peer achievement (e.g., Hanushek et al. (2003), Lavy, Paserman, and Schlosser (2012)), though the literature has reached no consensus (e.g., Gibbons and Telhaj (2008)).

up only 3% of the sample, it is not possible to estimate a specification to break out these three racial groups. However, I reestimate the model using the alternative definition of black/nonblack. While the estimated peer effects are not statistically different from the white–nonwhite specification, the standard errors for nonblack students increase markedly over the standard errors for whites. This suggests that mismeasurement of peer groups may increase the standard errors in the estimated peer effects.

Table 3. Heterogeneous reference groups.[a]

| Dependent Variable: Reading | Mean | | Median | |
|---|---|---|---|---|
| | White | Nonwhite | White | Nonwhite |
| *Endogenous Peer Effects* | | | | |
| Avg. white peer reading | 0.4990*** | 0.1630 | 0.4996** | 0.1730 |
| | [0.1701] | [0.2137] | [0.2535] | [0.2384] |
| Avg. nonwhite peer reading | −0.0019 | 0.6427***† | −0.0500 | 0.5422** |
| | [0.1880] | [0.1693] | [0.2945] | [0.2316] |
| *Contextual Peer Effects* | | | | |
| % White ach. level 1 or 2 | 0.7962*** | 0.2729 | 0.7765* | 0.2877 |
| | [0.2820] | [0.3535] | [0.4212] | [0.3928] |
| % White ach. level 3 | 0.3078*** | 0.1255 | 0.2904 | 0.1422 |
| | [0.1222] | [0.1572] | [0.1822] | [0.1765] |
| % Nonwhite ach. level 1 or 2 | −0.0174 | 0.7798*** | −0.0768 | 0.6547** |
| | [0.2478] | [0.2093] | [0.3839] | [0.2935] |
| % Nonwhite ach. level 3 | −0.026 | 0.1753*** | −0.0432 | 0.1249 |
| | [0.0877] | [0.0629] | [0.1362] | [0.0902] |
| % Nonwhite | −0.0457* | −0.0921***† | −0.0577 | −0.1115*** |
| | [0.0263] | [0.0290] | [0.0437] | [0.0327] |
| % Male | 0.0082 | 0.0812**† | 0.0145 | 0.0559 |
| | [0.0186] | [0.0352] | [0.0378] | [0.0476] |
| % Parents HS degree | −0.0685 | −0.3029***† | −0.0507 | −0.2714*** |
| | [0.0555] | [0.0623] | [0.0938] | [0.0898] |
| % Parents 4-year degree | −0.1624** | −0.4691***† | −0.1366 | −0.4133** |
| | [0.0827] | [0.1472] | [0.1385] | [0.2016] |
| *Policy Variables* | | | | |
| Accountability | −0.0242** | −0.0330* | −0.0210 | −0.0290 |
| | [0.0117] | [0.0191] | [0.0192] | [0.0241] |
| Achievement level 1 or 2 | −1.6660*** | −1.5988***† | −1.6366*** | −1.5905*** |
| | [0.0047] | [0.0061] | [0.0063] | [0.0066] |
| Achievement level 3 | −0.7748*** | −0.7309***† | −0.7502*** | −0.7104*** |
| | [0.0029] | [0.0053] | [0.0040] | [0.0060] |
| Accountable ∗ level 1 or 2 | 0.1604*** | 0.1023***† | 0.1959*** | 0.1550*** |
| | [0.0091] | [0.0101] | [0.0105] | [0.0120] |
| Accountable ∗ level 3 | 0.0740*** | 0.0453***† | 0.0545*** | 0.0220* |
| | [0.0050] | [0.0096] | [0.0068] | [0.0120] |
| $N$ | 344,885 | 207,323 | 344,885 | 207,323 |
| $R^2$ | 0.5951 | 0.5354 | | |

[a] Standard errors are given in brackets, clustered at the peer group level. Standard errors are calculated using 200 bootstrap replications. Individual characteristics, classroom inputs, school-by-year fixed effects, grade fixed effects, and constant as described in equation (5.2) are included but not shown in the table. *, significant at 10%; **, significant at 5%; ***, significant at 1%. † denotes that nonwhite parameter estimates are statistically significantly different from white parameters at the 10% level for mean regression.

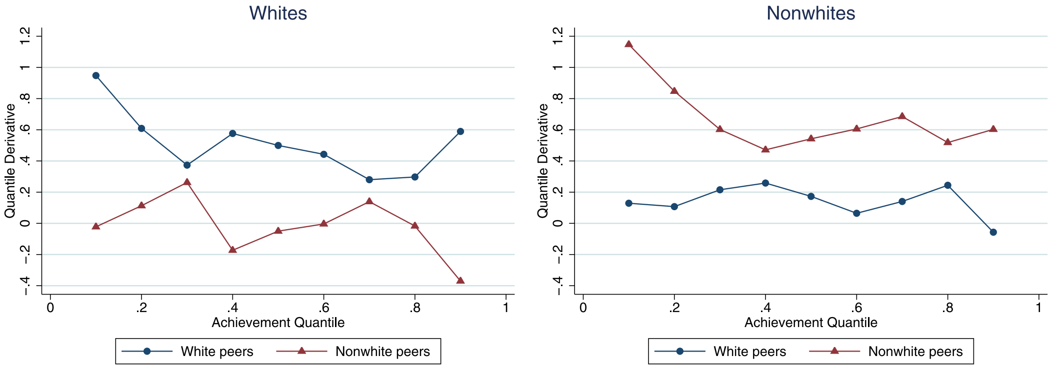


FIGURE 2. Effect of average peer achievement. The marginal effect of average white and nonwhite peer achievement across conditional quantiles is given for white and nonwhite students, using the two-stage quantile regression described in equation (5.2).

To provide some insight into magnitudes, Tables 4 and 5 present marginal effects of a 1 standard deviation increase in each of the peer variables for whites and nonwhites using the estimates from the two-stage quantile regression corresponding to those shown in Figure 2. The second column presents the average over all quantiles within a given race, while the remaining columns present the marginal effect for a given quantile and race. The marginal effect of a 1 standard deviation increase in white peer achievement is 0.22 for whites and 0.07 for nonwhites on average, while the marginal of a 1 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. Furthermore, 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. This suggests that failure to consider contemporaneous spillovers may severely understate the effect of peers, particularly for the lowest quantiles of students. The article by Graham (2008) is the only study, to my knowledge, that has estimated social multipliers in achievement. While he did not break out estimates by race or quantiles, he found that a 1 standard deviation in average peer achievement leads to 0.28 of a standard deviation increase in reading achievement, which is comparable to the estimates above.[26] Furthermore, Graham's (2008) study is arguably free from selection, as he relied on random assignment of students to classrooms from the Project STAR (Student/Teacher Achievement Ratio) experiment. This lends further credence that the sizable peer effects estimated in the present study are not driven by nonrandom assignment.

These estimated peer effects are comparable in magnitude to some of the more important determinants of student achievement found in the literature, such as teacher

[26]This is estimated for the average class size of 22, which is comparable to the average North Carolina class size of 23.

Table 4. Average marginal effects of peers for whites (dependent variable: standardized reading score).[a]

| | Mean | 0.1 Quantile | Median | 0.9 Quantile |
|---|---|---|---|---|
| Avg. white peer reading | 0.2171* | 0.4012*** | 0.2114** | 0.2495 |
| | [0.1326] | [0.1513] | [0.1073] | [0.1921] |
| Avg. nonwhite peer reading | −0.0068 | −0.0113 | −0.0252 | −0.1866 |
| | [0.1929] | [0.2048] | [0.1485] | [0.2954] |
| % White ach. level 1 or 2 | 0.1151 | 0.2226*** | 0.1089* | 0.1281 |
| | [0.0731] | [0.0835] | [0.0591] | [0.1059] |
| % White ach. level 3 | 0.0497 | 0.1066*** | 0.0460 | 0.0512 |
| | [0.0349] | [0.0401] | [0.0288] | [0.0504] |
| % Nonwhite ach. level 1 or 2 | −0.0074 | −0.0026 | −0.0187 | −0.1275 |
| | [0.1214] | [0.1282] | [0.0936] | [0.1867] |
| % Nonwhite ach. level 3 | −0.0076 | −0.0058 | −0.0101 | −0.0491 |
| | [0.0411] | [0.043] | [0.0319] | [0.0633] |
| % Nonwhite | −0.0088 | −0.0062 | −0.0109 | −0.0170 |
| | [0.0101] | [0.0104] | [0.0082] | [0.0149] |
| % Male | 0.0008 | 0.0037 | 0.0012 | −0.0046 |
| | [0.0038] | [0.0038] | [0.0033] | [0.0058] |
| % Parents HS degree | −0.0148 | −0.0318 | −0.0108 | 0.0107 |
| | [0.0251] | [0.0253] | [0.0200] | [0.0383] |
| % Parents 4-year degree | −0.0392 | −0.1030*** | −0.0323 | 0.0028 |
| | [0.0378] | [0.0382] | [0.0327] | [0.0550] |
| Teacher adv. degree | −0.0002 | 0.0009 | −0.0004 | −0.0014 |
| | [0.0025] | [0.0025] | [0.0021] | [0.0037] |
| Teacher experience | 0.0250* | 0.0076 | 0.0228* | 0.0376* |
| | [0.0140] | [0.0140] | [0.0122] | [0.0220] |
| Teacher experience$^2$ | −0.0176* | −0.0073 | −0.0150* | −0.0261 |
| | [0.0105] | [0.0110] | [0.0090] | [0.0169] |

[a]The marginal effects are for a 1 standard deviation increase in the peer variable using the two-stage quantile regression described in equation (5.2). Marginal effects are averaged over quantiles for the second column. *, significant at 10%; **, significant at 5%; ***, significant at 1%.

quality and class size. For instance, Rivkin, Hanushek, and Kain (2005) reported that a 1 standard deviation increase in teacher quality leads to approximately 0.095 of a standard deviation increase in reading achievement. Using findings from Project STAR, they reported 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 (0.40 for whites and 0.49 for nonwhites) is 4–5 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 or for whites at any of the quantiles. In all cases, the effect of a 1 standard deviation

TABLE 5. Average marginal effects of peers for nonwhites (dependent variable: standardized reading score).[a]

| | Mean | 0.1 Quantile | Median | 0.9 Quantile |
|---|---|---|---|---|
| Avg. white peer reading | 0.0681 | 0.0620 | 0.0832 | −0.0274 |
| | [0.1273] | [0.1531] | [0.1147] | [0.1861] |
| Avg. nonwhite peer reading | 0.2845*** | 0.4877*** | 0.2306** | 0.2563 |
| | [0.1201] | [0.1590] | [0.0985] | [0.1703] |
| % White ach. level 1 or 2 | 0.0429 | 0.0438 | 0.0517 | −0.0192 |
| | [0.0780] | [0.0937] | [0.0705] | [0.1134] |
| % White ach. level 3 | 0.0215 | 0.0235 | 0.0272 | −0.0123 |
| | [0.0368] | [0.0436] | [0.0338] | [0.0539] |
| % Nonwhite ach. level 1 or 2 | 0.1716** | 0.3069*** | 0.1373** | 0.1449 |
| | [0.0746] | [0.0981] | [0.0616] | [0.1059] |
| % Nonwhite ach. level 3 | 0.0360* | 0.0785*** | 0.0241 | 0.0247 |
| | [0.0210] | [0.0276] | [0.0174] | [0.0298] |
| % Nonwhite | −0.0191** | −0.0092 | −0.0248*** | −0.0178 |
| | [0.0090] | [0.0139] | [0.0073] | [0.0110] |
| % Male | 0.0075 | 0.0162** | 0.0050 | 0.0028 |
| | [0.0051] | [0.0066] | [0.0043] | [0.0068] |
| % Parents HS degree | −0.0594*** | −0.0963*** | −0.0530*** | −0.0473* |
| | [0.0199] | [0.0272] | [0.0175] | [0.0248] |
| % Parents 4-year degree | −0.0980** | −0.1661*** | −0.0863** | −0.0599 |
| | [0.0462] | [0.0598] | [0.0421] | [0.0575] |
| Teacher adv. degree | −0.0013 | 0.0012 | −0.0034 | −0.0001 |
| | [0.0024] | [0.0032] | [0.0021] | [0.0031] |
| Teacher experience | 0.0078 | −0.0173 | 0.0126 | 0.0182 |
| | [0.0194] | [0.0260] | [0.0161] | [0.0241] |
| Teacher experience$^2$ | −0.0041 | 0.0109 | −0.0073 | −0.0092 |
| | [0.0138] | [0.0186] | [0.0112] | [0.0171] |

[a]The marginal effects are for a 1 standard deviation increase in the peer variable using the two-stage quantile regression described in equation (5.2). Marginal effects are averaged over quantiles for the second column. *, significant at 10%; **, significant at 5%; ***, significant at 1%.

increase in percentage nonwhite is much smaller in magnitude than a 1 standard deviation increase in average peer achievement of the same race. Figure 3 further 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. To test this, I estimate the effect of different quantiles of overall classroom peer achievement on

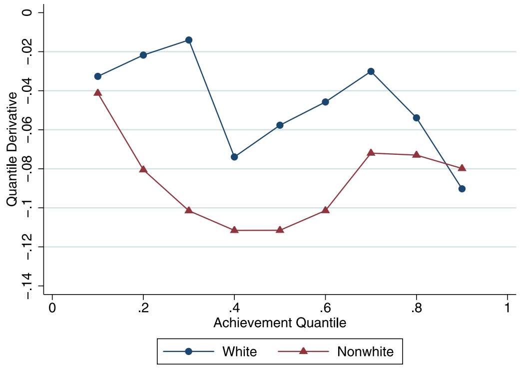


FIGURE 3. Effect of percentage nonwhite. The marginal effect of the percentage nonwhite peers across conditional quantiles is given for white and nonwhite students, using the two-stage quantile regression described in equation (5.2).

students at different quantiles of the achievement distribution.[27] 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 responds most to the median of peer achievement, and so forth. This is not the case. I find that the lower-achieving students are influenced more (relative to students at higher quantiles) from all quantiles of peer achievement (the 25th, median, and 75th percentiles). Students at the median benefit most from increases in median achievement, while students at the upper quantile are not affected by increases in the upper quantile of peer achievement but only by the median.

While these patterns suggest that the within-race spillovers are not driven primarily by achievement disparities across races, I also estimate a specification that replaces the mean peer achievement of each race with the 25th, median, and 75th 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 the only statistically significant spillovers in this specification come from within race peers, regardless of the moment of the peer achievement distribution. 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 75th 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.

Evidence in Fryer and Torelli (2010) and elsewhere suggests that peer effects for nonwhites vary based on the percentage of nonwhite students in the classroom. To test this, 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

[27]The instruments, and the percentage of students below and just above the threshold, interacted with student accountability, are significant predictors of the different quantiles of peer achievement, even the 75th percentile.

surprisingly, splitting the sample creates noisier estimates, but I still do not find any evidence of cross-racial spillovers in classes with either larger percentages or smaller percentages of nonwhite students.

### 6.1 *Sensitivity analysis*

*Instrumental variable* As discussed in Section 4, the instrument is not valid if it captures a response of teachers rather than a response of students. As mentioned above, my peer effect estimates are biased *downward* if teachers respond in ways predicted by studies on school accountability (e.g., Jacob (2005), Neal and Schanzenbach (2010), Reback (2008)). If teachers redistribute attention to lower-achieving or marginal students in proportion to the percentage of these students, then part of the increase in peer achievement arises from teacher effort. Recall that because I control 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 achievers and is a direct response to the policy (rather than an indirect response to the change in student effort as a response to the policy).

The present evidence provides some support that the peer effects are unlikely to be driven by teacher responses. It would be difficult to reconcile my findings of within-race peer spillovers with the alternative of teacher shifts in effort. 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.

Furthermore, we might expect teachers and schools to be forward looking in their response to student accountability by making any resource shift apply to all grades. Returning to Figure 1, the distribution of achievement for fourth graders, who were not held to the new accountability standards in either year, remains similar across the 2 years, suggesting that this is not the case. To the extent that any resource shifts occur in all grades, this helps my identification strategy, as I control for a different effect of percentage low achievers in 2001 and later.

I also look for evidence of direct responses of teachers to student accountability. First, if teachers face additional pressure to respond to student accountability, we might observe increased turnover of fifth grade teachers relative to fourth grade teachers. In Table 6, I define "turnover" first to be that a teacher either switches grades, switches schools, or leaves the sample (columns 1 and 2), and, second, use a stricter definition that a teacher changes schools or leaves the sample (columns 3 and 4). Columns 1 and 3 present a difference-in-difference analysis that examines whether fifth grade teachers (who are subject to student accountability) move relatively more after the introduction of student accountability than fourth grade teachers, after controlling for school fixed effects. I do not find evidence of this. Thus, in contrast to evidence showing increased teacher turnover in response to school accountability in North Carolina (Clotfelter, Vigdor, Ladd, and Diaz (2004)), there does not appear to be a similar teacher response to student accountability.

Table 6. Teacher mobility in response to student accountability.[a]

| | Change Grade or School (1) | Change Grade or School (2) | Change School (3) | Change School (4) |
|---|---|---|---|---|
| Student accountability | 0.079 | 0.252 | 0.081 | 0.248 |
| | [0.050] | [0.205] | [0.051] | [0.206] |
| Accountable × % ach. level 1 or 2 | | 0.010 | | −0.177 |
| | | [0.352] | | [0.354] |
| Accountable × % ach. level 3 | | −0.548 | | −0.422 |
| | | [0.417] | | [0.420] |
| % Ach. level 1 or 2 | | 0.625 | | 0.606 |
| | | [0.181]*** | | [0.182]*** |
| % Ach. level 3 | | −0.062 | | 0.026 |
| | | [0.209] | | [0.211] |
| *N* | 28,570 | 28,570 | 28,619 | 28,619 |

[a]Logit regressions run at the teacher level for 1998–2002. School, year, and grade fixed effects are included. In columns 2 and 4, interactions of grade 5 and post-2001 dummy variables with percentage at achievement level 1 or 2 and percentage at achievement level 3 are included. ***, significant at 1%.

Teachers who teach higher percentages of low achievers post-student accountability may face relatively more pressure. Columns 2 and 4 investigate whether there is increased mobility tied to student accountability for fifth grade teachers who have high percentages of students at achievement levels 1 or 2 relative to fourth grade teachers in the same period or fifth grade teachers in prior years, using a similar difference-in-difference strategy. The results show no statistical difference in mobility for teachers who face higher percentages of low achievers post-student accountability.

Schools might also respond in how they assign students to classrooms. I calculate a dissimilarity index to see whether the composition of classes changed following accountability, that is, whether there is a higher tendency to group certain types of students together. Formally, for a given observable characteristic, such as nonwhite (NW) or white (W), the dissimilarity index is calculated as $\sum_c |(\frac{T_c^{\text{NW}}}{T_s^{\text{NW}}} - \frac{T_c^{\text{W}}}{T_s^{\text{W}}})|$, where $T_c^{\text{NW}}$ refers to the total number of students in the class who are nonwhite and $T_s^{\text{NW}}$ refers to the total number of students in the school, grade, and year who are nonwhite (and similarly for white). I also calculate dissimilarity indices for males and females, students whose parents have a 4-year college degree or more of education versus those who do not and students at achievement levels 1 and 2 (not proficient) relative to achievement levels 3 and 4 (proficient) based on prior year test scores. Table 7 presents results from a difference-in-difference analysis that considers whether dissimilarity changed when students were subject to accountability. There is no evidence in any of the cases of regrouping in response to student accountability.

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. Table 8 shows results from the difference-in-difference estimation of

TABLE 7. Dissimilarity index.[a]

| | Nonwhite (1) | Male (2) | Parent 4 Year+ (3) | % Achievement Level 1 or 2 (4) |
|---|---|---|---|---|
| Student accountability | 0.002 | 0.001 | 0.012 | −0.008 |
| | [0.011] | [0.005] | [0.012] | [0.010] |
| $N$ | 552,208 | 552,208 | 552,208 | 552,208 |
| $R^2$ | 0.43 | 0.20 | 0.38 | 0.34 |

[a]School level regressions, 1998–2002, weighted by school size. School, year, and grade fixed effects are included. Each column corresponds to a separate regression, where the dependent variable is the dissimilarity index for the given category: nonwhite, male, parent 4-year degree or more, and % at achievement level 1 or 2 (based on prior year test scores).

TABLE 8. Accountability by class size.[a]

| | Overall (1) | Interacted With Small Class (2) |
|---|---|---|
| Accountability | 0.005 | −0.002 |
| | [0.008] | [0.010] |
| Accountable ∗ ach. level 1 or 2 | 0.138 | −0.035 |
| | [0.016]*** | [0.020]* |
| Accountable ∗ ach. level 3 | 0.080 | −0.005 |
| | [0.011]*** | [0.014] |
| Ach. level 1 or 2 | −1.86 | 0.023 |
| | [0.007]*** | [0.009]*** |
| Ach. level 3 | −0.888 | −0.001 |
| | [0.005]*** | [0.007] |
| $N$ | 552,208 | |
| $R^2$ | 0.6 | |

[a]The two columns refer to one stacked regression with school-by-year fixed effects for years 1998–2002. Column 2 reports variables interacted with small class size, defined as a class size smaller than the median size of 23. Included, but not reported, are an indicator for fifth grade, a constant, and achievement levels interacted with fifth grade and post-2001. *, significant at 10%; ***, significant at 1%.

the effect of student accountability across large and small classrooms (above and below the median class size of 23).[28] While students who are low achieving in $t-1$ do see relatively more benefits to their achievement in $t$ if they are in small classrooms (relative to large classrooms), they do not realize similar benefits with the introduction of student accountability. Estimates show that low achievers are, instead, marginally worse off in small relative to large classes after student accountability is introduced. This is unlikely to be because of differential resources, given that fourth grade acts as a control group and because of the inclusion of school-by-year fixed effects. Thus, contrary to what might be expected if the effect were deriving through teacher effort, if anything,

[28]These results are robust to more extreme measures of small versus large classrooms.

the achievement for low-achieving students increases more for students in *larger* classrooms with student accountability.

Finally, if findings are driven by teacher responses, we might also expect more experienced teachers or teachers with more training to be better able to raise achievement when student accountability policies are in place. To test for this, I take the main student achievement model described in Table 3 and interact teacher characteristics with student accountability. I also control for the potential that teacher characteristics matter differentially in fifth grade and/or after 2001. The interactions with student accountability are insignificant, suggesting that more able teachers (in measurable dimensions) are not more successful in raising achievement post-student accountability.

*Contextual effects: Sorting* 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. Recall that the instrumental variable strategy alleviates this concern for endogenous effects. To explore potential bias in 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 to classrooms based on observable characteristics.[29] 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 0.1 or the school has only one classroom per grade. This is about 72% of the schools in my sample.[30]

I reestimate the mean version of the peer effects model for the subsample of schools that appear to assign students to classrooms at random. Table 9 shows that the estimated peer effects (both contextual and endogenous) are qualitatively and quantitatively similar to the estimates on the main sample (Table 3).

To further consider potential endogeneity of contextual effects, I reestimate the mean regression from Table 3 using teacher and year fixed effects (rather than school-by-year fixed effects). Arguably, if teachers face similar peer groups over time, they might provide a better control for nonrandom assignment to classrooms. As shown in columns 3 and 4 of Table 9, results are qualitatively similar, though standard errors are much larger. I chose not to pursue the strategy of using teacher fixed effects, out of concern that it is not a sufficiently long panel of teachers.

The above tests provide supportive evidence that selection into classrooms within schools is not biasing estimates of contextual effects.

*Comparison to literature* As a point of comparison to previous literature, I also include estimates of the linear-in-means model that focus on average peer achievement but are not broken out by race. As shown in column 1 of Table 10, these results are comparable in magnitude to the within-race spillovers from peer achievement, 0.52. Column 2

[29] Vigdor and Nechyba (2009) used a similar intuition, as did Lavy and Schlosser (2011) in their balancing tests.

[30] As shown in Appendix Table A.2, this subset of schools is remarkably similar in terms of observables to the main sample.

TABLE 9. Mean regression: robustness.[a]

| | Apparent Random Assignment | | Teacher Fixed Effects | |
|---|---|---|---|---|
| | White (1) | Nonwhite (2) | White (3) | Nonwhite (4) |
| Avg. white peer reading score | 0.6018*** | 0.0729 | 0.4478 | 0.2408 |
| | [0.1351] | [0.1953] | [0.2135] | [0.1741] |
| Avg. nonwhite peer reading score | −0.0668 | 0.6871*** | −0.0584 | 0.4709 |
| | [0.1220] | [0.1413] | [0.2340] | [0.3365] |
| % White ach. level 1 or 2 | 0.9718*** | 0.1377 | 0.7328 | 0.3870 |
| | [0.2206] | [0.3285] | [0.3485] | [0.2769] |
| % White ach. level 3 | 0.3851*** | 0.0684 | 0.2788 | 0.1781 |
| | [0.0977] | [0.1440] | [0.1456] | [0.1249] |
| % Nonwhite ach. level 1 or 2 | −0.1015 | 0.8354*** | −0.0836 | 0.5581 |
| | [0.1605] | [0.1811] | [0.3082] | [0.4239] |
| % Nonwhite ach. level 3 | −0.0582 | 0.1834*** | −0.0377 | 0.1131 |
| | [0.0540] | [0.0552] | [0.1052] | [0.1239] |
| $N$ | 250,915 | 145,638 | 344,885 | 207,323 |
| $R^2$ | 0.5932 | 0.5370 | 0.6078 | 0.5530 |

[a]Standard errors are given in brackets, clustered at the peer group level. Columns 1 and 2 are the same regression as in Table 3, but restricted to schools that apparently randomly assign students to classrooms based on observables. Columns 3 and 4 include the same controls as in Table 3, but instead of school-by-year fixed effects, include teacher and year fixed effects. ***, significant at 1%.

shows that when, instead, lagged peer achievement (rather than contemporaneous) is included in the regression, the estimated peer effects are quite small in magnitude, 0.02, which is comparable to other findings using lagged peer achievement in the literature. Furthermore, column 4 compares estimated spillovers from contemporaneous peer achievement using grade-level peer effects. These estimates are slightly larger in magnitude, 0.68. In principle, it is unclear whether these should be larger or smaller, given that classroom peer groups may abstract away from important spillovers outside of the classroom or grade-level peer groups may dilute the effect if peers outside the classroom have no effect on student achievement. Together this evidence shows that the large peer spillovers are driven by the focus on contemporaneous peer achievement, instead of the choice to focus on the class instead of the grade peer group or to focus on race-specific peer groups.

The literature widely acknowledges that lagging peer achievement, that is, using a specification like that in column 2, does not solve the reflection problem. As well described by Ammermueller and Pischke (2009) and others, lagging the endogenous variable is equivalent to estimating a reduced form of the structural equation described in Section 4, equation (4.1). The coefficient on lagged peer achievement, instead of capturing the social multiplier or endogenous peer effect, captures a reduced form or "social" effect, in the language of Manski (1993), similar to other contextual variables in the reduced-form specification. Thus, the parameter can inform whether peer effects exist, but does not distinguish whether they derive through endogenous behaviors or exogenous characteristics.

TABLE 10. Mean regression class and grade peers ($N = 552{,}208$).[a]

| | Contemp. (1) | Class Lagged (2) | No Endog. (3) | Grade Contemp. (4) | Class Gains (5) |
|---|---|---|---|---|---|
| Average peer reading | 0.5247*** | 0.0276*** | | 0.6802*** | 0.453*** |
| | [0.1132] | [0.0104] | | [0.1451] | (0.108) |
| % Nonwhite | 0.0978** | −0.1064*** | −0.1118*** | 0.1117 | −0.0436** |
| | [0.0462] | [0.0138] | [0.0138] | [0.0744] | (0.0195) |
| % Male | 0.0231 | −0.0493*** | −0.0513*** | 0.0021 | −0.0307*** |
| | [0.0177] | [0.0131] | [0.0132] | [0.0337] | (0.0116) |
| % Parents HS degree | −0.1591*** | −0.0174 | −0.0125 | −0.1938*** | −0.0343** |
| | [0.0344] | [0.0181] | [0.0181] | [0.0412] | (0.0148) |
| % Parents 4-year degree | −0.2916*** | 0.0653*** | 0.0766*** | −0.4383*** | −0.0142 |
| | [0.0825] | [0.0205] | [0.0203] | [0.0558] | (0.0262) |
| % Reading ach. level 1 or 2 | 0.7752*** | −0.01 | −0.0708*** | 1.2350*** | −0.273*** |
| | [0.1831] | [0.0289] | [0.0182] | [0.1538] | (0.0505) |
| % Reading ach. level 3 | 0.2843*** | −0.0783*** | −0.1031*** | 0.4685*** | −0.140*** |
| | [0.0847] | [0.0194] | [0.0171] | [0.0713] | (0.0162) |
| Grade 5 × % ach. level 1 or 2 | −0.0193 | 0.0112 | 0.0079 | −0.0346 | −0.0690*** |
| | [0.0124] | [0.0158] | [0.0159] | [0.0273] | (0.0228) |
| Grade 5 × % ach. level 3 | 0.009 | 0.1083*** | 0.1081*** | 0.0053 | 0.0194 |
| | [0.0233] | [0.0172] | [0.0173] | [0.0442] | (0.0255) |
| Post-2001 × % ach. level 1 or 2 | −0.0932*** | −0.0725*** | −0.0731*** | −0.2123*** | −0.105*** |
| | [0.0150] | [0.0231] | [0.0232] | [0.0646] | (0.0222) |
| Post-2001 × % ach. level 3 | −0.0140 | −0.0527*** | −0.0550*** | −0.0044 | −0.0592*** |
| | [0.0145] | [0.0201] | [0.0203] | [0.0603] | (0.0173) |

[a]Standard errors are given in brackets, clustered at the peer group level. Not shown in table but included in regression are dummy variables for male, parent with high school degree, parent with 4-year degree, student accountability, lagged achievement level (also interacted with student accountability), teacher experience, experience$^2$, and teacher advanced degree. School-by-year fixed effects, grade fixed effects, and the constant also are included. Columns 1, 4, and 5 are estimated via two-stage least squares, with student accountability interacted with percentage at achievement levels 1 or 2 and level 3 in $t-1$ as the instrumental variable. Standard errors are calculated using 200 bootstrap replications. Columns 1–3 and 5 use class-level peer groups; column 4 uses grade level. **, significant at 5%; ***, significant at 1%.

The existing literature offers little guidance as to how to interpret lagged peer achievement in an equation that also controls for contemporaneous peer achievement. Given that my specification includes both lagged measures of peer achievement (through prior peer achievement levels), the question may be raised whether this specification really solves the reflection problem. My preferred interpretation is that contemporaneous peer achievement captures current behaviors that are simultaneously determined in equilibrium, and thus including lagged measures of peer achievement is like including other predetermined characteristics, that is, the $\bar{\mathbf{X}}$ in equation (4.1). However, one might be concerned that these variables are still endogenous in my main specification, beyond the concerns about nonrandom sorting addressed above. This is potentially particularly troubling given that prior peer achievement makes up part of my instrumenting strategy for identifying contemporaneous achievement spillovers.

I check that results are robust to these concerns in two ways. First, I reestimate the baseline regression using average gains in peer achievement rather than levels. The gains difference out the potential endogeneity of prior peer achievement. Column 5 of Table 10 shows that the effect of peer gains on achievement is not statistically significantly different from the average level effect reported in column 1: 0.45 compared to 0.52.

What does change across these two specifications is the sign of the contextual peer effects, including lagged peer achievement levels. The intuition for this is also straightforward. As discussed earlier, after conditioning on average peer achievement, higher levels of peer characteristics predict lower levels of peer effort. In column 1 this "negative" pressure deriving from the contextual effects partially proxying for peer effort appears to swamp the potential direct effect of these characteristics on achievement, leading to an apparently "counterintuitive" sign. In the model where classroom gains are used, the effect of this downward pressure is diminished (though arguably still present absent strong functional form assumptions) because prior peer achievement is picking up some of this negative proxy effect. Thus, the direct effect dominates and the contextual effect takes the sign generally expected in the literature. Importantly, this does not suggest that the sign of the contextual effects estimated in column 1 are wrong; rather they are entirely consistent with the theoretical model of peer achievement spillovers developed in the paper.

The second robustness check estimates contemporaneous peer achievement effects without conditioning on prior peer achievement. This helps us to check whether using prior peer achievement as part of the identification strategy is biasing estimates of the spillovers from contemporaneous peer achievement, for instance, through serial correlation in achievement. Recall that at the most basic level, my identification strategy centers around the assumption that high achievers are not affected directly by student accountability, whereas lower achievers who face the risk of failing are induced to work harder. Under this assumption, I can identify the effect of average peer achievement on high achievers because student accountability affects their achievement only through its effect on the effort of their classmates. In Table 11, I estimate the peer effect using this alternative strategy.

I must first define the high achievers who are not directly affected by the policy. I look for a threshold for reading achievement that keeps as much of the sample as possible, but shows no evidence of a direct effect of accountability on the high achievers. I find that setting the threshold slightly above the midpoint in achievement level 3 (in particular, a threshold of 0.7 between the lower and upper bound reading score for achievement level 3, or an average reading score of approximately 0.26) produces this result, as shown in column 1. This threshold is intuitively appealing, as students who are lower performers in achievement level 3 may still work hard to avoid failing, whereas students who performed comfortably well above the threshold would not be directly affected by the policy. Results are robust to increases in this threshold, though, as expected, they become noisier as the sample size shrinks.

TABLE 11. Mean regression with accountability as the instrument.[a]

| | Overall | High Achievers | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Average peer reading | | 0.514*** | 0.452*** |
| | | (0.0809) | (0.0681) |
| Accountability | −0.00273 | | |
| | (0.00476) | | |
| Accountable × low | 0.136*** | | |
| | (0.00471) | | |
| Low achieving | −1.091*** | | |
| | (0.00234) | | |
| % Nonwhite | −0.199*** | 0.229*** | 0.0221 |
| | (0.0150) | (0.0720) | (0.0327) |
| % Male | −0.0761*** | 0.0650*** | 0.00472 |
| | (0.0144) | (0.0249) | (0.0173) |
| % Parents HS degree | 0.0212 | −0.220*** | −0.101*** |
| | (0.0200) | (0.0470) | (0.0287) |
| % Parents 4-year degree | 0.181*** | −0.430*** | −0.124** |
| | (0.0218) | (0.111) | (0.0522) |
| % Reading ach. level 1 or 2 | | | 0.537*** |
| | | | (0.107) |
| % Reading ach. level 3 | | | 0.207*** |
| | | | (0.0518) |
| $N$ | 552,208 | 258,085 | 258,085 |

[a]Standard errors are given in brackets, clustered at the peer group level. The regression also includes dummy variables for male, parent with high school degree, parent with 4-year degree, teacher experience, experience$^2$, and teacher advanced degree. School-by-year fixed effects, grade fixed effects, and the constant also are included. Columns 2 and 3 are estimated via two-stage least squares, with student accountability as the instrumental variable. Standard errors are calculated using 200 bootstrap replications. Let lb denote the lower bound of reading achievement at achievement level 3 and let ub denote the upper bound. "Low-achieving" students are defined as those who are not high achieving by the previous definition. The threshold for "high" is students with prior reading achievement $\geq \text{lb} + 0.7 * (\text{ub} - \text{lb})$. Columns 2 and 3 are estimated on the subset of high students. **, significant at 5%; ***, significant at 1%.

Column 2 estimates the effect of average peer achievement on the reading achievement of high achievers, with no control for prior peer achievement and using student accountability as an instrument. The estimated effect of average peer achievement is 0.51, which is almost identical to the estimate of 0.52 using the primary identification strategy on the whole sample, as shown in column 1 in Table 10. Furthermore, controlling for lagged peer achievement levels does not lead to statistically significantly different estimates of the contemporaneous peer effect, as shown in column 3. Overall, these results provide strong support that conditioning on prior peer achievement levels and using them as part of the instrumenting strategy is not biasing estimates of the endogenous peer effect.

## 7. Desegregating peer groups

While the estimates above show that peer effects are important determinants of student achievement, it remains difficult to infer an actual effect of desegregation directly from the parameter estimates. Intuitively, the findings suggest that the potential gains from desegregation would be limited by lack of cross-racial spillovers. To the extent that desegregation creates more mixed-ability classrooms, we might also expect efficiency gains in terms of average achievement, given that lower-achieving students benefit relatively more from peer achievement than higher-achieving students. The magnitude of the gains is difficult to determine because of the need to account for social multiplier effects and changes in peer characteristics (though these effects are much smaller in magnitude).

Often peer effect studies attempt to infer an effect of desegregation from the reduced-form effects of percentage nonwhite holding other observable predetermined peer characteristics fixed. These strategies do not take into account the joint distribution of student characteristics, in that it would be impossible to change only the racial composition of classrooms and hold other peer characteristics fixed when these are correlated with race. This problem was also discussed by Graham, Imbens, and Ridder (2010). The challenge is similar in my context, but the potential effect is even more complicated because of social multipliers.

Furthermore, though the peer effects literature commonly assumes that reduced-form estimates of peer effects are sufficient to determine the effects of regrouping policies, Fruehwirth (2012) discussed why this may not be the case in many settings, like the current one, when there is likely to be matching between students and unobserved school quality in the data. Intuitively, holding resources fixed at some level means that nonwhite (white) students might receive higher (lower) resources on average than in the initially observed assignment. If this reallocation of resources creates social multiplier effects (that vary based on the composition of the classroom), then we need estimates of the social multiplier to separate an effect of racial integration from a resource effect. Importantly, this follows even though the reduced-form estimator obtains consistent estimates of the social effect of peers.

In the example below, I focus on fifth graders in 2001–2002. I consider an experiment of desegregating schools in Durham, a racially diverse school district in North Carolina (and home of Duke University). Table 12 shows average characteristics of classrooms for white and nonwhite students. Durham public schools have a large minority population (61% of fifth graders are nonwhite) and are also fairly segregated. The average class for a white Durham student is 46% nonwhite, while the average class for a nonwhite Durham student is 71% nonwhite.

To quantify the total effect of desegregation, I simulate the effect of creating racially diverse peer groups. As a baseline for comparison, I first estimate a predicted achievement, holding resources fixed at the average level for all students using observed peer groups. I then estimate the equilibrium achievement when students are randomly assigned to peer groups (effectively integrating classrooms), holding resources fixed at the average level. I use the parameter estimates from the two-stage quantile regression pro-

TABLE 12. Average characteristics of Durham public schools (grade 5, academic year 2001–2002, $N = 1685$).[a]

| Variable | Mean | Std. Dev. |
|---|---|---|
| Reading | 0.3431 | 0.8759 |
| % Nonwhite | 0.6142 | 0.4869 |
| % Parent with HS degree | 0.5887 | 0.4922 |
| % Parent with 4-year degree+ | 0.3697 | 0.4829 |
| *Nonwhite Classroom Characteristics* | | |
| Avg. white peer reading | 0.6108<br>[0.5087, 0.7462] | 0.5276 |
| Avg. nonwhite peer reading | 0.0266<br>[−0.0577, 0.0794] | 0.3262 |
| % Nonwhite | 0.7080<br>[0.7443, 0.6535] | 0.2243 |
| % Parent with HS degree | 0.6374<br>[0.6684, 0.4921] | 0.2049 |
| % Parent with 4-year degree+ | 0.3069<br>[0.2710, 0.4475] | 0.2264 |
| *White Classroom Characteristics* | | |
| Avg. white peer reading | 0.7959<br>[0.4960, 0.8755] | 0.4025 |
| Avg. nonwhite peer reading | 0.1355<br>[0.0186, 0.2518] | 0.4342 |
| % Nonwhite | 0.4605<br>[0.5147, 0.4374] | 0.2380 |
| % Parent with HS degree | 0.5022<br>[0.5598, 0.4762] | 0.1921 |
| % Parent with 4-year degree+ | 0.4647<br>[0.3809, 0.4918] | 0.1999 |

[a]Numbers in square brackets denote average classroom peer characteristics for students in the 10th and 90th percentiles of the unconditional achievement distribution.

cedure described in Section 5 and in Tables 4 and 5.[31] 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.

The left-hand panel of Figure 4 shows the change in achievement relative to predicted achievement from observed groupings (holding resources fixed) from randomly assigning students to peer groups in Durham. The figure describes average gains for students at given percentiles of the initial achievement distribution. Desegregation produces large gains for the low-achieving Durham students (as much as 0.75 standard deviations for the lowest-achieving nonwhite students and about 0.6 of a standard devia-

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

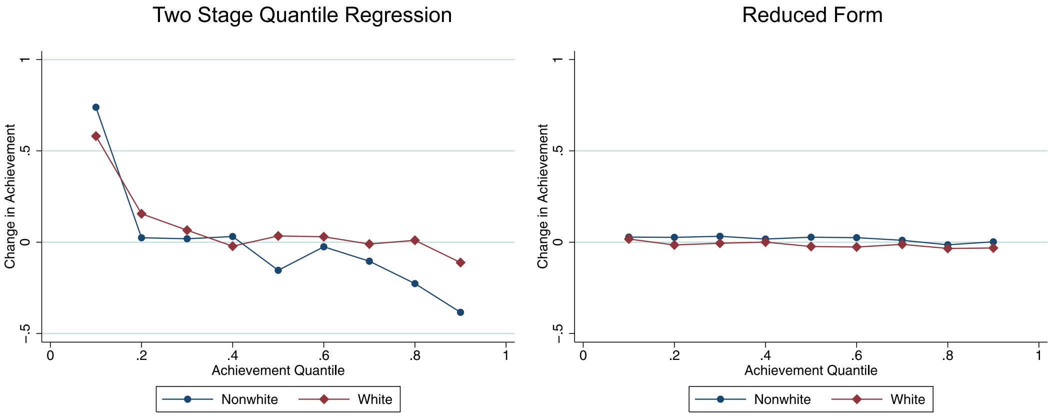


FIGURE 4. Achievement gains from desegregating Durham. Predictions are based on assigning each student the average school by year fixed effect and teacher quality, and randomly assigning fifth graders in 2001–2002 to classrooms in Durham. The $x$-axis refers to the student's position in the unconditional achievement distribution under observed groupings (but with equal resources). The $y$-axis denotes average changes for students at a given percentile of the initial achievement distribution. The left-hand panel uses two-stage quantile regression parameter estimates, as described in equation (5.2). The right-hand panel uses reduced-form parameter estimates, as described in equation (7.1).

tion for the lowest-achieving white students). Low achievers likely have the most to gain, in part because they generally suffer initially from lower "quality" peer groups (Table 12). Lower-achieving students also have the most to gain from improvements in peer quality given the shape of peer achievement spillovers evidenced in Figure 2.

In contrast, the highest-achieving students experience small losses to achievement (about −0.1 of a standard deviation for whites and −0.4 for nonwhites). The losses are smaller than for the lowest-achieving students, in part because of the smaller marginal effects derived from improvements in peer quality for higher-achieving students (Figure 2). On average, white students gain about 0.02 and nonwhites gain about 0.07 from desegregation.

Translating these estimates into effects on the achievement gap, the left-hand panel of Figure 5 shows changes in the achievement gap between whites and nonwhites at the different percentiles of the achievement distribution after desegregation. While the gap narrows by a little more than 0.1 of a standard deviation at the 10th percentile, it increases by about 0.8 of a standard deviation at the 90th percentile. The effects at the lower percentiles are driven by gains to nonwhites and whites. In contrast, the effect at the upper percentiles is driven by losses to both whites and nonwhites. Given these disparate gains and losses across the percentiles and districts, the overall achievement gap narrows only by 0.06 of a standard deviation, on average, from desegregation.

To highlight the importance of endogenous effects, I compare the above estimates to predictions from the flexible reduced-form quantile estimator using lagged peer

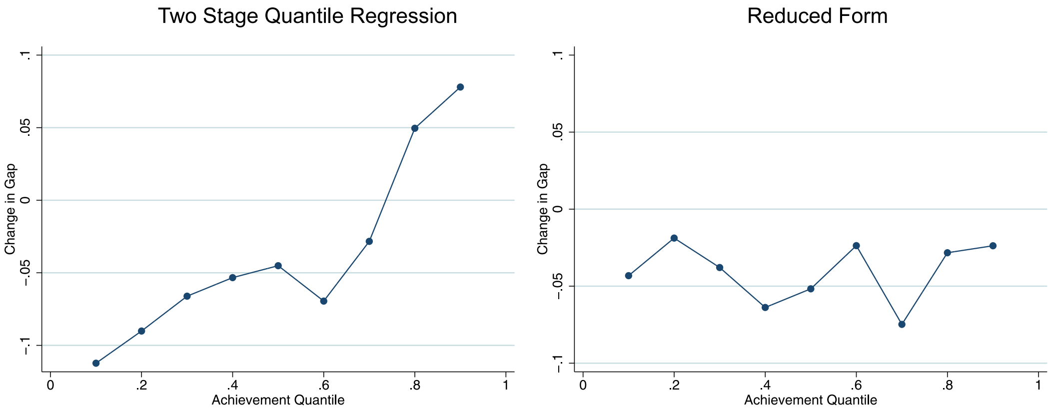


FIGURE 5. Changes in achievement gap from desegregating Durham. Predictions are based on assigning each student the average school by year fixed effect and teacher quality, and randomly assigning fifth graders in 2001–2002 to classrooms in Durham. The $x$-axis refers to the student's position in the unconditional achievement distribution under observed groupings (but with equal resources). The achievement gap at the $x$th percentile is calculated as the difference in the $x$th percentile of the white and nonwhite achievement distributions. The change in the gap is then taken from the difference in the gap at the $x$th percentile in the randomly assigned groupings relative to the observed groupings. The left-hand panel uses two-stage quantile regression parameter estimates, as described in equation (5.2). The right-hand panel uses reduced-form parameter estimates, as described in equation (7.1).

achievement of whites and nonwhites,

$$\begin{aligned} Y_{ict} = \beta_0 + \bar{Y}^{\mathrm{W}}_{-ict-1}\beta_1 + \bar{Y}^{\mathrm{NW}}_{-ict-1}\beta_2 + \mathbf{X}_{it}\boldsymbol{\beta}_3 + \bar{\mathbf{X}}_{-ict}\boldsymbol{\beta}_4 + \mathbf{P}_{it}\boldsymbol{\beta}_5 \\ + \mathbf{K}_{ct}\boldsymbol{\beta}_6 + \widehat{\mathrm{SchYr}}_{it}\beta_7 + u_{ict}, \end{aligned} \tag{7.1}$$

where peer characteristics include average peer parental education (high school or college plus), percentage nonwhite, and percentage male.[32] The school-by-year fixed effects are predicted from the mean version of this equation. I estimate the production function separately by race. Peer effect estimates are much smaller using this estimator, with the largest spillover from lagged peer achievement on the order of magnitude of 0.04.[33] The right-hand panel of Figure 4 shows that the reduced-form estimates predict very little change in achievement as a result of desegregation. On average, nonwhite students lose about −0.02 and white students gain about 0.02. Furthermore, the estimator predicts that the gap decreases at all percentiles of the achievement distribution, from as much as −0.02 to −0.07, with an average narrowing of −0.04. While the average happens not to be that different from that in the two-stage quantile regression, the reduced-form estimator fails to capture the distributional effects (the increases in the

[32]The equation does not control for percentage white–nonwhite achievement levels 1–2 and 3 (and interactions with grade 5 and post-2001) because of multicollinearity with lagged white–nonwhite peer achievement and because this specification is more closely connected to others in the literature.

[33]See Appendix Table A.3.

gap at the upper percentiles versus the decreases at the lower percentiles). Furthermore, the change in the gap in the two cases arises from quite different sources. For instance, whites experience losses at all percentiles under the reduced form, whereas the lower-achieving whites experience large gains under the two-stage quantile regression. The reduced form also predicts that the highest-achieving nonwhites experience similar gains as the lowest-achieving nonwhites, whereas the two-stage quantile regression predicts significant losses for the highest-achieving nonwhites and large gains for the lowest-achieving nonwhites. Thus, the reduced-form and two-stage quantile regression have quite disparate policy implications.

## 8. Conclusion

This paper uses an equilibrium model of student behavior to motivate a new approach to interpreting and identifying peer achievement spillovers. I find that the effect of average peer achievement is comparable in magnitude or larger than some of the more important determinants of student achievement found in the literature, such as teacher quality and class size (e.g., Rivkin, Hanushek, and Kain (2005)). This suggests that in ignoring behavioral spillovers deriving through contemporaneous achievement, studies that focus on prior peer achievement severely understate the effect of peers.

My identification and estimation strategy contributes new insight to the desegregation literature by allowing the effect of racial diversity to operate both directly through racial composition and through heterogeneity in responses to peer achievement by race and by percentiles of the achievement distribution. I find that white students appear to conform only to white peer achievement and nonwhites conform only to nonwhite peers. I also find that lower-achieving students respond relatively more to increases in average peer achievement (of the same race) than higher-achieving students. The direct effect of racial composition, which is the focus of prior research, is swamped in magnitude by the peer achievement spillovers.

Despite the large peer effects, I find that creating racially diverse peer groups would lead to only a small narrowing of the achievement gap, 0.04 of a standard deviation on average. In part, this is because of a lack of cross-racial spillovers. However, simply focusing on the mean effect of desegregation masks important distributional effects. For lower-achieving students, the gap narrows under desegregation by about 10%, and is driven by improvements to both nonwhite and white achievement. The gap increases at the upper percentiles by as much as 0.08 of a standard deviation, but this is driven by losses to both whites and nonwhites. I show that the predictions from reduced-form peer effect regressions both misstate the overall effect of desegregation on the achievement gap and fail to capture these important distributional effects.

This paper has focused on isolating one mechanism, namely peers, through which desegregation may help narrow the achievement gap. It is worth emphasizing that a full assessment of the effect of desegregation would need to take into account the general equilibrium effects of residential sorting and/or selection out of public schools in response to regrouping of students. Given the importance of peer achievement spillovers, future research on desegregation might fruitfully examine how to generate greater cross-racial spillovers and the potential for desegregation to change student interactions over time.

## References


Ammermueller, A. and J.-S. Pischke (2009), "Peer effects in European primary schools: Evidence from PIRLS." *Journal of Labor Economics*, 27, 315–348. [113]

Bishop, J. H., M. Bishop, L. Gelbwasser, S. Green, and A. Zuckerman (2003), "Nerds and freaks: A theory of student culture and norms." *Brookings Papers on Education Policy*, 6, 141–199. [89]

Bisin, A., A. Moro, and G. Topa (2011), "The empirical content of models with multiple equilibria in economies with social interactions." Working Paper 17196, National Bureau of Economic Research. [90]

Bramoulle, Y., H. Djebbari, and B. Fortin (2009), "Identification of peer effects through social networks." *Journal of Econometrics*, 150 (1), 41–55. [86, 88]

Brock, W. A. and S. N. Durlauf (2001a), "Discrete choice with social interactions." *Review of Economic Studies*, 68 (2), 235–260. [90, 93]

Brock, W. A. and S. N. Durlauf (2001b), "Interactions-based models." In *Handbook of Econometrics*, Vol. 5 (J. Heckman and E. Leamer, eds.), 3297–3380, Elsevier, Amsterdam. [89, 93]

Calvo-Armengol, A., E. Patacchini, and Y. Zenou (2009), "Peer effects and social networks in education." *Review of Economic Studies*, 76 (4), 1239–1267. [86, 88]

Chemerinsky, E. (2003), "The segregation and resegregation of American public education: The courts' role." *North Carolina Law Review*, 81 (4), 1597–1622. [86]

Chernozhukov, V. and C. Hansen (2005), "An IV model of quantile treatment effects." *Econometrica*, 73 (1), 245–261. [87, 101]

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2003), "Segregation and resegregation in North Carolina's public school classrooms." *North Carolina Law Review*, 81 (4), 1463–1512. [98]

Clotfelter, C. T., J. L. Vigdor, H. F. Ladd, and R. A. Diaz (2004), "Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers?" *Journal of Policy Analysis and Management*, 23 (2), 251–271. [109]

Epple, D. and R. Romano (2010), "Peer effects in education: A survey of the theory and evidence." In *Handbook of Social Economics*, Vol. 1B (J. Benhabib, A. Bisin, and M. O. Jackson, eds.), Chapter 20, 1053–1164, North-Holland, Amsterdam, The Netherlands. [98]

Figlio, D. N. (2007), "Boys named Sue: Disruptive children and their peers." *Education Finance and Policy*, 2 (4), 376–394. [89]

Fordham, S. and J. Ogbu (1986), "Black students' school success: Coping with the 'burden of acting white'." *The Urban Review*, 18, 176–206. [87]

Fruehwirth, J. C. (2012), "Can achievement peer effect estimates inform policy? A view from inside the black box." Working paper, University of Cambridge. [103, 117]

Fryer Jr., R. G. and P. Torelli (2010), "An empirical analysis of 'acting white'." *Journal of Public Economics*, 94 (5–6), 380–396. [87, 108]

Gibbons, S. and S. Telhaj (2008), "Peers and achievement in England's secondary schools." Discussion Paper 0001, Spatial Economics Research Centre, LSE. [85, 103]

Giorgi, G. D., M. Pellizzari, and S. Redaelli (2010), "Identification of social interactions through partially overlapping peer groups." *American Economic Journal: Applied Economics*, 2 (2), 241–275. [86]

Graham, B. S. (2008), "Identifying social interactions through conditional variance restrictions." *Econometrica*, 76 (3), 643–660. [105]

Graham, B. S., G. W. Imbens, and G. Ridder (2010), "Measuring the effects of segregation in the presence of social spillovers: A nonparametric approach." Working Paper 16499, National Bureau of Economic Research. [117]

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2003), "Does peer ability affect student achievement?" *Journal of Applied Econometrics*, 18 (5), 527–544. [87, 93, 98, 100, 103]

Hanushek, E. A., J. F. Kain, and S. G. Rivkin (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. [87, 103]

Hoxby, C. (2000), "Peer effects in the classroom: Learning from gender and race variation." Working Paper 7867, National Bureau of Economic Research. [87, 98]

Hoxby, C. M. and G. Weingarth (2005), "Taking race out of the equation: School reassignment and the structure of peer effects." Working Paper, Harvard University. [100]

Imbens, G. W. and W. K. Newey (2009), "Identification and estimation of triangular simultaneous equations models without additivity." *Econometrica*, 77 (5), 1481–1512. [87, 97, 98, 101]

Jacob, B. (2005), "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools." *Journal of Public Economics*, 89 (5–6), 761–796. [96, 109]

Lavy, V., M. D. Paserman, and A. Schlosser (2012), "Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom." *Economic Journal*, 122 (559), 208–237. [103]

Lavy, V. and A. Schlosser (2011), "Mechanisms and impacts of gender peer effects at school." *American Economic Journal: Applied Economics*, 3 (2), 1–33. [87, 98, 112]

Lazear, E. P. (2001), "Educational production." *Quarterly Journal of Economics*, 116 (3), 777–803. [89]

Manski, C. (1993), "Identification of endogenous social effects: The reflection problem." *Review of Economic Studies*, 60 (3), 531–542. [86, 93, 113]

Moffitt, R. A. (2001), "Policy interventions, low-level equilibria and social interactions." In *Social Dynamics* (S. N. Durlauf and H. P. Young, eds.), 45–82, Brookings Institution Press, Washington, DC. [93]

Neal, D. A. and D. W. Schanzenbach (2010), "Left behind by design: Proficiency counts and test-based accountability." *Review of Economics and Statistics*, 92 (2), 263–283. [96, 109]

Reback, R. (2008), "Teaching to the rating: School accountability and the distribution of student achievement." *Journal of Public Economics*, 92 (5–6), 1394–1415. [96, 109]

Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005), "Teachers, schools, and academic achievement." *Econometrica*, 73 (2), 417–458. [91, 106, 121]

Sacerdote, B. (2011), "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In *Handbook of the Economics of Education*, Vol. 3, Chapter 4, 249–277, Elsevier, Amsterdam. [85, 100]

Schaffer, M. E. (2005), "xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models." Statistical software components, Boston College Department of Economics. [101]

Sweeting, A. (2009), "The strategic timing of radio commercials: An empirical analysis using multiple equilibria." *RAND Journal of Economics*, 40 (4), 710–742. [90]

Vigdor, J. and T. Nechyba (2004), "Peer effects in elementary school: Learning from 'apparent' random assignment." Working paper, Duke University. [112]

Vigdor, J. and T. Nechyba (2007), "Peer effects in North Carolina Public Schools." In *Schools and the Equal Opportunity Problem* (P. Peterson and L. Woessmann, eds.), MIT Press, Cambridge. [87, 103]

---

Submitted July, 2012. Final version accepted September, 2012.