

NBER WORKING PAPER SERIES

PEER EFFECTS IN THE CLASSROOM:
LEARNING FROM GENDER AND RACE VARIATION

Caroline Hoxby

Working Paper 7867
<http://www.nber.org/papers/w7867>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2000

The author very gratefully acknowledges the help of John Kain, Daniel O'Brien and others at The Texas Schools Project, Cecil and Ida Green Center for the Study of Science and Society, University of Texas at Dallas. The data used are part of the Texas Schools Microdata Panel maintained by them. The author also gratefully acknowledges help from the Texas Education Agency and research assistance from Bryce Ward and Joshua Barro. The author is solely responsible for any errors. The views expressed herein are those of the author and not necessarily those of the National Bureau of Economic Research.

© 2000 by Caroline Hoxby. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Peer Effects in the Classroom: Learning From Gender and Race Variation

Caroline Hoxby

NBER Working Paper No. 7867

August 2000

ABSTRACT

Peer effects are potentially important for understanding the optimal organization of schools, jobs, and neighborhoods, but finding evidence is difficult because people are selected into peer groups based, in part, on their unobservable characteristics. I identify the effects of peers whom a child encounters in the classroom using sources of variation that are credibly idiosyncratic, such as changes in the gender and racial composition of a grade in a school in adjacent years. I use specification tests, including one based on randomizing the order of years, to confirm that the variation I use is not generated by time trends or other non-idiosyncratic forces. I find that students are affected by the achievement level of their peers: a credibly exogenous change of 1 point in peers' reading scores raises a student's own score between 0.15 and 0.4 points, depending on the specification. Although I find little evidence that peer effects are generally non-linear, I do find that peer effects are stronger *intra*-race and that some effects do not operate through peers' achievement. For instance, both males and females perform better in math in classrooms that are more female despite the fact that females' math performance is about the same as that of males.

Caroline Hoxby

Department of Economics

Harvard University

Cambridge, MA 02138

and NBER

choxby@harvard.edu

I. Introduction

Peers effects have long been of interest to social scientists because, if they exist, they potentially affect the optimal organization of schools, jobs, neighborhoods, and other forums in which people interact. Economists, in particular, are interested in peer effects because it is likely that at least some peer effects--which are, by definition, externalities--are not internalized. Thus, the existence of peer effects may create opportunities for social welfare-enhancing interventions, in form of prices that make people act as though they internalized the value of their own peer effects.

For example, the literature on school finance and control is currently absorbed by the question of whether students are affected by the achievement of their schoolmates.¹ If peer effects exist at school, then a school finance system that encourages an efficient distribution of peers will make human capital investments more efficient and will, thus, increase macroeconomic growth. Similar arguments are made regarding the organization of local government, which may encourage or discourage an efficient distribution of peer effects within neighborhoods. Indeed, a number of recent models of macroeconomic growth depend crucially on peer effects.² At a less high-flown level are questions like whether schools should eliminate tracking, under which students are exposed only to peers with similar achievement, and whether desegregation plans should assign students to schools outside their neighborhood or their district.³

There are two principal difficulties for theories that rely on peer effects. First, it is doubtful whether peer effects exist at all because there are formidable empirical obstacles to estimating them. Although some credible estimates of peer effects do exist, people often rely on evidence that is seriously

¹ See Nechyba [1996] and Epple and Romano [1998].

² Benabou [1996] and Kremer [1993] are examples.

³ A sampling of the peer effects literature might include, in addition to works mentioned elsewhere in this paper: Summers and Wolfe [1977], Banerjee and Besley [1990], Case and Katz [1991], Betts and Morell [1999], Zimmer and Toma [2000], and the chapters in Brooks-Gunn, Duncan, and Aber [1997].

biased by selection. For instance, if everyone in a group is high achieving, then many observers assume that achievement is an *effect* of belonging to the group instead of a *reason* for belonging to it. I return to this point below.

Second, the model of peer effects that is probably most popular in practice (the “baseline” model) is one in which peer effects have distributional consequences but no efficiency consequences. According to the baseline model, an individual’s outcome on a certain variable is affected linearly by the mean of his peers’ outcomes on that variable.⁴ For instance, under the baseline model, a student’s reading score would be affected linearly by the mean reading score of his classmates. Regardless of how one allocates peers, total societal achievement remains the same under the baseline model. In order to give one student a better peer, one must take that peer away from another student; the two effects exactly cancel. If one accepts the baseline model, then one is limited to peer effects questions that are distributional in nature, such as disparity in educational opportunities or income inequality.⁵ Many questions regarding peer effects, however, require a model that is either non-linear in peers’ mean achievement or in which other moments of the peer distribution matter. For instance, the argument for de-tracking is based on the idea that *both* less able and more able students benefit from being with one another in the classroom.⁶ Other models of learning impose the condition that more able individuals benefit *more* from a good peer. The pedagogical literature is inconsistent: both the “one bad apple” and the “one shining light” models are popular. Any theory in which economic growth depends on peer effects must employ a model other than the baseline model. Thus, although one might be tempted to dismiss

⁴ The baseline model is often expressed with an equation like the following:

$$y_{ij} = a + b\bar{y}_j^{-i} + X_{ij}c$$

where y_{ij} is some outcome for person i in group j , $\bar{y}_j^{-i}$ is the mean value of the outcome for all of the people in group j except for person i , and X_{ij} is a vector of other factors that affect person i ’s outcome.

⁵ See, for instance, Durlauf [1996].

⁶ See Argys, Rees, and Brewer [1996].

the baseline model as naive or restrictive, if one were to find empirically that the baseline model adequately described peer effects, some interesting theories would fall by the wayside.

The central problem with estimating peer effects in schools is that vast majority of cross-sectional variation in students' peers *is* generated by selection. Families self-select into schools based on their incomes, job locations, residential preferences, and educational preferences. A family may even self-select into a school based on the ability of an individual child. For instance, a family with a highly able child may choose to live near a school that has a program for gifted children. Moreover, families may influence the particular class to which their child is assigned within his school. If, for example, educationally savvy parents believe that a certain third grade teacher is best, they may get their children assigned to her class, creating a class in which parents care about education to an unusual degree. School staff can generate a great deal of additional selection. A school may assign children with similar achievement to the same classroom, in order to minimize teaching difficulty. Or, a school may place all of the "problem" students in a certain teacher's class because she is good at dealing with them. In short, one should assume that a child's being in a school *is* associated with unobserved variables that affect his achievement. One should also assume that there *are* unobserved variables associated with a child's being in a particular classroom, within his grade within his school.

In this paper, I take for granted that parents choose a school based on its population of peers and that parents and schools manipulate the assignment of students to classes within their grades. I introduce two empirical strategies that, even under these conditions, generate estimates of peer effects that are credibly free of selection bias. Both strategies depend on the idea that there is *some* variation in adjacent cohorts' peer composition *within a grade within a school* that is idiosyncratic and beyond the easy management of parents and schools.⁷ That is, even parents who make very active decisions about their

⁷ A student's "cohort" is determined by the year in which he reaches a given grade--for instance, students who enter kindergarten in fall 2000 are a "cohort."

child's schooling cannot perfectly predict how their child's actual cohort within a given public school will turn out. There are differences between adjacent cohorts that would be labeled "unexpected" even by econometricians who have far more information than parents have. Parents are unlikely to predict these "unexpected" differences perfectly. A parent may have a fairly accurate impression of the cohorts around his child's age and may pick a school on that basis, but it is expensive for a parent to react to a cohort composition "surprise" by changing schools. Moreover, so long as we focus on idiosyncratic variation in *cohort* composition, as opposed to classroom composition, we need not worry about schools and parents manipulating the assignment of students to classrooms. If a cohort is more female than the previous cohort, for instance, the school must allocate the "extra" females among its classrooms somehow. Inevitably, some students in the cohort will end up with a peer group that is more female than is typical.

In the first strategy, I attempt to identify idiosyncratic variation by comparing adjacent cohorts' gender and racial groups' shares. In the second strategy, I attempt to identify the idiosyncratic component of each group's achievement and determine whether the components are correlated. For both strategies, I am sensitive to the potential criticism that what appears to be idiosyncratic variation in groups' shares or achievement may actually be a time trend within a grade within a school. (This criticism does not affect estimates based on gender groups under strategy 1.) To address this criticism, I not only eliminate linear time trends: I also eliminate any school from the sample in which actual years explain more variation (in cohort composition or in achievement) than false, randomly assigned years.

I implement these empirical strategies using administrative data on third, fourth, fifth, and sixth graders in the state of Texas during the 1990s. The data cover the entire population of Texas students in public schools. Texas contains a very large number of elementary schools, which is fortunate because idiosyncratic variation in cohorts within a grade within a school is sufficiently uncommon that a large number of observations are needed to generate the needed number of "natural events."

The empirical strategies in this paper are, I would argue, an improvement on many previous methods of identifying peer effects in schools. Previous researchers have most often estimated models like the baseline model and used cross-sectional variation in schoolmates to identify effects. They have dealt with selection by controlling for observable variables, comparing siblings in families that move (so that the siblings experience different schools), examining children in magnet or desegregation programs, or estimating a selection model.⁸ In practice, these methods have generally proved unconvincing because there are unobservable variables that are correlated with peer selection, with moving, with participating in a magnet or other school program, or with the excluded variables that identify the selection model. Some of the most convincing estimates of peer effects come from policy or natural experiments at the college or neighborhood level. For instance, Zimmerman [1999] and Sacerdote [2000] estimate the effects of college roommates who are conditionally randomly assigned at Williams College and Dartmouth College, respectively. Rosenbaum [1995] and de Souza Briggs [1997] describe housing mobility programs, which are a promising source of information on neighborhood effects.⁹

Before proceeding to the empirical strategies, it is useful to clear about what peer effects include. Peer effects do include students teaching one another, but direct peer instruction is only the tip of the iceberg. A student's innate ability can affect his peers, not only through knowledge spillovers but through his influence on classroom standards. A student's environmentally determined behavior may affect his peers. For instance, a student who has not learned self-discipline at home may disrupt the

⁸ In particular, Boston's Metco program, in which inner-city minority children are sent to schools in the suburbs, has been much studied. The difficulty with estimates based on Metco is that children who enter the program (and do not attrit from it) are likely to have higher unobserved ability or motivation.

⁹ One must approach peer effects estimates from housing mobility program with some caution, however. Even in programs that randomize offers of housing mobility (such as the "Moving to Opportunity" program), families that apply may be unusually susceptible to peer effects, and families that attrit are less likely to have experienced good peer effects. In the Gautreaux program described by Rosenbaum and de Souza Briggs, being offered the change to move is not randomized among applicants, but there is some arbitrariness in the neighborhood to which the family moves. Selection bias is certainly reduced, relative to normal family moves observed in data like the Panel Survey of Income Dynamics or the National Longitudinal Survey of Youth, but size of the reduction is unclear.

classroom. Peer effects may follow lines like disability, race, gender, or family income: a learning disabled child may draw disproportionately on teacher time, racial or gender tension in the classroom may interfere with learning, richer parents may purchase learning resources that get spread over a classroom. Peer effects may even work through the way in which teachers or administrators react to students. For instance, if teachers react to black students by creating a classroom atmosphere in which students are expected to perform badly, then the effects of such systematic teacher behavior would be associated with black peers. In some cases, I am able to distinguish empirically among the various channels for peer effects. In general, however, the peer effects estimated in this paper (and in most research) embody multiple channels. When judging the magnitude of the results, it is important to keep the multiple channels in mind. Note that the baseline model does not assert that there is a single channel for peer effects: it asserts that mean peer achievement is a sufficient statistic for the multiple channels.

II. The Empirical Strategies

The essence of the two empirical strategies employed in this paper is simple. One needs a source of variation in the peers whom a student experiences that does not reflect self-selection or selection by other forces. Variation in peers between schools is suspect because families self-select into schools. The variation in peers between classrooms within a cohort within a school is suspect: students may be placed in classrooms based on schools' or parents' assessment of their abilities or of teachers' abilities. Variation within and between private schools is suspect because they have some control over admissions.

Fortunately, adjacent cohorts in a grade in a particular public school are a potential source of non-suspect variation. Even within a school that has an entirely stable population of families, biological variation in the genetic ability, timing, and gender of births would create idiosyncratic variation in the share of 6 year olds, say, who were female, white, innately able, and so on. It is this idiosyncratic variation that the empirical strategies in this paper attempt to exploit. The strategies use far more

information than parents have to identify variation between cohorts that is, I would argue, credibly idiosyncratic, unlikely to have been foreseen by parents, and unlikely to reflect unobserved neighborhood variables. Moreover, because the strategies exploit variation in *cohort* composition, as opposed to classroom composition, they are impervious to the effect of parents and schools selecting particular classrooms within a cohort within a grade within a school.

A. Empirical Strategy 1 - The Basics

There is little reason to suspect that variation between cohorts in gender composition, within a grade within a public school, is correlated with unobserved determinants of achievement. A school with entirely stable demographics has variation in cohorts' gender composition purely because of variation in the gender composition of births. The availability of single-sex private schools is one of the only forces that systematically affects the gender composition of public schools, but private schools tend to have effects that are grade-specific, not cohort-specific for a given grade in a given school. For instance, a single-sex private school may enroll children only through the fourth grade (which would probably cause a shift in gender composition between grades four and five in the local public school), but the private school is not likely have very different effects on adjacent cohorts within grade four within the local public school. Indeed, it is not merely plausible that variation in gender composition between cohorts within a grade within a school is essentially random, there is no public elementary school in the Texas data that shows evidence of a time pattern in gender composition.

Because cohort-to-cohort changes in the gender composition of a grade within a public school are, very plausibly, all due to random variation, empirical strategy 1 is most easily illustrated using gender composition. After presenting strategy 1 in its simplest form, I extend and modify it to cover between-cohort variation in racial composition within a grade within a school. Intuitively, in strategy 1, I see whether first differences in the achievement of adjacent cohorts within a grade within a school are systematically associated with first differences in the gender composition of those cohorts. If there are

no peer effects, the average achievement of male (or female) students should not be affected by the *share* of their peers who are female.

To formalize this intuition, consider the achievement of male students in grade g in school j in cohort c . Let the variable i index the group to which the students belong. In this case, $i \in \{male, female\}$. Let the variable A stand for achievement.

Define $\eta_{male,gj}$ to be the “true” mean achievement of males in grade g in school j in the absence of peer effects. Because each male student has some idiosyncratic component of achievement, any given cohort of males in grade g in school j may have average achievement that deviates from $\eta_{male,gj}$. Let $\epsilon_{male,gjc}$ represent this deviation. In other words, if there are no peer effects, then the average achievement of male students is, by definition:

$$(1a) \quad \bar{A}_{male,gjc} = \eta_{male,gj} + \epsilon_{male,gjc} .$$

By definition, $\epsilon_{male,gjc}$ is distributed with mean zero.¹⁰ Equation 1a assumes that true mean achievement is stable across cohorts; I relax this assumption below. Naturally, there is a parallel equation for females:

$$(1b) \quad \bar{A}_{female,gjc} = \eta_{female,gj} + \epsilon_{female,gjc} .$$

If there are peer effects, then equation 1a is insufficient because there are at least two ways in which the average achievement of males could be affected by the presence of female peers. First, to the extent that $\eta_{male,gj}$ is not equal to $\eta_{female,gj}$, peer achievement in a cohort varies systemically with the share of the cohort that is female. If students are influenced by their peers’ achievement, then the cohort’s gender composition would affect males’ achievement. Second, the prevalence of females could have some effect on achievement that does not operate through its effect on peer achievement. Females might, for instance, have a general effect on classroom culture. Equations that allow for peer effects (through peer achievement or other channels) are:

$$(2a) \quad A_{male,gjc} = \eta_{male,gj} + \beta p_{female,gjc} + \epsilon_{male,gjc} ,$$

¹⁰ It is also reasonable, under the circumstances, to assume that $\epsilon_{male,gjc}$ is normally distributed.

$$(2b) \quad A_{female,gjc} = \eta_{female,gj} + \gamma p_{female,gjc} + \epsilon_{female,gjc} ,$$

where $p_{female,gj}$ is the share of the cohort that is female. If there are no peer effects, then one should not be able to reject the null hypothesis that $\beta=0$ nor reject the null hypothesis that $\gamma=0$. That is, under the null of no peer effects, any given cohort of males may have average achievement that differs from that of other male cohorts in their grade in their school, but their achievement should not vary *systematically* with the share of students who are female.

When males and females are the groups, there is no definitive test for whether one group affects the other solely through its effect on peer achievement. Nevertheless, there are “plausibility” tests that happen to work well in practice. Moreover, there are definitive tests available when groups are defined along racial lines. See below for a discussion of this issue.

Naturally, one can write less restrictive versions of equations 2a and 2b that allow for nonlinear effects of $p_{female,gj}$. Nonlinear effects might occur if, say, it is not peers’ mean achievement that matters, but the achievement of the top quintile of peers. Alternatively, nonlinear effects might occur if females do not affect classroom culture until they are 60 percent, say, of a classroom. Below, I investigate nonlinearities but, for now, let us stick with linear equations, which are already general enough to subsume typical specifications of peer effects.

If one first differences equations 2a and 2b, one obtains the basic estimating equations for strategy 1:

$$(3a) \quad A_{male,gjc} - A_{male,gj,c-1} \equiv \Delta A_{male,gjc} = \beta_1 (p_{female,gjc} - p_{female,gj,c-1}) + (\epsilon_{male,gjc} - \epsilon_{male,gj,c-1}) \equiv \beta_1 \Delta p_{female,gjc} + \Delta \epsilon_{male,gjc} ,$$

$$(3b) \quad \Delta A_{female,gjc} = \gamma \Delta p_{female,gjc} + \Delta \epsilon_{female,gjc} .$$

The “true” basic achievement of males and females is assumed to be constant across adjacent cohorts in a grade in a school, so it drops out.

B. Extending Strategy 1 to Racial Groups

Schools classify students into five “racial” groups: Native American, Asian, black, Hispanic,

and white (“Anglo” in Texas). There are versions of equations 4a and 4b for racial groups, but, before writing them, consider a concern that arises when one extends strategy 1 to racial groups. A school might have a trend in the share of its students who are black, say. The trend might be associated with trends in other local variables that are unobserved and that affect achievement. Cohort-to-cohort changes in the share of students who are black will reflect the trend and will, moreover, be correlated with cohort-to-cohort changes in the unobserved variables. One might estimate an effect of cohort racial composition and naively interpret it as a peer effect when, in fact, it combines peer effects and the effects of the unobserved variables.

The data used in this paper have short panels (6 to 9 school years, depending on the grade). As shown below, the data evince changes in racial composition that are tiny compared to the apparently arbitrary cohort-to-cohort fluctuations in racial composition that are exploited by strategy 1. Nevertheless, I modify strategy 1 to address the problem of unobserved variables correlated with trends in racial composition.

First, I estimate linear trends for each racial group in each grade in each school. That is, a regression with a constant and a time variable is estimated for Asian students in grade 3 in school 1, another is estimated for black students in grade 3 in school 1, and so on for a total of about 48,000 regressions (about 3000 schools times 4 grades times 4 racial groups). I use the estimated residuals from these regressions as instruments for actual racial composition. Intuitively, I calculate each cohort’s “unexpected shock” in percent Asian, percent black, *et cetera*; and I use the cohort-to-cohort changes in the “shocks” as instruments for the actual cohort-to-cohort changes in racial composition. Formally, the counterparts of equations 2a and 3a are:

$$(4) \quad A_{Anglo,gjc} = \eta_{Anglo,gj} + \delta_1 p_{NativeAm,gjc} + \delta_2 p_{Asian,gjc} + \delta_3 p_{black,gjc} + \delta_4 p_{Hispanic,gjc} + \epsilon_{Anglo,gjc} ,$$

$$(5) \quad \Delta A_{Anglo,gjc} = \delta_1 \Delta p_{NativeAm,gjc} + \delta_2 \Delta p_{Asian,gjc} + \delta_3 \Delta p_{black,gjc} + \delta_4 \Delta p_{Hispanic,gjc} + \Delta \epsilon_{Anglo,gjc} .$$

Equations 4 and 5 show Anglo achievement as the dependent variable, but there are obviously parallel

equations with Native American, Asian, black, or Hispanic achievement as the dependent variable.

Equation 5 is estimated by instrumental variables where the instruments are:

$$(6) \quad \{\Delta\hat{\theta}_{NativeAm,gjc}, \Delta\hat{\theta}_{Asian,gjc}, \Delta\hat{\theta}_{black,gjc}, \Delta\hat{\theta}_{Hispanic,gjc}\},$$

which come from least squares estimation of the following equations:

$$(7a) \quad P_{NativeAm,gjc} = \alpha_{NativeAm,gj} + \phi_{NativeAm,gj} C + v_{NativeAm,gjc} \quad \forall g, j \text{ combinations}$$

$$(7b) \quad P_{Asian,gjc} = \alpha_{Asian,gj} + \phi_{Asian,gj} C + v_{Asian,gjc} \quad \forall g, j \text{ combinations}$$

$$(7c) \quad P_{black,gjc} = \alpha_{black,gj} + \phi_{black,gj} C + v_{black,gjc} \quad \forall g, j \text{ combinations}$$

$$(7d) \quad P_{Hispanic,gjc} = \alpha_{Hispanic,gj} + \phi_{Hispanic,gj} C + v_{Hispanic,gjc} \quad \forall g, j \text{ combinations}$$

The identifying assumption for this first modification to strategy 1 is that, for the short period in question, the time trends in racial composition can be adequately summarized by *linear* trends. For the vast majority of schools, this assumption appears to hold in practice. Nevertheless, one might argue that the modification does not far enough to eliminate potential omitted variables bias. Thus, I also use an alternative method that is almost certainly overkill.

The alternative method, which I call “drop if more than random,” works as follows. I flag a school as exhibiting a time trend in some racial group’s share if keeping the years in chronological order gives the school a more discernable time pattern than misassigning the years randomly. I drop all schools that--by this standard--exhibit a time trend in any racial group’s share, and I then use the reduced sample to estimate equation 5 (and the parallel equations for other races) by ordinary least squares.

More precisely, the “drop if more than random” procedure works as follows. I estimate, for each racial group in each grade in each school, a regression that has a constant and a quartic in the true year (cohort) of the data.¹¹ I then randomly reorder the cohorts for each regression five times, subject to the

¹¹ A quartic function in time is the highest power that is estimable for most of the grades in the sample.

constraint that the random reordering cannot equal the true order.¹² After each random reordering, I estimate, for each racial group in each grade in each school, a regression that has a constant and a quartic in the false order of the data. If the R-squared (share of variation explained) for the regression with *true* time is at least 1.05 times the smallest of the R-squared coefficients from the five regressions with false time, I flag the school as one with a time trend. The threshold is a stringent one, and--in general--this is a procedure that probably discards too many schools, especially since any racial group or grade can cause a school to be dropped.

The two methods just discussed for dealing with possible time trends can be applied to the gender group regressions just as easily as the racial group regressions. In practice, however, instrumental variables and “drop if more than random” results for gender groups are virtually identical to the results obtained from straightforward estimation of the first-difference equations. Evidently, schools do not have time trends in gender composition.

C. Do Gender and Racial Group Effects Work Solely through Peer Achievement?

Recall that the prevalence of a gender or racial group can have peer effects through at least two channels. First, to the extent that the groups have different values of η_{igj} , peer achievement in a cohort varies systemically with group shares in the cohort. Second, the prevalence of a group may have an effect on achievement that does not operate through its effect on peer achievement. We can test whether racial group effects work solely through peer achievement using the following method.

Obtain instrumental variables estimates or the “drop if more than random” estimates of equation 5; call these δ_1 , δ_2 , δ_3 , and δ_4 . Note that:

$$(8) \quad \Delta \bar{\eta}_{gjc} = (\eta_{NativeAm,gj} - \eta_{Anglo,gj}) \Delta p_{NativeAm,gjc} + (\eta_{Hispanic,gj} - \eta_{Anglo,gj}) \Delta p_{Asian,gjc} + (\eta_{black,gj} - \eta_{Anglo,gj}) \Delta p_{black,gjc} + (\eta_{Hispanic,gj} - \eta_{Anglo,gj}) \Delta p_{Hispanic,gjc} .$$

¹² Specifically, I assign a random number to each cohort and reorder the data according to the random number. If the random order happens to be the true order, I assign new random numbers to each cohort and reorder again. The process continues until data for each regression are in false, random order.

That is, a given increase in the share of a racial group increases peer achievement by a amount that varies with the difference between its η_{ig} and the η_{ig} of the base group, which is the Anglo group in this case.

One can estimate the difference between each group's η_{ig} and $\eta_{Anglo,g}$ of the base group by subtracting the implied estimate of $\eta_{Anglo,g}$

$$(9) \quad \hat{\eta}_{Anglo,g} = \text{weighted mean} (A_{Anglo,gjc} - \hat{\delta}_1 p_{NativeAm,gjc} - \hat{\delta}_2 p_{Asian,gjc} - \hat{\delta}_3 p_{black,gjc} - \hat{\delta}_4 p_{Hispanic,gjc}) ,$$

from the implied estimate of $\eta_{black,g}$, and so on.¹³ Each racial group's implied estimate of η_{ig} is computed using a equation like equation 9.

Translate the estimated coefficients on $p_{NativeAm,gj}$, $p_{Asian,gj}$, $p_{black,gj}$, and $p_{Hispanic,gj}$ into estimated coefficients on peer achievement by dividing each coefficient by the increase in peer achievement that a increase of 1.0 in a group's share would imply. For instance, suppose that Asians typically score 3 points higher in math than Anglos. Then, if the share of Asians rose by 10 percent and the share of Anglos dropped by 10 percent between two cohorts in a school, the underlying level of peer achievement (before peer effects) would rise by 0.10 times 3 points. Thus, if the coefficient on Asians' share were divided by 3, it would be the effect of raising peer achievement by 1 point.

Since the coefficient on each racial group's share can be translated into the common basis of peer achievement, one can test whether peer achievement is the sole channel for racial groups' peer effects by testing the hypothesis that the "translated" coefficients are equal (using an F-test). Put another way, if racial group composition has peer effects purely by changing peer achievement, then it should not matter whether peer achievement changes through a change in Asians' share, blacks' share or Hispanics' share--so long as the effect on $\bar{\eta}$ is the same. If one sees that a racial group has effects that are greatly in excess of what its plausible effects through peer achievement are, one should suspect that the group also has effects on peer achievement that operate through channels such as classroom culture, changes in teachers' behavior towards students, *et cetera*.

¹³ Equation 9 comes from applying the coefficient estimates from equation 5 to equation 4

When the groups are males and females, there is no neat test of whether a group's peer effects all operate through peer achievement. Nevertheless, one can still use "plausibility" tests based on common sense. For instance, an increase in the share of females that generates an 1 point increase in $\bar{\eta}$ might raise or lower the achievement of males by a fraction of a point or by a few points. If male achievement changes by *many* points, it is implausible that the entire effect of females as peers operates through peer achievement. Such "plausibility" tests happen to work well in practice.

D. Bells and Whistles for Strategy 1

There are a few minor empirical issues that deserve mention. First, the test itself and the testing arrangements vary slightly from year to year, so all of the estimating equations include year effects that are grade specific but common to all schools. If, for instance, the fourth grade test was unusually difficult in one year, then the difficulty would be common to the entire state and would be picked up by the year effect in the fourth grade equations. For visual simplicity, the year effects do not appear in the estimating equations written above, but in fact they are always included.

Second, the observations are group averages, and the groups vary in size. Larger groups' averages are likely to have smaller variance around the true mean. Weighted regression is the usual solution for this type of heteroskedasticity, and I employ weights throughout.

Third, although I have estimated versions of equation 5 in which the dependent variable is the achievement of Native American or Asian students, the number of students in these groups is so small that the resulting estimates are imprecise. Except when it is useful for clarity, I do not show estimates for Native American or Asian students' achievement.

Fourth, after examining the linear effects of group composition variables, I look for non-linear effects.

E. Empirical Strategy 2 - The Basics

The second empirical strategy also makes use of cohort-to-cohort differences in students, within grades, within schools; but it exploits information ignored in strategy 1. In strategy 2, I attempt to isolate the idiosyncratic component of each group's achievement (where a group is, as usual, a gender or racial group in a cohort in a grade in a school) and then test whether the idiosyncratic components of actual peers are correlated. For instance, if the females in the 1996-97 cohort of third graders in school 1 have unusually low achievement, does one find that the males in the 1996-97 cohort of third graders in school 1 have unusually low achievement too? If the Hispanic students in the 1994-95 cohort of fifth graders in school 100 have unusually high achievement, does one find that the Anglo, black, and Asian students in the 1994-95 cohort of fifth graders in school 100 have unusually high achievement too? For this strategy to make sense, one must obtain an estimate of the idiosyncratic component of each group's achievement that is *independent* of the estimates with which one plans to correlate it.

Formally, the procedure for strategy 2 works as follows. Obtain an estimate of each group's idiosyncratic achievement by estimating the regression:

$$(10) \quad A_{igc} = \lambda_{igj}^0 + \lambda_{igj}^1 c + \lambda_{igj}^2 P_{male,igc} + \lambda_{igj}^3 P_{black,igc} + \lambda_{igj}^4 P_{Hispanic,igc} + \xi_{igc} ,$$

for each group i in each grade g in each school j .¹⁴ For instance, one regression has, as its dependent variable, the reading scores of black third graders in school 1. An estimated residual from one of the above regressions is--literally--the portion of the achievement of cohort c in group i in grade g in school j that cannot be explained by a constant (specific to igj), a linear time trend (specific to igj), and the observed gender and racial composition of the cohort. Take the estimated residual to be an unbiased estimate of the idiosyncratic component of achievement of cohort c in group i in grade g in school j ; and note that the residual is estimated independently of the residuals for other groups in cohort c in grade g in school j . That is, the procedure does not, in any way, impose a correlation between residuals of different

¹⁴ This amounts to about 84,000 regressions for reading scores and the same number for math scores: about 3000 schools times 4 grades times 7 groups (2 gender groups and 5 racial groups).

groups who share the same classroom. The regression includes variables indicating the shares of the cohort that are female, black, and Hispanic because the results of strategy 1 suggest that these variables have systematic effects.

Rather than simply estimate pair-wise correlations among the residuals, it is best to estimate regressions that can take account of multiple “other” groups and state-wide year effects (because, as noted above, the test varies slightly from year to year). In addition, the regressions need to account for the fact that the idiosyncratic achievement of a group that forms a small share of a school’s students would not be expected to have the same peer effect as the idiosyncratic achievement of a group that forms a large share. If one multiplies each group’s idiosyncratic achievement by its median group share (that is, the median among the cohorts observed), however, one allows each student’s idiosyncratic achievement to have an equal effect. This is a reasonable basic specification and gives us regressions of the form:

$$(11) \quad \hat{\xi}_{black,gjc} = \theta_1 p_{NativeAm,gj}^{median} \hat{\xi}_{NativeAm,gjc} + \theta_2 p_{Asian,gj}^{median} \hat{\xi}_{Asian,gjc} + \theta_3 p_{Hispanic,gj}^{median} \hat{\xi}_{Hispanic,gjc} + \theta_4 p_{Anglo,gj}^{median} \hat{\xi}_{Anglo,gjc} + I^{cohort} \theta_5 + \omega_{black,gjc} ,$$

for examining correlations among racial groups and gives us regressions of the form:

$$(12) \quad \hat{\xi}_{male,gjc} = \theta_6 p_{female,gj}^{median} \hat{\xi}_{female,gjc} + I^{cohort} \theta_7 + \omega_{male,gjc} ,$$

for examining correlations among gender groups. I^{cohort} is the vector of indicator variables for cohorts that generates the state-wide year effects.

If there are no peer effects, one should not be able to reject the null hypothesis that $\theta_1=0$, $\theta_2=0$, $\theta_3=0$, $\theta_4=0$, and $\theta_6=0$. The interpretation of the coefficient θ_1 is, for instance, the effect on a black student’s achievement of having his Native American cohort-mates score one point higher on average (under the assumption that each student has an effect proportional to his share of the class). The interpretations of θ_2 , θ_3 , θ_4 , and θ_6 are similar. Moreover, if the idiosyncratic achievement of a student affects his peers in the same way regardless of his race or gender, then one should not be able to reject

the null hypothesis that $\theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_6$.

It is arbitrary that equation 11 is written with black students' idiosyncratic achievement as the dependent variable and that equation 12 is written with male students' idiosyncratic achievement as the dependent variable. Mainly for convenience, I show not only the results of equations 11 and 12, but also the results of parallel equations, with other racial groups' and females' idiosyncratic achievement as the dependent variables. Naturally, the results of the parallel equations do not contain much new information--they are mainly a way of rewriting the same information so that comparisons are easy.

F. Additional Notes on Strategy 2

There are two concerns about strategy 2. The first one is related to time trends. Equation 10, which is used to estimate idiosyncratic achievement, assumes that any time trend in each group's achievement can be captured by a linear term. One may be concerned, however, about time trends that are not captured by the linear term. Thus, after applying strategy 2 in its basic form, I use the "drop if more than random" method and apply strategy 2 on the reduced sample of schools that do not appear to have nonlinear time trends in achievement.

The second concern about strategy 2 is that estimated idiosyncratic achievement includes not only the effects of idiosyncratic student achievement (which one wants to exploit), but also the effects of common shocks that affect a particular cohort in a grade in a school. For instance, if a unusually good teacher is hired and teaches third grade for one year, her effect will be a common shock on the cohort of students who experience her teaching. Since all of the racial and gender groups in the cohort would presumably experience her teaching, it would appear that their idiosyncratic student achievement is correlated because of peer effects, when in fact they have simply experienced a common teaching shock. Note that an unusually good teacher who teaches third grade for the whole period would not cause such a problem: her effect would be absorbed in the fixed effect for third graders in the school. A third grade teacher who improved her teaching over the period would have her effect absorbed by the linear time

trend or would cause her school to be dropped under the “drop if more than random” method. Similarly, the substitution of a better for a worse third grade teacher part of the way through the period would almost certainly cause the school to be dropped under the “drop if more than random” method. Thus, one should be primarily concerned about teacher shocks of one or two years. One might also worry about transitory shocks like a building project that disrupts a classroom, unusual testing conditions like excessively hot weather, and so on.

There are two ways in which I test whether the peer effects apparently estimated in equations 11 and 12 are really the effects of common shocks. First, I attempt to determine the importance of peripatetic teachers by limiting the sample to schools with low teacher turnover over the period (fewer than 10 percent of the teacher slots in the school turn over in each six-year period). Second, I investigate whether the idiosyncratic *third grade* achievement of a group is correlated with the idiosyncratic *fifth grade* achievement of their peers. Such between-grade regressions are ideal for eliminating common shocks with transitory effects (such as test conditions), but not common shocks with lasting effects (such as a peripatetic teacher whose instruction has lasting effects). The standard for the between-grade test should be whether one can reject the null of no correlation, not whether the between-grade correlation is as strong as the same-grade correlation. After all, there are numerous reasons, apart from common shocks with a transitory effect, why between-grade correlation should be lower than same-grade correlation: the composition of a cohort changes as children migrate into and out of the school, a group that performs idiosyncratically well on third grade material need not perform equally well on fifth grade material, and so on.¹⁵

Furthermore, the variables for strategy 2 are estimated residuals, which are erroneous measures of true idiosyncratic achievement. The measurement error will generate attenuation bias, which will

¹⁵ One cannot use third grade to sixth grade comparisons because many students change schools between fifth and sixth grades, thereby disrupting cohort composition.

become particularly obvious in the between-grade regressions that eliminate common shocks with transitory effects. Put another way, the estimated residuals will contain classical measurement error and measurement errors that represent common shocks with transitory effects. The classical error will be uncorrelated across groups and will cause the estimates to be downward biased. The errors that represent common shocks will cause the estimates to be upward biased. The same-grade estimates may be either upward or downward biased because attenuation and common shocks work in opposite directions. The between-grade estimates will definitely be downward biased because they suffer only from attenuation bias. Measurement error will particularly affect the residuals estimated for Native Americans because so few students are in the group. One should not expect to learn much from the coefficients on the Native American residuals. The same problem affects the residuals estimated for Asians, to a lesser extent. Therefore, in interpreting the strategy 2 results, I focus on the idiosyncratic achievement of black, Hispanic, and Anglo students.

III. Data

The empirical strategies described require data on students' achievement on a standardized metric, by gender and racial group, in several adjacent cohorts. In addition, the empirical strategies call for cohorts that are relatively small (so that idiosyncratic variation in individual students' gender, race, and achievement does not get averaged out) and for many schools (since the share of observations with "natural events" is small). Cohorts also need to have integrity as peer groups. Cohorts have integrity in the elementary grades, but do not always have integrity in the secondary grades, where some classes are organized by material instead of by grade (for instance, Algebra II instead of grade 9 math).

The data requirements are fulfilled by a dataset drawn from the Texas Schools Microdata Sample, which is managed by the Texas Schools Project. The Microdata Sample uses administrative data on the population of students in Texas public schools, which are gathered by the Texas Education

Agency. Beginning with the 1990-91 school year, Texas began to administer a state-wide achievement test called the Texas Assessment of Academic Skills (TAAS) to elementary school students. TAAS is one of a generation of state-wide tests written by Harcourt-Brace Educational Measurement, the largest standardized test maker in the United States and the purveyor of such well-known tests as the Stanford 9 and Metropolitan Achievement Test. Although, like other state-wide tests, TAAS contains elements that are specific to the curriculum that Texas advocates, TAAS is a fairly typical standardized test with questions that are extremely similar (if not identical) to questions that Harcourt-Brace uses in other standardized tests.

In this paper, I use test data on grades three, four, five, and six. Grade three has been tested from 1990-91 to the present; grade four from 1992-93 to the present; and grades five and six from 1993-94 to the present. Table 1 display data on Texas schools and demographics for third graders, from 1990-91 to 1998-99. In a typical year during this period, there were about 3,300 schools in Texas that enrolled third graders and the size of the median cohort was about 80 students. Third graders were typically 48.7 percent female, 0.3 percent Native American, 2.3 percent Asian, 15.0 percent black, 33.1 percent Hispanic, and 49.3 percent Anglo. There were no apparent time trends in the shares of third graders who were female or Native American. There were slight upward trends in the shares of third graders who were Asian (2.2 to 2.5 percent over the period), black (14.8 to 15.7 percent over the period), and Hispanic (30.7 to 34.9 percent). There was a mild downward trend in the share of third graders who were Anglo (52.2 to 46.4 percent). Appendix Table 1 shows comparable statistics for grades four, five, and six, which are very similar (naturally, because most of the students are the same).

Table 2 shows statistics on the reading scores of third graders from 1990-91 to 1998-99. Over the period, the TAAS reading test had a mean of about 29.5 points and a standard deviation of about 2.3 points. The average female scored 1.1 points--or about half a standard deviation--higher than the average male. Compared to the average Anglo student, the average Native American student scored 1.5

points lower, the average Asian student scored 0.7 points higher, the average black student scored 3.6 points lower, and the average Hispanic student scored 2.9 points lower. Note that the black-Anglo and Hispanic-Anglo score gaps are substantial: 1.6 and 1.3 standard deviations, respectively. There is an upward trend in the scores of all groups over the period: the average score rose from 28.5 to 31.3 points. Some score improvement typically occurs over the first few years of test administration, simply owing to comfort with the test. The improvement in Texas scores accelerated over time, however, and the last few years' improvement are most likely to due to true learning of the material tested by the examinations--particularly as Texas distributed its curriculum (towards which the tests are oriented) only in the last few years.

Table 3 contains similar information for the TAAS math tests. The math test had a mean of 35.6 and a standard deviation of 2.9 over the period. There was a slight upward trend in scores: an average gain of 0.1 points per year. The average female scored 0.1 points higher than the average male--a difference of only 0.03 standard deviations. Compared to the average Anglo student, the average Native American student scored 1.9 points lower, the average Asian student scored 1.3 points higher, the average black student scored 4.7 points lower, and the average Hispanic student scored 3.2 points lower. The black-Anglo and Hispanic-Anglo score gaps are substantial: 1.6 and 1.1 standard deviations, respectively.

Appendix Tables 2 and 3 display reading and math test statistics for fourth, fifth, and sixth graders. The results are very similar to those for the third grade tests, except that the fourth, fifth, and sixth grade tests have slightly larger standard deviations. The standard deviations are 3.4 for reading and 4.2 for math in the fourth grade; 2.7 for reading and 3.8 for math in the fifth grade, and 3.1 for reading and 4.6 for math in the sixth grade.

Finally, Appendix Table 4 shows Asian-Anglo, black-Anglo, and Hispanic-Anglo score gaps for schools with different basic racial composition. For instance, the table displays the Hispanic-Anglo score

gap for schools that less than 10 percent, 10 to 25 percent, 25 to 60 percent, and more than 60 percent Hispanic. Interestingly enough, the score gaps tend to be similar across schools with different racial composition. This fact is convenient to know later, when we consider non-linear peer effects.

IV. Results of Strategy 1

Table 4 shows an example of the variation used by strategy 1. It displays statistics on the first differences in gender and racial shares for the 1994-95 school year versus the 1993-94 school year. Third grade cohorts are used. The racial shares are *detrended* (with a linear time trend) before the first differences are calculated. Thus, the table shows the instruments for equation 5.

Consider the first differences in percent female, for instance. A standard deviation in the variable is 11 percentage points. At the 1st percentile are cohorts with percent female that is 30 percentage points lower than the previous cohorts; at the 99th percentile are cohorts with percent female that is 28 percentage points higher. Clearly, the distribution of the first-differences is symmetric (as it should be). Since gender composition is highly centered around 49 percent female, we can see that most of the variation in gender composition that is exploited by strategy 1 is in cohorts that range from 20 to 80 percent female. There are a few all male and a few all female cohorts in the data, but such occurrences are naturally very rare.¹⁶

The first-differences in percent black, Hispanic, and Anglo have standard deviations of 6, 8, and 9 percentage points, respectively. At the 1st percentile are cohorts with black, Hispanic, and Anglo shares that are--respectively--17, 23, and 25 percentage points lower than the previous cohorts'. Since the distributions of the first differences are highly symmetric (as they should be if the detrending is working as intended), the 99th percentile is almost a mirror image of the 1st percentile.

Overall, Table 4 shows a large amount of cohort-to-cohort variation, within grade, within school.

¹⁶ All of these occurrences take place in schools with normal gender composition overall.

The cohort-to-cohort variation dwarfs the time trends shown in Table 1, and it is the foundation of strategy 1.

A. The Effect of Having A More Female Peer Group

Table 5 displays the effect of having a peer group that is more female (less male). The results are based on weighted least squares estimates of equations 3a and 3b. The structure of the table is similar to that of the tables that follow, so it is useful to describe it here. Each cell shows the estimated coefficient on the change in the share of the cohort that is female; and, thus, each cell represents a separate regression. The share of the cohort that is male is the “omitted share.” Neither Table 5 nor any of the tables that follow show the estimated year effects. The year effects are significant but simply pick up the year-to-year differences in the test across the state, as displayed in Tables 2 and 3. Each cell in Table 5 shows the coefficient first, with one asterisk if it is statistically significant at the 0.05 level and two asterisks if it is statistically significant at the 0.01 level. The standard error on the coefficient is in parentheses. In the square brackets is a translation of the coefficient into the effect of a change in peers’ mean test scores, where the change in the mean is due solely to the change in the share of the cohort that is female. To make this translation, one uses the estimated difference between the genders’ true underlying test scores (that is, test scores before peer effects). The translation is useful for testing the hypothesis that peer effects operate purely through peers’ achievement.

Table 5 shows that both females and males tend to perform better in reading when they are in more female classes. For instance, the coefficient on the change in the female share is 0.374 for female third graders’ reading scores, implying that females’ scores rise by 0.0374 points for every 10 percentage point change in the share of their class that is female. Males’ scores rise by 0.0471 points for 10 percentage change in the share of their class that is female. To put this in perspective, an all-female class would score about one-fifth of a standard deviation higher in reading, all else equal. The effects for fourth, fifth, and sixth grade reading scores are similar. The translation of the results into effects of mean

peer achievement provide a different perspective: being surrounded by peers who--for exogenous reasons--score 1 point higher on average raises a student's own score by 0.3 to 0.5 points, depending on the grade. The translation suggests that peer effects are substantial.

Table 5 also shows that both female and male students perform better in math when they are in more female classes. Female third graders' scores rise by 0.0381 points for every 10 percentage point change in the share of their class that is female. The effect is larger for higher grades: female sixth graders' scores rise by 0.0640 points for every 10 percentage point change in the share of their class that is female. A parallel effect exists for males' scores. Male third graders score 0.0396 points higher and male sixth graders score 0.0808 points higher for every 10 percentage point change in the share of their class that is female. Because the average female scores only a little higher than the average male, however, translating the scores into the effect of peers' mean achievement generates implausibly large effects. If one were to take the translated effects in square brackets literally, one would conclude that being surrounded by peers whose math scores were exogenously 1 point higher on average would raise a student's own score by 1.7 to 6.8 points, depending on the grade. These effects are so large that they suggest that peer effects do not operate purely through peers' mean achievement in math.

There are a few alternative channels that might explain the effect of females on math scores. First, since learning math requires reading and reading scores are higher in more female classes, females may affect subjects like math *through* their (quite plausible) peer effect on reading. Second, more female classes may simply have fewer disruptive students or a more learning-oriented culture. Third, classroom observers argue that pressure to be feminine makes girls unenthusiastic about math. Perhaps in female-dominated classrooms, females do not experience much pressure and therefore remain enthusiastic about math--allowing the teacher to teach it better to all students. In any case, it is clear that the baseline model of peer effects is inadequate: peer effects do not operate solely through peers' mean achievement in the same subject.

I investigate possible non-linearities in the effect of having a more female peer group in Table 6. The table displays estimates from a simple variant of equations 3a and 3b: the change in the female share is interacted with an indicator for whether the initial cohort was 0 to 33 percent female, 33 to 66 percent female, or 66 to 100 percent female. One can just discern a pattern in the point estimates. The effect of a change in the female share tends to be largest in classes that are initially at least 66 percent female.¹⁷ This suggests either that (1) the effects of peers' mean achievement is rising in their level of achievement or (2) the "atmospheric" effects of females in the classroom are especially strong in classrooms where females are in a supermajority.

B. The Effect of Having a Peer Group with Different Racial Composition

For interpreting the next set of results, it is worthwhile to remember that the effects under discussion are not the effects *inherently* associated with a racial group, but include the effects of variables that are statistically associated with a racial group in Texas, such as family income, parents' education, and home language. In particular, the effects should not be interpreted as effects of a group's innate ability.

Table 7a shows the effect of having a peer group with various racial compositions. The table displays weighted, instrumental variables estimates of equation 5 (and its parallel for other races). Each *column* represents a separate regression and shows the coefficients on the changes in the Native American, Asian, black, and Hispanic shares. The Anglo share is the omitted share. Table 7a shows results for third graders; Appendix Tables 5a, 6a, and 7a show parallel results for fourth, fifth, and sixth graders.

The broad result to draw from Table 7a is that black, Hispanic, and Anglo third graders all tend to perform worse in reading and math when they are in classes that have a larger share of black

¹⁷ Females' scores in math are probably an exception to this statement.

students.¹⁸ For instance, for every 10 percentage point change in the share of their class that is black, black students' reading scores fall by 0.2501 points, Hispanic students' reading scores fall by 0.0983 points, and Anglo students' reading scores fall by 0.0620 points. For the same 10 percentage point change in the share of their class that is black, black students' math scores fall by 0.1863 points, Hispanic students' reading scores fall by 0.0861 points, and Anglo students' reading scores fall by 0.0427 points. It is interesting that the effects of black peers appear to have the greatest effect on other black students; this difference in the size of the effect is largely confirmed by the results for grades four, five, and six. Recalling that black students have the lowest scores on both the reading and math tests, one can see that these results can be interpreted as effects of peer achievement. If one translates the results, one finds that being surrounded by peers who exogenously score 1 point lower on average has the following effects: it lowers a black student's own score by 0.676 points in reading and 0.402 points in math; it lowers an Hispanic student's own score by 0.266 points in reading and 0.185 points in math; and it lowers an Anglo students' own score by 0.168 points in reading and 0.092 points in math. The translation suggests that the effect of mean peer achievement varies from small (0.092) to substantial (0.676), and that the most substantial effects of mean peer achievement are *intra*-racial group.

There are other noteworthy effects in Table 7a and its parallel tables for fourth, fifth, and sixth grades (Appendix Tables 5a, 6a, and 7a). In the fourth, fifth, and sixth grades, Hispanic students perform worse in reading and math and Anglo students perform worse in math when they are in classes that have a larger share of Hispanic students. For instance, for every 10 percentage point change in the share of their class that is Hispanic, Hispanic fifth graders' reading scores fall by 0.1420 points and their math scores fall by 0.2047 points. For the same change in the Hispanic share, Anglo fifth graders' math scores fall by 0.0612 points. If one translates the results, one finds that being surrounded by peers who

¹⁸ As mentioned above, I do not show results in which Native American or Asian students' achievement is the dependent variable. These groups form such small shares of the student population that such results must be based on relatively few observations. The results are, nevertheless, available from the author.

exogenously score 1 point lower on average has the following effects: it lowers an Hispanic student's own score by 0.439 points in reading and 0.587 points in math; it lowers an Anglo student's own score by 0.176 points in math. Again, the results suggest that the effect of mean peer achievement varies, and are greatest for peers within the racial group generating the change in achievement.

There are a few coefficients on the change in the share of students who are Native American that are statistically significantly different from zero. Each of these significant coefficients is negative, a finding that is in keeping with the mean peer achievement interpretation of the coefficient.¹⁹ In addition, there are a few coefficients on the change in the share of students who are Asian that are statistically significantly different from zero. Each of these significant coefficients is positive and in a *math* regression. For instance, for every 10 percentage point change in the share of their class that is Asian, Anglo fifth graders' math scores rise by 0.0718 points and Anglo sixth graders' math scores rise by 0.2022 points. The effects of the Asian share are in keeping with mean peer achievement interpretations because the Asian-Anglo score gap is positive and relatively large in math (0.62 of a standard deviation in the fourth, fifth, and sixth grades).

The last lines of Table 7a and Appendix Tables 5a, 6a, and 7a show the p-value for the F-test of the hypothesis that changes in mean peer achievement have an equal effect regardless of which race generated them. In other words, having translated each coefficient into an effect of peers' mean achievement, one can test whether it is only peers' mean achievement that matters or also the composition of the peer group. The p-values indicate that the null hypothesis of equal effect tends to be rejected when black students' achievement is the dependent variable. The rejection is mainly caused by black students' achievement being disproportionately affected by the share of their cohort that is black. When Anglo students' achievement is the dependent variable, the null hypothesis tends not to be

¹⁹ Even when the coefficient on the change in the share of students who are Native American is statistically significant, it has a large standard error. It is not useful to interpret the point estimate of such coefficients, particularly in light of the small number of Native American students who generate the results.

rejected, suggesting that changes in mean peer achievement tend to affect Anglo students in the same way regardless of which racial minority group's share is responsible for the change. When Hispanic students' achievement is the dependent variable, the test results vary by grade and test. The null hypothesis is likely not to be rejected for math, but it is rejected about half the time for reading.

Table 7b shows alternative estimates of the effect of having a peer group with various racial compositions. The table displays least squares estimates of equation 5 (and its parallel for other races) that are computed using the reduced sample generated by the “drop if more than random” method. Almost two-thirds of the observations are dropped in this very stringent test for time trends. Despite the reduction in the sample, the results of Table 7b are generally similar to those of Table 7a, which assume that the time trends can be captured by linear terms.²⁰ In addition, Appendix Tables 5b, 6b, and 7b--which contain “drop if more than random” results for fourth, fifth, and sixth graders--display estimates that are similar to the parallel estimates that assume that the time trends can be captured by linear terms. Broadly, Table 7b and Appendix Tables 5b, 6b, and 7b suggest that black, Hispanic, and Anglo students perform worse in both reading and math when they are in a cohort that has a larger share of black students. The negative effect is stronger for black and Hispanic students than for Anglo students. There is also some evidence in the tables that Hispanic and Anglo students have lower scores (especially in math) when they are in a cohort that is more Hispanic. The negative effect of the Hispanic share is greatest for Hispanic students. A few coefficients suggest that the Asian share has a positive effect on Anglo students' achievement in math. The p-values at the bottom of each table have a pattern that is similar to the pattern described above for Table 7a and Appendix Tables 5a, 6a, and 7a.

The fact that intra-race peer effects appear to be stronger than between-race peer effects suggests one inadequacy of the baseline model of mean peer achievement, but what about general non-linearities? In Table 8, I investigate non-linearities in the effect of racial composition. The table displays estimates

²⁰ The standard errors are, however, uniformly larger in Table 7b than in Table 7a.

from a variant of equation 5 in which the change in the black share is interacted with an indicator for whether the initial cohort is 0 to 33 percent black, 33 to 66 percent black, or 66 to 100 percent black.

Also, the change in the Hispanic share is interacted with an indicator for whether the initial cohort is 0 to 33 percent Hispanic, 33 to 66 percent Hispanic, or 66 to 100 percent Hispanic.

Although the standard errors on some coefficients are large, there are three discernable patterns in the point estimates. The negative effect of the black share on black students is strongest in cohorts that between 33 and 66 percent black. The negative effect of the black share on Anglo students is largest in cohorts that are at least 33 percent black (it is unclear whether the effect is greater in the 33 to 66 percent or the 66 to 100 percent range). The negative effect of the Hispanic share on Hispanic students only appears in cohorts that are 0 to 33 percent Hispanic. In fact, the Hispanic share has a statistically significant, *positive* effect on the achievement of Hispanic students in cohorts that are 66 to 100 percent Hispanic. There are few possible interpretations of this sign reversal. First, greater availability of Hispanic peers may be helpful in cohorts that are already mainly Hispanic because each student who has difficulty speaking English is more likely to find a bilingual student to translate for him, help him learn English, and so on. Second, a more Hispanic cohort may be helpful for Hispanic students because it makes teachers sensitive to providing instruction that can be absorbed by language-minority students or because it forces a school to provide language services (such as English as a Second Language). Third, some schools, when faced with an unusually Hispanic cohort, may segregate their Spanish speaking students in a particular class because there are enough such students to fill a class. It is possible that such segregation generates higher achievement among Hispanic students (even if it is undesirable for other reasons).

V. Results of Strategy 2

Recall that the variables used in strategy 2 are groups' idiosyncratic achievement, where the

idiosyncratic component of achievement is, in practice, the residual from a school-grade-gender specific regression of test scores on a time trend, cohort gender composition, and cohort racial composition. The coefficients are effects of peers' test scores, so "translations" in square brackets are not needed. Also, because the variables are the product of the residuals themselves and the relevant group's median share, each coefficient can be interpreted as the effect of being surrounded by peers who score 1 point higher. Finally, recall that the variables for strategy 2 are estimates that contain measurement error, especially for Native American and Asian students. It is unclear whether measurement error causes the same-grade estimates to be biased (because attenuation bias and common shocks with transitory effects are offsetting), but the between-grade estimates are definitely downward biased.

Strategy 2 is concerned with the *correlation* among groups' residuals, so it is arbitrary which group's residuals are assigned to be the dependent variable in the regressions. Regressions are used for convenience since year effects must be estimated, but they are not meant to imply that females' residuals, say, *cause* males' residuals, anymore than males' residuals cause females' residuals. Partly to keep this point clear and partly for convenience of comparison, the tables "cycle" the dependent variable among the groups.

Table 9 exemplifies the structure of the tables that contain the results of strategy 2. In Table 9, each *cell* represents a different regression, and the regression is described by the two left-hand columns and the two right-hand column headings. In each regression, year effects were also estimated, but they are not shown.

A. The Effect of Peer Achievement, Take 1: Groups are Defined by Gender

Table 9 shows the effect of peer achievement, using residuals estimated for male and female groups. Clearly, these groups are mutually exclusive, so the residuals on the left- and right-hand side of each regression were estimated independently. In the top panel of Table 9, males' residuals are regressed on the residuals of females who were actually their peers. In the bottom panel of Table 9, females'

residuals are regressed on the residuals of the males who were actually their peers. For all of the same-grade regressions, one gender's idiosyncratic achievement has a positive, highly statistically significant effect on the idiosyncratic achievement of their peers from the other gender group. The point estimates are in a rather narrow range, especially for reading. In grades three through six, being surrounded by peers who score one point higher in reading raises a student's own score by 0.3 to 0.4 points. Put another way, the two gender groups' idiosyncratic achievements are correlated with a correlation coefficient of approximately 0.3 to 0.4, excluding the correlation generated by year-specific factors like the test itself. In math, being surrounded by peers who score one point higher raises a third grader's own score by about 0.6 points, raises a fourth grader's own score by about 0.5 points, and raises a fifth or sixth grader's own score by about 0.4 points.

To test whether the residuals are correlated due to common shocks, such as unusual test conditions, I regress fifth graders' residuals on the third grade residuals of their peers. These estimates are displayed in the bottom row of each panel of Table 9. The third grade residuals do have a statistically significant effect on the fifth grade residuals, which suggests that peer effects compose at least part of the same-grade correlation. The point estimates in the between-grade regressions are in the range of 0.06 to 0.08, but they are almost certainly underestimates because of attenuation bias and because migration of students limits between-grade correlation.

Table 10 contains two specification tests. The top panel tests whether the correlation between residuals is generated by teachers who teach only one or two years. (Recall that teachers who teach for longer periods will show up as fixed effects or time trends of some sort.) The sample used in the top panel includes only schools that have low teacher turnover (fewer than 10 percent of the slots turn over in each six-year period). The coefficients in the top panel of Table 10 are quite similar to those in Table 9, which suggests that teacher shocks do not account for much of the correlation. In fact, the correlations in the top panel of Table 10 are slightly higher than those in Table 9. It may be that the schools in the low

turnover sample are generally more stable so that the residuals are more precisely estimated and the coefficients suffer less from attenuation bias.

The bottom panel of Table 10 attempts to test whether the apparent peer effects in Table 9 are caused by insufficient controls for time trends. In particular, one might worry that the time trends for achievement are non-linear for some groups. The estimates in the bottom panel are computed using the reduced sample generated by the “drop if more than random” method.

The coefficients in the bottom panel of Table 10 are quite similar to those in Table 9, which suggests that non-linear time trends do not account for much of the correlation. In fact, the correlations in the bottom panel of Table 10 are slightly higher than those in Table 9, suggesting that schools with no apparent time trend may be more stable generally so that coefficients suffer less from attenuation bias.

B. The Effect of Peer Achievement, Take 2: Groups are Defined by Race

Table 11 shows the effect of peer achievement in reading, using residuals estimated for the five racial groups. Because the groups are mutually exclusive, the residuals for groups who are actually peers were estimated independently. Each *row* is a regression, and the table cycles the dependent variable through the races.²¹

For all of the same-grade regressions shown in Table 11, the idiosyncratic reading achievement of black, Hispanic, and Anglo students is positively, statistically significantly correlated. The pattern of coefficients also suggests that the idiosyncratic reading achievement of Asian students is positively, statistically significantly correlated with the reading achievement of black, Hispanic, and Anglo students, but that measurement error in the Asian residuals causes their coefficients to vary widely.²²

²¹ The table does not use Native American or Asian residuals as the dependent variable because the sample would be so small. The sample varies with the choice of the dependent variable because some schools do not contain any black students, other schools do not contain any Hispanic students, and so on.

²² The coefficients on Asian residuals vary widely, which suggests that measurement error (both classical and due to common shocks) generates a large share of the total variation.

Nevertheless, the idiosyncratic achievement of Asian students is positively, statistically significantly

The estimated effect of peers' reading achievement varies somewhat from regression to regression, but the precisely estimated coefficients suggest that being surrounded by peers who score 1 point higher in reading raises a student's own reading score by 0.3 to 0.8 points. For most of the same-grade regressions, the p-value in the right-hand column shows that one cannot reject the hypothesis that idiosyncratic achievement of peers has the same effect, regardless of the racial group from which the peers come. These tests suggest that the racial origin of peer achievement is not important, except perhaps within racial groups (strategy 2 cannot be used to analyze this issue). The tests also suggest that the effects of peer achievement are not highly non-linear. Black students, for instance, typically have low scores in their classes, so if variation in the low range mattered more than variation in the middle range, the coefficient on black students' residuals would be greater than the coefficient on Hispanic students' residuals.

The coefficients from the between-grade regressions are displayed in the bottom row of each panel of Table 11. The coefficients are statistically significant for the racial groups with residuals that are reasonably well estimated: black, Hispanic, and Anglo students. These statistically significant coefficients are in the range of 0.06 to 0.09, and one must keep in mind that they are almost certainly underestimates. They suggest, however, that peer effects compose at least part of the same-grade correlation. It is unlikely that common shocks account for the entire correlation.

Table 12 uses math achievement, but otherwise repeats the exercise shown in Table 11. In general, the math results are similar to the reading results. The same-grade correlations are slightly

correlated with the idiosyncratic achievement of Anglo students in all grades. In the third and fourth grades (but not in the fifth or sixth grades), the idiosyncratic achievement of Asian students is positively, statistically significantly correlated with the achievement of black and Hispanic students.

The third and fourth grades have longer panels and, thus, more precisely estimated residuals. More precise residuals probably account for the statistical significance of Asian residuals in the third and fourth, but not the fifth and sixth, grades.

The fact that Asian residuals are correlated with Anglo residuals even in the fifth and sixth grades, where the panels are short, suggests that the Asian residuals are more precisely estimated in schools that contain Anglos, but few black or Hispanic students. These are precisely the schools that get included in the regression when Anglo residuals are the dependent variables but get excluded when black or Hispanic residuals are the dependent variable.

larger for math than for reading, but the fifth-grade-to-third-grade correlations are similar in math and reading. In all but one regression, one cannot reject the hypothesis that idiosyncratic achievement of peers has the same effect, regardless of the racial group from which the peers come. Not only does this suggest that the racial origin of peer achievement is not important (except perhaps within racial groups), it also suggests that the effects of peer achievement are not highly non-linear.

Table 13 contains the specification tests based on schools with low teacher turnover and schools with no apparent time trends. The results in Table 13 are for math, so the results should be compared to those in Table 12. The top panel of Table 13 employs the low turnover sample to test whether the correlation between residuals is generated by teachers who teach only one or two years. The coefficients in the panel are very similar to those in Table 12, which suggests that teacher shocks do not account for much of the correlation.

The bottom panel of Table 13 uses the reduced sample generated by the “drop if more than random” method to test whether insufficient controls for time trends generate the apparent peer effects. The estimates in the panel are similar to those in Table 12, which suggests that non-linear time trends do not account for much of the correlation that has been attributed to peer effects.

Finally, Table 14 tests for non-linear effects of other groups’ achievement using a variant of equation 12 in which there is a quadratic in females’ residual achievement. The coefficients on the linear term are nearly identical to those in Table 9 (which restricted the effect to be linear) and the coefficients on the quadratic terms are all small (in the range of 0.001 to 0.008) and statistically insignificantly different from zero. These results do not provide any evidence of non-linearities; nor did results for racial groups or cubic specifications.

Let us assess the results of strategy 2 overall. The estimated peer effects based on gender groups are between 0.3 and 0.4, but only some of the statistically significant estimates based on racial groups are in the same range--about two-thirds are higher. The higher estimates may be overestimates caused by

common shocks with transitory effects. The between-grade estimates in which such common shocks are eliminated range between 0.6 and 0.9, but they are almost certainly underestimates of true peer effects, not only because of attenuation bias but also because the migration of a few low-achieving or high-achieving students can change a peer group's idiosyncratic component of achievement. In short, strategy 2 generates unambiguous evidence about the existence of peer effects, but the range of estimates is somewhat wide: 0.10 to 0.55 is a plausible summary of the range, given the various results and known biases.

VI. Conclusions

In this paper, I empirically investigate whether there are peer effects in the classroom. Schools are only one possible location for peer influence to occur, but they are possibly an important location. I attempt to identify the effects of peers as they work through all channels. Although one channel for peer effects is students instructing one another, peer effects may also work through classroom disruption, changes in classroom atmosphere, or resources that some students bring with them from home. Peer effects may even work through channels like the way in which teachers react to some students. In the paper, I make some effort to distinguish among the channels by which peer effects operate, but my primary purpose is to establish the existence and direction of peer effects. In particular, I attempt to judge the adequacy of the baseline model of peer effects, which states that a student's own achievement is affected linearly by the mean achievement of his peers.

The primary contribution of the paper is two empirical strategies that, I would argue, generate estimates of peer effects that are credibly free of selection bias. Selection has traditionally plagued estimates of peer effects, with parents' behavior and schools' behavior being potent sources of selection bias in classroom-based estimates of peer effects. Both empirical strategies depend on the idea that, although parents may choose a school based on its population of peers and schools may assign a child to

a classrooms based on his achievement, there is *some* variation between cohorts' peer composition *within a grade within a school* that is idiosyncratic and beyond the easy management of parents and schools. In the first strategy, I attempt to identify idiosyncratic variation by comparing adjacent cohorts' gender and racial groups' shares. In the second strategy, I attempt to identify the idiosyncratic component of each group's achievement and determine whether the components are correlated. For both strategies, I am sensitive to the potential criticism that what appears to be idiosyncratic variation in groups' shares or achievement may actually be a time trend within a grade within a school. (This criticism does not affect estimates based on gender groups under strategy 1.) To address this criticism, I not only eliminate linear time trends: I also eliminate any school from the sample that appears to have a non-linear time pattern. To do this, I determine whether actual years explain more of a school's variation than false, randomly assigned years.

The peer effect estimates generated by the two strategies are reasonably similar. One useful way to state the estimates is in terms of test scores: the effect on a student's own test scores of being surrounded by peers who score 1 point higher. If one translates the peer effect estimates from strategy 1 into test scores, then strategy 1 generates estimates in the range of 0.15 to 0.40. Strategy 2 tends to generate estimates in the range of 0.10 to 0.55.

In addition, by exploring patterns in the estimates generated by the two strategies, I find evidence that the baseline model of peer effects is inadequate. Although I find little evidence that peer achievement has effects that are *generally* non-linear, I do find that peer achievement is not the sole channel for peer effects. The prevalence of females has a positive effect on male math scores that could not plausibly come through females' effect on mean peer achievement in math. Also, the Hispanic share has a positive effect on certain Hispanic students' scores that could not be an effect of mean peer achievement since raising the Hispanic share lowers mean peer achievement. In addition, some results suggest that peer effects are stronger inside racial groups than between racial groups.

References

- Argys, Laura M., Daniel I. Rees, Dominic J. Brewer, "Detracking America's Schools: Equity at Zero Cost?" *Journal of Policy Analysis & Management*, Vol. 15, No. 4 (Fall 1996), 623-45.
- Banerjee, Abhijit V., and Timothy Besley, "Peer Group Externalities and Learning Incentives: A Theory of Nerd Behavior," John M. Olin Program for the Study of Economic Organization and Public Policy Working Paper No. 68 (December 1990), 1-38.
- Benabou, Roland, "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance," *American Economic Review*. Vol. 86, No. 3 (June 1996), 584-609.
- Betts, Julian R, and Darlene Morell, "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects," *Journal of Human Resources*, Vol. 34, No. 2 (Spring 1999), 268-93.
- Brooks-Gunn, Jeanne, Greg J. Duncan, and J. Lawrence Aber, eds., *Neighborhood Poverty: Context and Consequences for Children*. New York: Russell Sage Foundation, 1997.
- Case, Anne C., and Lawrence F. Katz, "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths," NBER Working Paper No. W3705, May 1991.
- de Souza Briggs, Xavier, "Moving Up versus Moving Out: Neighborhood Effects in Housing Mobility Programs," *Housing Policy Debate*, Vol. 8, No. 1 (1997), 195-234.
- Durlauf, Steven N., "Neighborhood Feedbacks, Endogenous Stratification, and Income Inequality," in William A. Barnett, Giancarlo Gandolfo, and Claude Giancarlo, eds., *Dynamic Disequilibrium Modeling: Theory and Applications: Proceedings of the Ninth International Symposium in Economic Theory and Econometrics*. International Symposia in Economic Theory and Econometrics series. Cambridge: Cambridge University Press, 1996, 505-34.
- Epple, Dennis, and Richard E. Romano, "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects," *American Economic Review*, Vol. 88, No. 1 (March 1998), 33-62.
- Kremer, Michael, "The O-Ring Theory of Economic Development," *The Quarterly Journal of Economics*, Vol. 108, No. 3 (August 1993), 551-75.
- Nechyba, Thomas, "Public School Finance in a General Equilibrium Tiebout World: Equalization Programs, Peer Effects, and Private School Vouchers," National Bureau of Economic Research Working Paper No. 5642, June 1996, 1-34.
- Rosenbaum, James E., "Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program," *Housing Policy Debate*, Vol. 6, No. 1 (1995), 231-69.
- Sacerdote, Bruce, "Peer Effects with Random Assignment: Results for Dartmouth Roommates," Dartmouth College, 2000.

Summers, Anita A, and Barbara L. Wolfe, "Do Schools Make a Difference?" *American Economic Review*, Vol. 67, No. 4 (September 1977), 639-52.

Zimmer, Ron W, and Eugenia F. Toma, "Peer Effects in Private and Public Schools across Countries," *Journal of Policy Analysis & Management*, Vol. 19, No. 1 (Winter 2000), 75-92.

Zimmerman, David, "Peer Effects on Academic Outcomes: Evidence from a Natural Experiment," Williams College, 1999.

Table 1
Number and Size of Third Grades and Demographics of Third Graders in Texas

	Number of Schools with a 3rd Grade	Size of the Median 3rd Grade Cohort	Percent of Texas 3rd Graders who are:							
			Female	Native American	Asian	Black	Hispanic	Anglo	Eligible for Free Lunch	Eligible for Reduced Price Lunch
1990-91	3265	79	48.7	0.2	2.2	14.8	30.7	52.2	41.6	(included in free lunch)
1991-92	3161	79	48.6	0.2	2.1	14.9	30.5	52.2	42.3	
1992-93	3201	77	48.7	0.4	2.2	14.9	30.5	52.0	36.8	5.8
1993-94	3256	85	48.7	0.3	2.1	14.1	34.9	48.6	43.8	6.3
1994-95	3285	84	48.7	0.3	2.2	14.1	35.8	47.6	45.1	6.5
1995-96	3329	78	48.6	0.3	2.4	15.2	33.6	48.5	44.3	7.1
1996-97	3408	76	48.7	0.3	2.5	15.4	33.2	48.5	43.3	7.7
1997-98	3439	77	48.8	0.3	2.6	15.7	33.7	47.7	43.4	7.9
1998-99	3512	77	48.9	0.3	2.5	15.7	34.9	46.4	42.8	8.1

Source: Author's calculations based on Texas Schools Microdata Panel. See Appendix Table 1 for comparable results for fourth, fifth, and sixth grades.

Table 2
Reading Scores of Third Graders

	standard deviation (All)	mean test score of third graders who are:										
		All	Female	Male	Native American	Asian	Black	Hispanic	Anglo	Not Disadvan- taged	Eligible Free Lunch	Eligible Reduced Lunch
1990-91	2.3	28.5	29.2	27.9	28.7	30.3	26.6	26.7	30.2	30.1	26.4	(included in free lunch)
1991-92	2.4	28.8	29.4	28.1	28.6	30.6	26.7	26.8	30.4	30.3	26.6	
1992-93	2.6	28.0	28.7	27.4	27.8	29.8	25.9	25.9	29.7	29.5	25.5	27.7
1993-94	2.2	29.5	30.1	29.0	29.1	31.5	27.3	28.1	31.1	31.2	27.6	29.3
1994-95	2.4	29.8	30.4	29.3	29.9	32.2	27.5	28.4	31.4	31.5	27.8	29.7
1995-96	2.4	29.6	30.1	29.1	30.3	31.7	27.2	28.2	31.2	31.4	27.5	29.5
1996-97	2.5	29.5	30.1	28.9	29.0	32.0	27.3	28.0	31.1	31.4	27.3	29.4
1997-98	2.1	30.3	30.8	29.8	29.9	32.5	28.4	29.1	31.6	31.9	28.4	30.1
1998-99	2.1	31.3	31.8	30.9	31.0	33.1	29.0	30.4	32.7	32.7	29.6	31.3

Source: Author's calculations based on Texas Schools Microdata Panel. See Appendix Table 2 for comparable results for fourth, fifth, and sixth grades.

Table 3
Math Scores of Third Graders

	standard deviation (All)	mean test score of third graders who are:										
		All	Female	Male	Native American	Asian	Black	Hispanic	Anglo	Not Disadvan- taged	Eligible Free Lunch	Eligible Reduced Lunch
1990-91	2.6	35.9	35.9	36.0	36.5	38.4	33.3	33.9	37.7	37.4	33.7	(included in free lunch)
1991-92	2.3	36.4	36.4	36.4	35.9	38.8	34.2	34.7	37.9	37.6	34.6	
1992-93	2.6	35.7	35.7	35.7	35.7	38.2	33.1	34.0	37.3	36.9	33.7	35.4
1993-94	3.0	33.1	33.2	33.0	32.3	36.5	29.6	31.5	35.1	35.0	30.8	32.7
1994-95	3.1	34.8	34.9	34.7	34.8	38.2	31.5	33.2	36.7	36.7	32.6	34.6
1995-96	3.1	35.4	35.5	35.3	35.9	38.8	32.1	33.9	37.2	37.4	33.0	35.3
1996-97	2.7	36.5	36.6	36.4	35.9	39.7	33.5	35.3	38.1	38.2	34.4	36.4
1997-98	2.5	36.1	36.1	36.1	35.8	39.2	33.3	34.9	37.7	37.8	34.1	35.9
1998-99	2.4	37.0	36.8	37.2	36.8	39.7	33.7	36.2	38.6	38.5	35.2	36.9

Source: Author's calculations based on Texas Schools Microdata Panel. See Appendix Table 3 for comparable results for fourth, fifth, and sixth grades.

Table 4
The Variation of Interest: Cohort-to-Cohort Changes in the Gender, Race, and Disadvantaged Shares of Third Graders
difference between 1994-95 and 1993-94 used as an example

statistic	first difference between adjacent cohorts in:								
	percent female	percent native american <i>detrended</i>	percent Asian <i>detrended</i>	percent black <i>detrended</i>	percent Hispanic <i>detrended</i>	percent Anglo <i>detrended</i>	percent nondisadvantaged <i>detrended</i>	percent free lunch <i>detrended</i>	percent reduced price lunch <i>detrended</i>
standard deviation	11	2	2	6	8	9	11	11	5
1st percentile	-30	-3	-6	-17	-23	-25	-33	-30	-14
5th percentile	-16	-2	-3	-8	-11	-12	-16	-15	-8
10th percentile	-11	-1	-2	-5	-8	-9	-11	-11	-6
90th percentile	11	1	2	5	8	9	11	11	5
95th percentile	15	2	3	8	11	12	15	15	8
99th percentile	28	3	7	16	22	26	31	32	15

Source: Author's calculations based on Texas Schools Microdata Panel.

Table 5
The Effect of Having a More Female Peer Group
Third through Sixth Grade Regressions using First-Difference Variables (first differences between adjacent cohorts in a school)

each Cell represents a separate regression
and shows coefficient on change in the share of the cohort that is female

	dependent variable is change in mean reading score of students who are:		dependent variable is change in mean math score of students who are:	
	female	male	female	male
third grade	0.374** (0.149) [0.337]**	0.471** (0.174) [0.424]**	0.381* (0.195) [6.561]*	0.396* (0.204) [6.832]*
fourth grade	0.315* (0.153) [0.424]	0.189 (0.215) [0.254]	0.509* (0.266) [2.545]	0.422 (0.258) [2.110]
fifth grade	0.413* (0.188) [0.516]*	0.402* (0.204) [0.503]*	0.603* (0.281) [6.030]*	0.044 (0.294) [0.404]
sixth grade	0.330* (0.158) [0.314]*	0.323* (0.169) [0.308]*	0.640* (0.352) [1.684]*	0.808* (0.419) [2.126]*

Notes: Standard errors in parentheses. The coefficient is significantly different from zero at the 0.01 level if there are two asterisks, at the 0.05 level if there is one asterisk. In square brackets: translation of coefficients into the implied effect of the change in peers' test scores that would occur purely through the change in the share of the cohort that is female. To make this translation, one uses the estimated difference between the genders' true underlying test scores (that is, test scores before peer effects). Method is weighted least squares. The weights account for heteroskedasticity: the dependent variable is a group average. Number of observations is 22,496 in third grade regressions, 19,084 in fourth grade regressions, 14,974 in fifth grade regressions, and 9,743 in sixth grade regressions. An observation is a gender group in a cohort in a school. The dependent variables for third graders have the following means (and standard deviations): 30.1 (2.4) for females in reading, 29.0 (2.8) for males in reading, 35.7 (2.9) for females in math, 35.6 (3.1) for males in math. See Appendix Tables 2 and 3 for descriptive statistics on the dependent variables for other grades. Author's calculations based on Texas Schools Microdata Panel.

Table 6
Non-Linear Effects of Gender Composition?
Effect of a Change in the Share of the Cohort that is Female, for Various Ranges of Percent Female
each Column within a Grade represents a separate regression

			dep. var. is mean reading score of students who are:		dep. var. is mean math score of students who are:	
			female	male	female	male
third grade	effect of a change in the share of the cohort that is female, where the cohort is:	cohort is 0 to 33 percent female	-0.417 (0.332)	-0.233 (0.358)	-0.656 (0.723)	-1.789** (0.658)
		cohort is 33 to 66 percent female	0.569** (0.166)	0.331* (0.168)	0.460* (0.197)	0.451* (0.205)
		cohort is 66 to 100 percent female	-0.512 (0.314)	0.809* (0.384)	-0.350 (0.631)	1.767* (0.773)
fourth grade	effect of a change in the share of the cohort that is female, where the cohort is:	cohort is 0 to 33 percent female	-0.378 (0.889)	-0.214 (0.539)	0.050 (1.194)	-0.486 (0.690)
		cohort is 33 to 66 percent female	0.415* (0.205)	0.122 (0.272)	0.529** (0.236)	0.412 (0.349)
		cohort is 66 to 100 percent female	1.042 (0.820)	0.970* (0.490)	1.622 (1.103)	1.355** (0.682)
fifth grade	effect of a change in the share of the cohort that is female, where the cohort is:	cohort is 0 to 33 percent female	-0.333 (0.516)	-0.226 (0.456)	-0.898 (0.773)	-1.604* (0.715)
		cohort is 33 to 66 percent female	0.253 (0.269)	0.092 (0.286)	1.002** (0.401)	0.582 (0.396)
		cohort is 66 to 100 percent female	1.002* (0.502)	1.105** (0.367)	-0.066 (0.811)	1.832* (0.889)
sixth grade	effect of a change in the share of the cohort that is female, where the cohort is:	cohort is 0 to 33 percent female	0.325 (0.907)	-0.513 (0.843)	0.307 (0.586)	-1.414 (1.284)
		cohort is 33 to 66 percent female	0.259 (0.270)	0.363 (0.306)	0.676** (0.329)	0.886* (0.460)
		cohort is 66 to 100 percent female	0.613* (0.311)	0.734* (0.372)	-0.602 (0.733)	1.458* (0.716)

See notes for previous table.

Table 7a
The Effect of Having Peers from Various Racial Groups
Third Grade Regressions using First-Difference Variables (first differences between adjacent cohorts in a school)

each Column represents a separate regression
and shows coefficients on changes in the share of the cohort who belong to various racial groups

independent variable	dep. var. is change in mean reading score of 3rd graders who are:			dep. var. is change in mean math score of 3rd graders who are:		
	black	Hispanic	Anglo	black	Hispanic	Anglo
change in share of 3rd graders who are Native Am	-1.699 (2.207) [1.019]	0.030 (1.473) [-0.018]	-2.791** (0.600) [1.674]**	2.355 (2.666) [-1.266]	-3.109 (1.742) [1.672]	-0.701 (0.747) [0.377]
change in share of 3rd graders who are Asian	-0.420 (1.099) [-0.663]	-0.634 (0.975) [-1.003]	-0.209 (0.474) [-0.331]	0.417 (1.343) [0.298]	0.553 (1.159) [0.394]	0.377 (0.592) [0.269]
change in share of 3rd graders who are black	-2.501** (0.412) [0.676]**	-0.983* (0.432) [0.266]**	-0.620** (0.243) [0.168]**	-1.863** (0.510) [0.402]**	-0.861* (0.423) [0.185]**	-0.427** (0.201) [0.092]**
change in share of 3rd graders who are Hispanic	-0.420 (0.434) [0.143]	0.056 (0.282) [-0.019]	-0.277 (0.180) [0.078]	-0.155 (0.534) [0.050]	-0.003 (0.340) [0.001]	0.094 (0.225) [-0.030]
p-value: all races have equal effect	0.0003	0.0705	0.0002	0.0585	0.1240	0.4137

Notes: Standard errors in parentheses. The coefficient is significantly different from zero at the 0.01 level if there are two asterisks, at the 0.05 level if there is one asterisk. In square brackets: translation of coefficients into the implied effect of the change in peers' test scores that would occur purely through the change in the share of the cohort that belongs to the racial group. To make this translation, one uses the estimated difference between the racial group's and Anglo's true underlying test scores (that is, test scores before peer effects). Method is instrumental variables with weights. The weights account for heteroskedasticity: the dependent variable is a group average. The instruments are *detrended* changes in the share of third graders who belong to a racial group. The number of observations varies with the racial group whose achievement is the dependent variable: 15,178 for black, 20,368 for Hispanic, 20,127 for Anglo. An observation is a racial group in a cohort in a school. Author's calculations based on Texas Schools Microdata Panel.

Table 7b
Coefficient on Change in the Share of *Third* Graders who belong to Various Racial Groups
Third Grade Regressions using Reduced Sample of Schools that Do Not Show Evidence of Time Trends

each Column represents a separate regression
and shows coefficients on changes in the share of the cohort who belong to various racial groups

independent variable	dep. var. is change in mean reading score of 3rd graders who are:			dep. var. is change in mean math score of 3rd graders who are:		
	black	Hispanic	Anglo	black	Hispanic	Anglo
change in share of 3rd graders who are Native Am	-1.258 (4.061) [0.755]	2.441 (2.701) [-1.464]	-9.539** (1.000) [5.722]**	-0.570 (4.936) [0.307]	-4.759 (3.150) [2.559]	-5.986** (1.225) [3.219]**
change in share of 3rd graders who are Asian	0.413 (1.714) [0.653]	-1.467 (1.556) [-2.319]	0.164 (0.711) [0.259]	4.189* (2.084) [2.991]	0.708 (1.816) [0.506]	0.527 (0.871) [0.376]
change in share of 3rd graders who are black	-2.814** (0.648) [0.761]**	-2.929** (0.656) [0.792]**	-0.678* (0.322) [0.184]*	-1.139* (0.526) [0.245]*	-1.517* (0.766) [0.327]*	-0.577* (0.254) [0.124]*
change in share of 3rd graders who are Hispanic	-0.731 (0.681) [0.249]	-1.058** (0.450) [0.361]**	-0.108 (0.291) [0.037]	-0.903 (0.828) [0.289]	-0.104 (0.526) [0.033]	0.349 (0.357) [-0.112]
p-value: all races have equal effect	0.0437	0.0206	0.0001	0.0242	0.1427	0.4137

Notes: Standard errors in parentheses. The coefficient is significantly different from zero at the 0.01 level if there are two asterisks, at the 0.05 level if there is one asterisk. In square brackets: translation of coefficients into the implied effect of the change in peers' test scores that would occur purely through the change in the share of the cohort that belongs to the racial group. To make this translation, one uses the estimated difference between the racial group's and Anglo's true underlying test scores (that is, test scores before peer effects). Method is weighted least squares, in which the weights account for heteroskedasticity: the dependent variable is a group average. The number of observations is reduced from the number in the previous table because the sample includes only schools that do not show evidence of time trends (the standard of evidence is "drop if more than random"--see text). The number of observations is: 5,608 for black achievement, 6,875 for Hispanic achievement, and 6,928 for Anglo achievement. An observation is a racial group in a cohort in a school. Author's calculations based on Texas Schools Microdata Panel.

Table 8
Non-Linear Effects of Racial Composition?
Effect of a Change in the Share of the Cohort that is Black or Hispanic, for Various Ranges of Percent Black or Hispanic
each Column represents a separate regression

		dep. var. is mean reading score of third graders who are:			dep. var. is mean math score of third graders who are:		
		black	Hispanic	Anglo	black	Hispanic	Anglo
effect of change in share of 3rd graders who are black, where cohort is:	0 to 33 percent black	-0.827 (0.531)	-0.357 (0.470)	-0.189 (0.254)	-0.313 (0.634)	-1.107* (0.550)	0.008 (0.311)
	33 to 66 percent black	-2.503** (0.507)	-1.362 (1.184)	-0.933* (0.461)	-2.412** (0.605)	1.192 (0.792)	-1.146* (0.562)
	66 to 100 percent black	0.111 (0.615)	-1.062 (0.439)	-2.625* (1.261)	1.347 (0.734)	-0.538 (1.384)	-1.090 (1.538)
effect of change in share of 3rd graders who are Hispanic, where cohort is:	0 to 33 percent Hispanic	-0.222 (0.492)	-1.063** (0.439)	-0.115 (0.210)	0.740 (0.587)	-1.346** (0.514)	-0.081 (0.256)
	33 to 66 percent Hispanic	-0.351 (0.590)	0.143 (0.367)	-0.099 (0.289)	-0.683 (0.704)	0.226 (0.429)	0.240 (0.352)
	66 to 100 percent Hispanic	1.600 (1.035)	0.678* (0.330)	0.147 (0.582)	0.694 (1.235)	0.813* (0.403)	0.096 (0.708)

See notes for previous table. Specification is the same, except that the change in the share of students who are black (and Hispanic) is interacted with three indicator variables for the share of the cohort that is black (and Hispanic).

Table 9
Effect of Females' Unexpected Performance on the Unexpected Performance of their Male Peers (and *vice versa*)

each Cell represents a separate regression

dependent variable: residual from a school-grade-gender specific regression of test scores on a time trend and cohort gender and racial composition
explanatory variables: year indicator variables, residual from a school-grade-gender specific regression of test scores on a time trend and cohort gender and racial composition (residual is multiplied by group's share of cohort)

dependent variable	explanatory variable of interest	coefficient for reading regression	coefficient for math regression
male 3rd graders' residuals	female 3rd graders' residuals	0.444** (0.029)	0.622** (0.021)
male 4th graders' residuals	female 4th graders' residuals	0.414** (0.031)	0.489** (0.024)
male 5th graders' residuals	female 5th graders' residuals	0.325** (0.033)	0.423** (0.032)
male 6th graders' residuals	female 6th graders' residuals	0.330** (0.036)	0.388** (0.031)
male 5th graders' residuals	female 3rd graders' residuals	0.081** (0.018)	0.056** (0.020)
female 3rd graders' residuals	male 3rd graders' residuals	0.385** (0.024)	0.609** (0.019)
female 4th graders' residuals	male 4th graders' residuals	0.352** (0.025)	0.479** (0.026)
female 5th graders' residuals	male 5th graders' residuals	0.316** (0.032)	0.398** (0.031)
female 6th graders' residuals	male 6th graders' residuals	0.285** (0.031)	0.384** (0.031)
female 5th graders' residuals	male 3rd graders' residuals	0.079** (0.017)	0.055** (0.020)

Notes: An observation is at the school-cohort-grade-gender group level. Each cell represents a separate regression which includes year indicator variables as well as the variable of interest shown. Method is least squares with robust standard errors that allow for school clustering. There are 28,733 observations for third grade cohorts, 18,536 observations for fourth grade cohorts, 14,899 observations for fifth grade cohorts, and 12,048 observations for sixth grade cohorts.

Table 10
Are Ostensible Peer Effects Really Teacher Effects or Time Trends?
Specification Tests for Effect of Females' Unexpected Performance on the Unexpected Performance of their Male Peers

each Cell represents a separate regression

Specification is identical to that of previous table. Only sample differs.

Sample is schools with low teacher turnover

dependent variable	explanatory variable of interest	coefficient for reading regression	coefficient for math regression
male 3rd graders' residuals	female 3rd graders' residuals	0.570** (0.020)	0.745** (0.014)
male 4th graders' residuals	female 4th graders' residuals	0.556** (0.020)	0.582** (0.018)
male 5th graders' residuals	female 5th graders' residuals	0.514** (0.058)	0.552** (0.049)
male 6th graders' residuals	female 6th graders' residuals	0.535** (0.023)	0.576** (0.022)

Sample is schools with no apparent time trend

dependent variable	explanatory variable of interest	coefficient for reading regression	coefficient for math regression
male 3rd graders' residuals	female 3rd graders' residuals	0.592** (0.072)	0.639** (0.049)
male 4th graders' residuals	female 4th graders' residuals	0.572** (0.076)	0.501** (0.066)
male 5th graders' residuals	female 5th graders' residuals	0.564** (0.203)	0.533** (0.134)
male 6th graders' residuals	female 6th graders' residuals	0.613** (0.087)	0.554** (0.067)

Notes: See notes for previous table. In schools with low teacher turnover, fewer than 10 percent of teaching slots turn over in each six-year period. A school is classified as having no apparent time trend if a regression that is quartic in time does not explain at least 1.05 times as much variation in student performance when actual years are used than when a false year is randomly assigned.

Table 11
Effect of Racial Groups' Unexpected Reading Performance on the Unexpected Reading Performance of their Peers from Another Racial Group

each Row represents a separate regression based on Reading scores

dependent variable: residual from a school-grade-race specific regression of test scores on a time trend and cohort gender and racial composition
explanatory variables: year indicator variables; residuals from school-grade-race specific regressions of test scores on a time trend and cohort gender and racial composition

each residual is multiplied by its group's share of the cohort, so that if all races had an equal effect, their coefficients would be identical

dependent variable	explanatory variables of interest	coefficient on the residual of students who are:					p-value: all races have equal effect
		Native Amer	Asian	black	Hispanic	Anglo	
black 3rd graders' residuals	3rd graders' residuals	-0.512 (2.127)	0.783** (0.322)		0.652** (0.058)	0.806** (0.069)	0.435
black 4th graders' residuals	4th graders' residuals	0.948 (1.920)	1.553** (0.362)		0.600** (0.063)	0.678** (0.097)	0.087
black 5th graders' residuals	5th graders' residuals	-0.368 (0.816)	0.769 (0.571)		0.401** (0.095)	0.435** (0.103)	0.701
black 6th graders' residuals	6th graders' residuals	2.652 (6.772)	1.080 (0.713)		0.558** (0.118)	0.551** (0.155)	0.900
black 5th graders' residuals	3rd graders' residuals	0.013 (5.013)	0.098 (0.321)		0.075* (0.034)	0.081* (0.039)	0.956
Hispanic 3rd graders' residuals	3rd graders' residuals	1.270 (1.162)	1.375** (0.301)	0.827** (0.073)		0.651** (0.049)	0.031
Hispanic 4th graders' residuals	4th graders' residuals	1.278* (0.617)	1.009** (0.316)	0.757** (0.079)		0.556** (0.062)	0.113
Hispanic 5th graders' residuals	5th graders' residuals	1.486 (0.926)	0.501 (0.444)	0.716** (0.102)		0.376** (0.087)	0.073
Hispanic 6th graders' residuals	6th graders' residuals	-0.546 (0.369)	1.106 (0.805)	0.885** (0.175)		0.550** (0.087)	0.003
Hispanic 5th graders' residuals	3rd graders' residuals	0.022 (4.835)	0.508 (0.305)	0.087* (0.041)		0.060* (0.027)	0.862
Anglo 3rd graders' residuals	3rd graders' residuals	1.188 (1.860)	0.782** (0.220)	0.584** (0.061)	0.454** (0.040)		0.043
Anglo 4th graders' residuals	4th graders' residuals	0.298 (0.689)	0.869** (0.357)	0.441** (0.074)	0.413** (0.043)		0.640
Anglo 5th graders' residuals	5th graders' residuals	1.051 (0.861)	0.705** (0.273)	0.335** (0.097)	0.288** (0.053)		0.394
Anglo 6th graders' residuals	6th graders' residuals	1.025 (0.801)	1.300** (0.409)	0.637** (0.124)	0.400** (0.066)		0.062
Anglo 5th graders' residuals	3rd graders' residuals	0.045 (0.648)	0.074 (0.146)	0.059* (0.033)	0.048* (0.018)		0.920

Each row represents a separate regression which includes year indicator variables as well as the variables of interest shown. Method is least squares with robust standard errors that allow for school clustering. 28,733 observations for third grade cohorts, 18,536 observations for fourth grade cohorts, 14,899 observations for fifth grade cohorts, 12,048 observations for sixth grade cohorts.

Table 12
Effect of Racial Groups' Unexpected Math Performance on the Unexpected Math Performance of their Peers from Another Racial Group

each Row represents a separate regression based on Math scores

dependent variable: residual from a school-grade-race specific regression of test scores on a time trend and cohort gender and racial composition
explanatory variables: year indicator variables; residuals from school-grade-race specific regressions of test scores on a time trend and cohort gender and racial composition

each residual is multiplied by its group's share of the cohort, so that if all races had an equal effect, their coefficients would be identical

dependent variable	explanatory variables of interest	coefficient on the residual of students who are:					p-value: all races have equal effect
		Native Amer	Asian	black	Hispanic	Anglo	
black 3rd graders' residuals	3rd graders' residuals	0.480 (1.893)	1.864** (0.327)		0.825** (0.052)	1.055** (0.052)	0.001
black 4th graders' residuals	4th graders' residuals	-0.026 (2.702)	1.660** (0.545)		0.633** (0.070)	0.784** (0.074)	0.064
black 5th graders' residuals	5th graders' residuals	-0.450 (9.358)	1.725** (0.608)		0.448** (0.090)	0.672** (0.097)	0.093
black 6th graders' residuals	6th graders' residuals	-0.875 (6.221)	1.086 (0.628)		0.710** (0.104)	0.721** (0.139)	0.937
black 5th graders' residuals	3rd graders' residuals	0.017 (6.147)	0.234 (0.355)		0.086* (0.039)	0.069* (0.035)	0.807
Hispanic 3rd graders' residuals	3rd graders' residuals	1.075 (1.138)	1.426** (0.296)	0.856** (0.053)		0.898** (0.040)	0.206
Hispanic 4th graders' residuals	4th graders' residuals	1.109 (1.677)	1.261** (0.370)	0.767** (0.069)		0.748** (0.059)	0.394
Hispanic 5th graders' residuals	5th graders' residuals	1.835 (3.856)	1.750** (0.636)	0.701** (0.102)		0.610** (0.066)	0.243
Hispanic 6th graders' residuals	6th graders' residuals	-0.087 (0.808)	1.264 (0.791)	0.740** (0.121)		0.567** (0.074)	0.384
Hispanic 5th graders' residuals	3rd graders' residuals	0.018 (5.651)	0.356 (0.320)	0.082* (0.036)		0.055* (0.027)	0.745
Anglo 3rd graders' residuals	3rd graders' residuals	1.077 (0.886)	1.252** (0.229)	0.747** (0.060)	0.632** (0.032)		0.137
Anglo 4th graders' residuals	4th graders' residuals	1.501 (0.803)	1.113** (0.267)	0.589** (0.071)	0.556** (0.047)		0.140
Anglo 5th graders' residuals	5th graders' residuals	0.461 (2.017)	1.256** (0.337)	0.464** (0.085)	0.435** (0.056)		0.131
Anglo 6th graders' residuals	6th graders' residuals	1.034 (0.802)	1.036 (0.646)	0.600** (0.205)	0.600** (0.078)		0.806
Anglo 5th graders' residuals	3rd graders' residuals	0.034 (1.144)	0.119 (0.267)	0.080* (0.038)	0.044* (0.021)		0.731

Notes: Each row represents a separate regression which includes year indicator variables as well as the variables of interest shown. Method is least squares with robust standard errors that allow for school clustering. 28,733 observations for third grade cohorts, 18,536 observations for fourth grade cohorts, 14,899 observations for fifth grade cohorts, 12,048 observations for sixth grade cohorts.

Table 13
Are Ostensible Peer Effects Really Teacher Effects or Time Trends?
Specification Tests for Effect of Racial Groups' Unexpected Performance on the Unexpected Performance of their Peers from Another Racial Group

each Row represents a separate regression based on Math scores

Specification is identical to that of previous table. Only sample differs.

Sample is schools with low teacher turnover		coefficient on the residual of students who are:				p-value: all races have equal effect
dependent variable	explanatory variables of interest	Native Amer	Asian	black	Hispanic	
Anglo 3rd graders' residuals	3rd graders' residuals	0.685 (1.902)	1.811** (0.362)	0.818** (0.095)	0.692** (0.058)	0.015
Anglo 4th graders' residuals	4th graders' residuals	1.187 (2.241)	1.394** (0.317)	0.573** (0.095)	0.593** (0.069)	0.284
Anglo 5th graders' residuals	5th graders' residuals	0.763 (2.387)	1.318** (0.343)	0.635** (0.096)	0.488** (0.062)	0.023
Anglo 6th graders' residuals	6th graders' residuals	1.641* (0.686)	1.811** (0.570)	0.549* (0.284)	0.734** (0.107)	0.206

Sample is schools with no apparent time trend		coefficient on the residual of students who are:				p-value: all races have equal effect
dependent variable	explanatory variables of interest	Native Amer	Asian	black	Hispanic	
Anglo 3rd graders' residuals	3rd graders' residuals	-0.938 (3.182)	1.582** (0.565)	0.815** (0.127)	0.595** (0.061)	0.110
Anglo 4th graders' residuals	4th graders' residuals	1.248 (1.044)	1.378** (0.481)	0.551** (0.108)	0.513** (0.071)	0.034
Anglo 5th graders' residuals	5th graders' residuals	0.882 (2.232)	0.927* (0.399)	0.382** (0.099)	0.387** (0.060)	0.511
Anglo 6th graders' residuals	6th graders' residuals	0.905 (0.649)	1.477* (0.774)	0.468* (0.229)	0.525** (0.087)	0.632

Notes: See notes for previous table. In schools with low teacher turnover, fewer than 10 percent of teaching slots turn over in each six-year period. A school is classified as having no apparent time trend if a regression that is quartic in time does not explain at least 1.05 times as much variation in student performance when actual years are used than when a false year is randomly assigned.

Table 14
Non-Linear Peer Effects?
Quadratic Specifications for Effect of Groups' Unexpected Performance on the Unexpected Performance of their Peers from Another Group

each Row represents a separate regression based on *Reading* scores

dependent variable	explanatory variable of interest	coefficient on linear term	coefficient on quadratic term
male 3rd graders' reading residuals	female 3rd graders' reading residuals	0.445** (0.029)	0.004 (0.007)
male 4th graders' reading residuals	female 4th graders' reading residuals	0.415** (0.031)	0.004 (0.004)
male 5th graders' reading residuals	female 5th graders' reading residuals	0.324** (0.036)	0.008 (0.008)
male 6th graders' reading residuals	female 6th graders' reading residuals	0.330** (0.036)	0.004 (0.004)

each Row represents a separate regression based on *Math* scores

dependent variable	explanatory variable of interest	coefficient on linear term	coefficient on quadratic term
male 3rd graders' math residuals	female 3rd graders' math residuals	0.621** (0.021)	-0.002 (0.004)
male 4th graders' math residuals	female 4th graders' math residuals	0.489** (0.024)	-0.002 (0.004)
male 5th graders' math residuals	female 5th graders' math residuals	0.423** (0.032)	-0.004 (0.008)
male 6th graders' math residuals	female 6th graders' math residuals	0.387** (0.031)	-0.003 (0.004)

Notes: An observation is at the school-cohort-grade-gender group level. Each row represents a separate regression which includes year indicator variables as well as the variable of interest shown. Method is least squares with robust standard errors that allow for school clustering. There are 28,733 observations for third grade cohorts, 18,536 observations for fourth grade cohorts, 14,899 observations for fifth grade cohorts, and 12,048 observations for sixth grade cohorts.

Appendix Table 1
 Number and Size of Fourth, Fifth, and Sixth Grades
 and Demographics of Fourth, Fifth, and Sixth Graders in Texas, early and late 1990s

	Number of Schools with this Grade	Size of the Median Cohort in this Grade	Percent of Texas Students in this Grade who are:							
			Female	Native American	Asian	Black	Hispanic	Anglo	Free Lunch	Reduced Price Lunch
4th Grade 1992-93	3,172	86	48.6	0.3	2.1	14.0	35.0	48.7	42.5	5.9
4th Grade 1998-99	3,482	79	48.9	0.4	2.6	15.3	35.8	46.0	42.9	8.2
5th Grade 1993-94	3,064	83	48.6	0.2	2.2	13.9	35.2	48.5	42.2	6.1
5th Grade 1998-99	3,278	77	48.7	0.3	2.6	14.7	36.4	46.0	42.6	7.9
6th Grade 1993-94	2,103	84	48.6	0.2	2.2	14.1	35.0	48.5	39.7	5.9
6th Grade 1998-99	2,240	79	48.6	0.3	2.5	14.6	37.4	45.3	41.5	7.3

Source: Author's calculations based on Texas Education Agency data.

Appendix Table 2
Reading Scores of Fourth, Fifth, and Sixth Graders in Texas, early and late 1990s

	standard deviation (All)	Mean Reading Score of Students in this Grade who are:										
	All	Female	Male	Native American	Asian	Black	Hispanic	Anglo	Not Disadvnt	Free Lunch	Reduced Lunch	
4th Grade 1992-93	3.5	27.6	28.2	27.1	27.3	30.4	24.1	24.9	30.4	30.2	24.3	27.0
4th Grade 1998-99	2.3	34.3	34.8	33.9	34.2	36.5	32.0	33.1	35.9	36.0	32.4	34.2
5th Grade 1993-94	2.5	30.2	30.7	29.8	30.1	32.9	27.5	28.5	32.2	32.1	27.8	29.8
5th Grade 1998-99	2.3	34.1	34.3	33.9	34.4	36.0	32.0	32.6	35.9	35.9	32.0	33.9
6th Grade 1993-94	2.9	28.9	29.4	28.5	29.0	32.1	25.9	26.6	31.3	31.1	25.9	28.5
6th Grade 1998-99	2.4	32.6	33.2	32.1	32.7	34.9	30.7	30.7	34.6	34.5	30.1	32.4

Appendix Table 3
Math Scores of Fourth, Fifth, and Sixth Graders in Texas, early and late 1990s

	standard deviation (All)	Mean Math Score of Students in this Grade who are:										
		All	Female	Male	Native American	Asian	Black	Hispanic	Anglo	Not Disadvnt	Free Lunch	Reduced Lunch
4th Grade 1992-93	4.1	35.8	36.1	35.6	35.8	40.8	31.2	33.3	38.7	38.5	32.4	35.2
4th Grade 1998-99	2.9	42.4	42.4	42.3	41.8	45.8	38.7	41.6	44.0	44.1	40.3	42.3
5th Grade 1993-94	3.6	38.1	38.3	37.9	37.7	43.5	33.5	36.0	40.6	40.5	35.0	37.5
5th Grade 1998-99	2.9	43.4	43.4	43.5	43.4	47.2	39.5	42.4	45.3	45.2	41.2	43.2
6th Grade 1993-94	4.2	40.4	41.1	39.7	39.4	46.6	35.1	37.5	43.6	43.2	36.4	39.7
6th Grade 1998-99	3.3	46.6	46.9	46.4	46.6	50.5	42.8	44.7	49.1	48.8	43.8	46.5

Source: Author's calculations based on Texas Education Agency data.

Appendix Table 4
Scores of Third Graders in 1994-95

	in schools that are:				in schools that are:			
	less than 1 percent black	1 to 6 percent black	6 to 20 percent black	more than 20 percent black	less than 10 percent Hispanic	10 to 25 percent Hispanic	25 to 60 percent Hispanic	more than 60 percent Hispanic
Asian-Anglo reading differential	0.8	1.0	1.2	1.1	1.0	0.9	0.7	1.2
black-Anglo reading differential	-2.6	-2.4	-3.1	-3.1	-2.9	-3.1	-3.2	-2.3
Hispanic-Anglo reading differential	-2.0	-1.8	-1.8	-1.5	-1.3	-1.9	-2.1	-1.7
Asian-Anglo math differential	1.8	1.7	1.6	1.9	1.9	1.6	1.1	2.1
black-Anglo math differential	-3.1	-3.8	-4.6	-4.2	-4.1	-4.4	-4.4	-3.9
Hispanic-Anglo math differential	-2.4	-2.4	-2.4	-1.8	-1.8	-2.5	-2.7	-1.9

Source: Author's calculations based on Texas Schools Microdata Panel.

Appendix Table 5a
The Effect of Having Peers from Various Racial Groups
Fourth Grade Regressions using First-Difference Variables (first differences between adjacent cohorts in a school)

each Column represents a separate regression
and shows coefficients on changes in the share of the cohort who belong to various racial groups

independent variable	dep. var. is change in mean reading score of 4th graders who are:			dep. var. is change in mean math score of 4th graders who are:		
	black	Hispanic	Anglo	black	Hispanic	Anglo
change in share of 4th graders who are Native Am	-9.105** (2.718) [4.823]**	0.118 (2.455) [-0.063]	-0.870 (1.351) [0.461]	-12.468** (3.448) [5.256]**	2.432 (3.151) [-1.025]	-1.562 (1.804) [0.658]
change in share of 4th graders who are Asian	1.285 (1.389) [1.373]	0.421 (1.273) [0.450]	-0.227 (0.627) [-0.242]	3.432* (1.762) [1.403]*	-0.021 (1.634) [-0.009]	0.436 (0.837) [0.178]
change in share of 4th graders who are black	0.293 (0.546) [-0.064]	-1.201* (0.560) [0.262]*	0.224 (0.327) [-0.048]	-0.400 (0.693) [0.067]	-2.999** (0.720) [0.502]**	-1.037** (0.436) [0.174]**
change in share of 4th graders who are Hispanic	0.380 (0.593) [-0.112]	-0.817* (0.377) [0.241]*	-0.029 (0.247) [0.009]	0.374 (0.752) [-0.106]	-1.657** (0.483) [0.472]**	-0.668* (0.329) [0.190]*
p-value: all races have equal effect	0.0045	0.9911	0.8146	0.0005	0.5874	0.9362

Notes: Standard errors in parentheses. The coefficient is significantly different from zero at the 0.01 level if there are two asterisks, at the 0.05 level if there is one asterisk. In square brackets: translation of coefficients into the implied effect of the change in peers' test scores that would occur purely through the change in the share of the cohort that belongs to the racial group. To make this translation, one uses the estimated difference between the racial group's and Anglo's true underlying test scores (that is, test scores before peer effects). Method is instrumental variables with weights. The weights account for heteroskedasticity: the dependent variable is a group average. The instruments are *detrended* changes in the share of fourth graders who belong to a racial group. The number of observations varies with the racial group whose achievement is the dependent variable: 12,962 for black achievement, 17,435 for Hispanic achievement, 17,049 for Anglo achievement. An observation is a racial group in a cohort in a school. Author's calculations based on Texas Schools Microdata Panel.

Appendix Table 5b
Coefficient on Change in the Share of *Fourth* Graders who belong to Various Racial Groups
Fourth Grade Regressions using Reduced Sample of Schools that Do Not Show Evidence of Time Trends

each Column represents a separate regression
and shows coefficients on changes in the share of the cohort who belong to various racial groups

	dep. var. is change in mean reading score of 4th graders who are:			dep. var. is change in mean math score of 4th graders who are:		
independent variable	black	Hispanic	Anglo	black	Hispanic	Anglo
change in share of 4th graders who are Native Am	-7.897 (5.431) [4.183]	0.618 (3.802) [-0.327]	-0.451 (2.010) [0.239]	-12.031* (6.019) [5.072]*	4.820 (4.915) [-2.031]	-1.178 (2.693) [0.496]
change in share of 4th graders who are Asian	0.099 (2.005) [0.106]	1.612 (1.840) [1.722]	0.573 (0.939) [0.613]	3.125 (2.521) [1.278]	2.080 (2.378) [0.850]	1.249 (1.258) [0.511]
change in share of 4th graders who are black	0.269 (0.748) [-0.059]	-1.459* (0.770) [0.318]*	-0.144 (0.455) [0.031]	-0.833 (0.940) [0.138]	-2.755** (0.997) [0.461]**	-1.115* (0.510) [0.187]*
change in share of 4th graders who are Hispanic	-0.636 (0.809) [0.187]	-0.905* (0.440) [0.267]*	-0.248 (0.360) [0.073]	-0.696 (1.017) [0.198]	-1.524* (0.701) [0.435]*	-0.075* (0.482) [0.022]*
p-value: all races have equal effect	0.2769	0.8735	0.8953	0.1696	0.5968	0.7900

Notes: Standard errors in parentheses. The coefficient is significantly different from zero at the 0.01 level if there are two asterisks, at the 0.05 level if there is one asterisk. In square brackets: translation of coefficients into the implied effect of the change in peers' test scores that would occur purely through the change in the share of the cohort that belongs to the racial group. To make this translation, one uses the estimated difference between the racial group's and Anglo's true underlying test scores (that is, test scores before peer effects). Method is weighted least squares, in which the weights account for heteroskedasticity: the dependent variable is a group average. The number of observations is reduced from the number in the previous table because the sample includes only schools that do not show evidence of time trends (the standard of evidence is "drop if more than random"--see text). The number of observations is: 5,955 for black achievement, 7,310 for Hispanic achievement, and 7,127 for Anglo achievement. An observation is a racial group in a cohort in a school. Author's calculations based on Texas Schools Microdata Panel.

Appendix Table 6a
The Effect of Having Peers from Various Racial Groups
Fifth Grade Regressions using First-Difference Variables (first differences between adjacent cohorts in a school)

each Column represents a separate regression
and shows coefficients on changes in the share of the cohort who belong to various racial groups

independent variable	dep. var. is change in mean reading score of 5th graders who are:			dep. var. is change in mean math score of 5th graders who are:		
	black	Hispanic	Anglo	black	Hispanic	Anglo
change in share of 5th graders who are Native Am	-2.294 (3.642) [1.513]	2.518 (2.591) [-1.660]	0.529 (0.775) [-0.349]	0.298 (5.300) [-0.137]	0.952 (3.700) [-0.439]	1.540 (1.188) [-0.710]
change in share of 5th graders who are Asian	1.465 (1.362) [2.032]	1.688 (1.209) [2.343]	0.301 (0.601) [0.418]	1.046 (1.981) [0.431]	1.852 (1.726) [0.764]	0.718* (0.364) [0.296]*
change in share of 5th graders who are black	-1.279** (0.546) [0.323]**	-0.604* (0.310) [0.152]*	-0.582 (0.320) [0.147]	-2.753** (0.794) [0.443]**	-0.995* (0.473) [0.160]*	-0.279 (0.492) [0.045]
change in share of 5th graders who are Hispanic	0.402 (0.603) [-0.124]	-1.420** (0.375) [0.439]**	-0.334 (0.241) [0.103]	-0.252 (0.877) [0.072]	-2.047** (0.536) [0.587]**	-0.612* (0.310) [0.176]*
p-value: all races have equal effect	0.0271	0.0480	0.0745	0.3591	0.0320	0.3095

Notes: Standard errors in parentheses. The coefficient is significantly different from zero at the 0.01 level if there are two asterisks, at the 0.05 level if there is one asterisk. In square brackets: translation of coefficients into the implied effect of the change in peers' test scores that would occur purely through the change in the share of the cohort that belongs to the racial group. To make this translation, one uses the estimated difference between the racial group's and Anglo's true underlying test scores (that is, test scores before peer effects). Method is instrumental variables with weights. The weights account for heteroskedasticity: the dependent variable is a group average. The instruments are *detrended* changes in the share of fifth graders who belong to a racial group. The number of observations varies with the racial group whose achievement is the dependent variable: 10,119 for black achievement, 13,749 for Hispanic achievement, 13,328 for Anglo achievement. An observation is a racial group in a cohort in a school. Author's calculations based on Texas Schools Microdata Panel.

Appendix Table 6b
Coefficient on Change in the Share of *Fifth* Graders who belong to Various Racial Groups
Fifth Grade Regressions using Reduced Sample of Schools that Do Not Show Evidence of Time Trends

each Column represents a separate regression
and shows coefficients on changes in the share of the cohort who belong to various racial groups

independent variable	dep. var. is change in mean reading score of 5th graders who are:			dep. var. is change in mean math score of 5th graders who are:		
	black	Hispanic	Anglo	black	Hispanic	Anglo
change in share of 5th graders who are Native Am	-1.326 (4.289) [0.874]	4.730 (3.052) [-3.120]	0.653 (0.860) [-0.430]	-0.583 (6.221) [0.269]	-0.361 (4.402) [0.167]	1.587 (1.327) [-0.733]
change in share of 5th graders who are Asian	1.892 (1.683) [2.625]	1.389 (1.509) [1.927]	0.423 (0.746) [0.587]	3.345 (2.441) [1.380]	0.981 (2.176) [0.405]	0.761* (1.151) [0.314]*
change in share of 5th graders who are black	-1.270* (0.650) [0.318]*	-1.814** (0.670) [0.458]**	-0.009 (0.379) [0.002]	-1.823* (0.904) [0.293]*	-2.357** (0.966) [0.379]**	-0.704 (0.585) [0.113]
change in share of 5th graders who are Hispanic	1.184 (0.722) [-0.366]	-2.023** (0.465) [0.626]**	-0.486 (0.304) [0.151]	-1.850 (1.047) [0.530]	-2.889** (0.671) [0.829]**	-1.314** (0.469) [0.377]**
p-value: all races have equal effect	0.0026	0.1848	0.4549	0.0084	0.1082	0.1419

Notes: Standard errors in parentheses. The coefficient is significantly different from zero at the 0.01 level if there are two asterisks, at the 0.05 level if there is one asterisk. In square brackets: translation of coefficients into the implied effect of the change in peers' test scores that would occur purely through the change in the share of the cohort that belongs to the racial group. To make this translation, one uses the estimated difference between the racial group's and Anglo's true underlying test scores (that is, test scores before peer effects). Method is weighted least squares, in which the weights account for heteroskedasticity: the dependent variable is a group average. The number of observations is reduced from the number in the previous table because the sample includes only schools that do not show evidence of time trends (the standard of evidence is "drop if more than random"--see text). The number of observations is: 6,087 for black achievement, 7,714 for Hispanic achievement, and 7,522 for Anglo achievement. An observation is a racial group in a cohort in a school. Author's calculations based on Texas Schools Microdata Panel.

Appendix Table 7a
The Effect of Having Peers from Various Racial Groups
Sixth Grade Regressions using First-Difference Variables (first differences between adjacent cohorts in a school)

each Column represents a separate regression
and shows coefficients on changes in the share of the cohort who belong to various racial groups

	dep. var. is change in mean reading score of 6th graders who are:			dep. var. is change in mean math score of 6th graders who are:		
independent variable	black	Hispanic	Anglo	black	Hispanic	Anglo
change in share of 6th graders who are Native Am	-0.978 (2.757) [0.514]	4.582 (3.314) [-2.406]	4.066 (2.904) [-2.135]	-8.068* (4.176) [2.902]*	-2.285 (4.933) [0.822]	3.620 (2.742) [-1.303]
change in share of 6th graders who are Asian	0.559 (1.876) [-1.684]	1.220 (1.784) [3.668]	1.160 (0.912) [3.492]	0.245 (2.840) [0.113]	0.358 (2.655) [0.164]	2.022* (1.033) [0.926]*
change in share of 6th graders who are black	-1.978** (0.719) [0.422]**	-0.628 (0.768) [0.134]	-0.645* (0.321) [0.138]*	-2.000* (1.005) [0.273]*	-0.662 (1.142) [0.085]	-0.940* (0.441) [0.128]*
change in share of 6th graders who are Hispanic	-0.107 (0.767) [0.023]	-0.936* (0.482) [0.209]*	-0.024 (0.330) [0.006]	-0.224 (1.163) [0.042]	-1.915** (0.754) [0.357]**	-0.457* (0.520) [0.085]*
p-value: all races have equal effect	0.0865	0.0643	0.0454	0.0938	0.4014	0.3015

Notes: Standard errors in parentheses. The coefficient is significantly different from zero at the 0.01 level if there are two asterisks, at the 0.05 level if there is one asterisk. In square brackets: translation of coefficients into the implied effect of the change in peers' test scores that would occur purely through the change in the share of the cohort that belongs to the racial group. To make this translation, one uses the estimated difference between the racial group's and Anglo's true underlying test scores (that is, test scores before peer effects). Method is instrumental variables with weights. The weights account for heteroskedasticity: the dependent variable is a group average. The instruments are *detrended* changes in the share of sixth graders who belong to a racial group. The number of observations varies with the racial group whose achievement is the dependent variable: 6,558 for black achievement, 8,739 for Hispanic achievement, 8,920 for Anglo achievement. An observation is a racial group in a cohort in a school. Author's calculations based on Texas Schools Microdata Panel.

Appendix Table 7b
Coefficient on Change in the Share of *Sixth* Graders who belong to Various Racial Groups
Sixth Grade Regressions using Reduced Sample of Schools that Do Not Show Evidence of Time Trends

each Column represents a separate regression
and shows coefficients on changes in the share of the cohort who belong to various racial groups

independent variable	dep. var. is change in mean reading score of 6th graders who are:			dep. var. is change in mean math score of 6th graders who are:		
	black	Hispanic	Anglo	black	Hispanic	Anglo
change in share of 6th graders who are Native Am	-3.884 (2.991) [2.040]	1.708 (4.450) [-0.897]	-3.303 (2.303) [-1.734]	-12.539** (4.567) [4.511]**	-1.792 (6.543) [0.645]	-5.439 (3.604) [1.956]
change in share of 6th graders who are Asian	-1.973 (2.338) [5.935]	0.097 (2.311) [0.292]	1.085 (1.175) [3.262]	-3.794 (3.566) [-1.737]	2.915 (3.398) [1.335]	1.426 (1.838) [0.652]
change in share of 6th graders who are black	-2.922** (0.867) [0.623]**	-0.413 (0.917) [0.088]	-1.241** (0.522) [0.265]**	-2.092* (1.050) [0.286]*	-0.582 (1.348) [0.079]	-2.013** (0.816) [0.275]**
change in share of 6th graders who are Hispanic	-0.867 (0.939) [0.194]	-1.442* (0.633) [0.322]*	-0.283 (0.426) [0.063]	-0.010 (1.436) [0.002]	-2.525** (0.928) [0.470]**	-0.195 (0.667) [0.036]
p-value: all races have equal effect	0.0589	0.1508	0.0561	0.0181	0.1509	0.1400

Notes: Standard errors in parentheses. The coefficient is significantly different from zero at the 0.01 level if there are two asterisks, at the 0.05 level if there is one asterisk. In square brackets: translation of coefficients into the implied effect of the change in peers' test scores that would occur purely through the change in the share of the cohort that belongs to the racial group. To make this translation, one uses the estimated difference between the racial group's and Anglo's true underlying test scores (that is, test scores before peer effects). Method is weighted least squares, in which the weights account for heteroskedasticity: the dependent variable is a group average. The number of observations is reduced from the number in the previous table because the sample includes only schools that do not show evidence of time trends (the standard of evidence is "drop if more than random"--see text). The number of observations is: 4,005 for black achievement, 5,219 for Hispanic achievement, and 5,209 for Anglo achievement. An observation is a racial group in a cohort in a school. Author's calculations based on Texas Schools Microdata Panel.