

Educational Peer Effects and the Chicago Public Schools*

Lars Lefgren

Department of Economics
Brigham Young University
130 Faculty Office Building
Provo, UT 84602-2363

Phone: (801) 422-5169
Fax: (801) 422-2844
E-mail: l-lefgren@byu.edu

* I would like to thank the Consortium on Chicago School Research and the Chicago Public Schools for providing the data used in this study. I am grateful to Eric Bettinger, Richard Butler, Julie Cullen, Mark Duggan, Michael Greenstone, Jonathan Guryan, Brian Jacob, Steve Levitt, Brigitte Madrian, Kevin Murphy, Mark Showalter, Robert Topel, and two anonymous referees for helpful feedback. All remaining errors are my own.

Educational Peer Effects and the Chicago Public Schools

Abstract

Using data on third and sixth grade students in the Chicago Public Schools, I examine peer effects using variation in school tracking policies. In tracked schools, high ability students receive the benefit of being placed in classes with high ability peers. The opposite is the case for low ability students. If peer effects were important, one would expect students with high initial ability in tracked schools to outperform similar students in untracked schools. By the same token, students with low initial ability in tracked schools should lag behind their counterparts in untracked schools. Using an identification strategy that takes advantage of this intuition, I find peer effects to be quite small, though generally positive and statistically significant. The preferred estimates suggest that a one standard deviation increase in initial peer ability raises math and reading achievement by no more than .024 standard deviations. Alternatively, moving a student from the 10th to the 90th percentile class in my sample yields about two weeks of additional learning. The findings are robust to a number of potential concerns.

JEL Classification: I21 and J24

Key Words: educational economics, human capital, peer effects

1. Introduction

People have long believed that the quality of a student's schoolmates is an important determinant of academic performance (and by extension, other life outcomes¹). In his influential study, Coleman (1966) asserted that peer quality was one of the only factors that could influence student outcomes besides family background. It would appear that families agree, given the substantial amounts that they are willing to pay for peer quality (see Black, 1999). Furthermore, economists have spent considerable energy developing theoretical models that take as given the importance of peer ability.² The belief that peer quality matters is so strong that programs that change the allocation of students across schools and classes tend to be extremely controversial. These programs include vouchers, school choice, and tracking.

In order to predict and evaluate the effects of these programs, it is necessary to understand how student achievement is affected by peer ability. Early examinations of peer effects implicitly assumed random assignment of students to schools and classes (after conditioning on observed covariates).³ Unfortunately, identification is complicated by the fact that students are *not* allocated randomly to schools and classes. Thus peer quality may proxy for unobserved student and family characteristics. Furthermore, peer characteristics may reflect the environment to which a student is exposed (see Manski,

¹ Murnane, Willett, and Levy (1995) present evidence that achievement on standardized tests is correlated to subsequent earnings.

² De Bartolome (1990) and Epple and Romano (1998) discuss theoretical models of school choice with vouchers and peer effects. Epple, Newlon, and Romano (2000) incorporate peer effects into a model of tracking and the competition between public and private schools. Epple and Romano (2000) provide an analysis of school choice incorporating peer effects. Rothschild and White (1995) present a model that incorporates peer effects into educational production and pricing. Benabou (1996) presents a growth model based on peer effects that operate at both a local and national level.

³ Such works include Henderson et al. (1978), Summers and Wolfe (1977), and Zimmer and Toma (2000).

1993). For these reasons, non-experimental estimates of peer effects are generally thought to be biased upwards.

Recently, a number of researchers have adopted a variety of creative methodologies to overcome these obstacles. Some of the best examples include Boozer and Cacciola (2001), Cullen, Jacob, and Levitt (2000 and 2003), Angrist and Lang (2002), Hanushek, Kain, Markman, and Rivkin (2001), and Hoxby (2000).⁴ Even among these studies, however, the results vary greatly depending on the specific methodology used. In particular, Boozer and Cacciola (2001) use variation in peer quality caused by differing class sizes to estimate classroom peer effects that are quite large. Using differences in school quality induced by residential location and magnet school lotteries, Cullen et al. (2000 and 2003) find no academic benefit associated with attending a school with better peers. Angrist and Lang (2002) find that exogenous changes in classroom composition have at most transitory effects on the achievement of minority students. Hanushek et al. (2001) and Hoxby (2000) take advantage of transitory fluctuations in school composition to estimate moderate to large peer effect estimates.

The current study takes advantage of variation in how schools allocate students to classes to identify the importance of class-level peer effects. More specifically, in tracked schools, students are placed into classes on the basis of initial ability. In untracked schools, classes contain children of varying ability levels. Thus high ability students in tracked schools will have better peers than their counterparts in untracked schools. If peer effects are important, students with high initial ability in tracked schools

⁴ In addition to the work on peer effects, there exists a related literature on the effects of track placement on student achievement. Recent work includes Hoffer (1992), Argys et al. (1996), Betts and Schkolnik (2000), and Figlio and Page (2000). Slavin (1987, 1990) and Kulik and Kulik (1982) provide meta-

should outperform similar students in untracked schools. Conversely, students with low initial ability in tracked schools should lag behind their counterparts in untracked schools. Using an instrumental variables strategy that incorporates this intuition and data on over 150,000 Chicago third and sixth graders, I find that that a one standard deviation increase in initial peer ability raises math and reading achievement by no more than .024 standard deviations. Alternatively, moving a student from the 10th to the 90th percentile class in my sample yields about two weeks of additional learning. These results are not driven by observable differences between tracked and untracked schools. The findings are also robust to a number of other potential concerns.

This identification strategy has two primary advantages relative to the prior efforts. First, the source of variation used to identify the peer effects is well within the control of policymakers. Thus the local average treatment effect is likely to be particularly relevant for understanding how the allocation of students across classes affects academic achievement. Second, the approach allows one to take advantage of the abundant naturally occurring variation in school tracking policies and initial student achievement—this permits me to identify even relatively modest peer effects.

I will continue the analysis by outlining my identification strategy in greater depth. I next describe the data source and the empirical implementation of my identification strategy. I will follow by discussing the estimation results and robustness checks. I then conclude.

2. Identification Strategy

analyses of older work. Unfortunately, there is little consensus regarding the effect of track placement on student performance.

The identification of peer effects presents formidable challenges to researchers. The primary obstacle arises from the fact that individuals are not randomly allocated to either schools or classes. Thus a student's peers are likely to be correlated to the child's own unobserved ability, willingness to work, and parental ambition and resources. Furthermore, using measures of peer ability that are contemporaneous to a student's own outcome measure leads to another source bias from "common shocks". In such a case, apparent peer effects may reflect nothing more than the fact that both the student and her peers face the same set of circumstances. These problems are discussed in depth by Manski (1993).

The interpretation of even well identified peer effect estimates is complicated by the multiple mechanisms through which peer effects could operate. First, one's peer group might have a significant effect on the content of the curriculum and the efficiency of instruction. Second, social interaction might lead to situations in which a student can benefit from being in a class with a particular peer group (e.g. low performing peers may provide a bad example or simply be more disruptive on average). Third, a student's peer group might affect the allocation of educational resources the student receives (e.g. the best teachers may be assigned to the classes with the best teachers). The ignorance regarding the mechanisms through which peer effects operate complicates the choice of functional form. For example, a single disruptive student could reduce the learning of an entire class. Alternatively, a teacher may target instruction for the bottom (or top) 30th percentile of the class ability distribution.

These functional form issues are part and parcel of almost any examination of peer effects. I focus on peer effects as they operate through mean classroom standardized

test performance. I do this in part because my identification strategy requires that I summarize student ability using a one-dimensional index. Given that the literature frequently focuses on mean peer achievement, it seems reasonable to use it as well. Additionally, mean peer ability is likely to be highly correlated to many alternative measures of classmate quality (e.g. the quantiles of the peer ability distribution). It is worth mentioning if homogeneous classrooms increase the efficiency of instruction, the variance of peer ability could have important effects on academic performance. While I will provide some evidence that my findings are unlikely to be biased due the effects of tracking on classroom heterogeneity, I will not make this a focus of the paper.

The identification strategy I employ will address the non-random placement of students across schools with the use of school*year fixed effects. I will address the non-random placement of students across classrooms using variation in peer ability that is generated by the interaction of *observed* initial achievement and the school's tracking policy. A simple example effectively outlines the intuition behind this strategy. Suppose that there are two types of students: high ability and low ability. Similarly, there are two types of schools: tracked and untracked. In untracked schools, both high and low ability students have similar peers. In tracked schools, students are assigned to classrooms on the basis of academic achievement. Thus high ability students have high ability peers and vice versa. Under these circumstances, peer ability is an increasing function of the *interaction* between a student's own achievement and the school's tracking policy. I use this interaction as an instrument for peer ability.

I will execute this identification strategy using two-stage least squares (2SLS). Naturally, the empirical implementation differs in a number of respects from this simple

example. In particular, initial ability is continuous and I control for a number of student-level covariates. I also include school*year fixed effects which subsume the average effect of tracking on student achievement along with all other factors that are constant across students within a given school and year. The second stage is given by the following learning equation:

$$(1) \quad reading_{isct} = X_{isct} \Gamma + \beta_1 reading_{isct-1} + \beta_2 math_{isct-1} + \beta_3 avg_ability_{isct-1} + FE_s + \varepsilon_{isct},$$

where *reading* and *math* are test performance of student *i* in school-year *s* (e.g. Ray School 1995), class *c*, and period *t*. *X* is a vector of observable student characteristics, *avg_ability* is the average ability of student *i*'s peers (as measured by the prior year's test scores), *FE_s* is a set of school*year fixed effects, and ε is an error term. β_3 represents the effect of average classmate ability on student performance. I will also estimate the corresponding equation for math performance.

The first stage is given by:

$$(2) \quad avg_ability_{isct-1} = X_{isct} P + \pi_1 reading_{isct-1} * tracked_s + \pi_2 math_{isct-1} * tracked_s + \pi_3 reading_{isct-1} + \pi_4 math_{isct-1} + FE_s + \xi_{isct}$$

where the variables are the same as described above except for *tracked* which in my analysis is a continuous variable that indicates the degree to which classes are segregated on the basis of initial ability. Recall that the direct effect of tracking is subsumed in the school*year fixed effects. Note that the variation in peer ability is generated by the interaction of *observed* initial achievement with the tracking status of the school. I do not use any variation in peer ability that is driven by unobserved ability, motivation, or parental influence. Because I control for the direct effect of initial achievement in the learning equation, in order for my instruments to be valid, it must be that the interaction

of observed initial ability and tracking status affects student achievement only through the allocation of students to peer groups. In a previous version of the paper I present formal conditions under which this identification strategy yields unbiased peer effect estimates (see Lefgren, 2001).

3. Data

To implement my estimation strategy I use administrative data from the Chicago Public Schools (ChiPS). These data are particularly well suited for my analysis. The large majority of students were tested and the system kept records of student performance as long as they remained in the district. I am able to track students over time using a unique identifier. Since elementary students within the ChiPS stay together throughout the day—even if a single teacher doesn't teach all subjects,⁵ it is possible to identify a student's classmates. Because the ChiPS collect data on all students in the system, accurate measures of peer ability can be constructed. The amount of data allows for precise estimation of peer effects.

Performance measures include reading and math scores on the Iowa Test of Basic Skills (ITBS) for every year that a student was in the school system and took the test. I use the grade equivalent (GE) metric for my analysis. It is calibrated based on the median student in the nation. The number to the left of the decimal point indicates grade and the number to the right of the decimal point indicates month in grade. Thus the median student in the nation in the eighth month of third grade would have a score of 3.8.

⁵ Special education and bilingual students may be sent to another class for part of the day to receive supplemental instruction.

Average peer quality measures are constructed with second or fifth grade ITBS performance depending on the specification. For the purpose of constructing peer ability measures, I use the average of each student's ITBS reading and math performance. All measures of peer ability exclude the performance of the reference student.

These administrative data also contain information on a student's gender, race, caretaker, and neighborhood characteristics. Room assignment is used to determine a student's classmates. Included in the sample are individuals who were in the third or sixth grades between 1993 and 1996. When examining students in the third grade cohorts, I use second and first grade reading and math controls. For sixth grade students, I use fifth and fourth grade ability measures. For students missing first or fourth grade reading scores, I include a dummy variable that is 1 if the score is missing. An analogous variable exists for missing math scores. I also control for race, gender, race-gender interactions, age, special education status, non-parent caretaker, foster parent, and the concentration of poverty and average social status in a student's census block.⁶

I drop all students who are missing either second or fifth grade test scores.⁷ I also exclude classes from my analysis when more than one third of students are classified as special education, more than one third of students are missing reading or math scores, or class size is less than ten students.⁸ It is likely that children in small classes with a large

⁶ Average social status is constructed using block-level census data on the mean level of education as well as the percent of employed individuals who are managers, executives, and professionals. The concentration of poverty measure is also computed using block-level census data on percent of individuals over 18 who are employed and the percent of individuals over the poverty line (the measure is reverse coded so a higher number indicates more poverty).

⁷ Not surprisingly, these students tend to be lower performing in both math and reading (about .25 to .5 years behind as measured by third or sixth grade test performance) and more mobile. For third grade, students with missing initial test scores are a somewhat less likely to be in schools that engage in tracking.

⁸ These exclusions affect about 10 percent of third grade students and about 9 percent of sixth grade students. As would be expected, the children excluded from the sample are more likely to be Hispanic or special education students. Excluded children with initial test scores perform worse in both reading and math than included children. The differences are particularly stark for the sixth grade cohort.

number of special education students may not be comparable to the broader student body and receive a very different treatment than other low achieving students. Similarly, classes in which a sizable fraction of test scores are missing frequently contain large numbers of limited English proficiency students who again may not be comparable to other students in my sample. Overall, the sample consists of 80,003 third grade and 94,230 sixth grade students.

Table 1 reports summary statistics for the students in my sample. It is apparent that they are much more likely to be minority students than would be the case nationally. Well over 80 percent of students are either black or Hispanic.⁹ Given that the national average performance for third and sixth grade students in the eighth month of the school year is 3.8 and 6.8 respectively, it is evident that the ChiPS students are somewhat low achieving. Between 7 and 9 percent of children with valid exam scores are classified as special education. Over 6 percent of children live with a non-parent relative and more than 3 percent live with foster parents.

In order to implement my identification strategy, it is necessary to know the extent to which students are sorted to classrooms on the basis of initial ability. The administrative data do not provide explicit information regarding the tracking policies of each school. Furthermore, school-reported measures of tracking tend to differ across respondents (see Figlio and Page, 2000) and informal tracking often occurs even when no formal measure is in place (see Betts and Schkolnik, 2000). I therefore construct a tracking proxy using test scores for the children in each school. My proxy describes the correlation between a student's own initial ability and the ability of her classmates. I

construct this measure by regressing initial ITBS scores (the average of math and reading performance) within a school in a given year on a set of class dummy variables.¹⁰ The r-squared from this regression is my tracking proxy. The higher the r-squared value, the closer on average a student's ability will be to his classmate's ability. If the r-squared is zero, assignment is orthogonal to initial ability. An r-squared of one would indicate that each student's initial ability equals her classmate's ability.¹¹ It is worth noting that there is a fair amount of persistence in each school's r-squared tracking measure over time. The year to year correlation is .59 for third grade and .71 for sixth grade. This suggests that my tracking measure is reflects policy decisions that persist over time—not simply the random variation in the quality of different classes.

Figure 1 shows the empirical distribution of r-squared values across schools for third grade students (the picture is similar for sixth graders). Apparently many schools engage in virtually no tracking. Figure 2 represents a school with an r-squared value of close to zero. The allocation of students across the two classes appears random. The school represented in Figure 3 has an r-squared value of .5. There is very little overlap in the distribution of initial test scores suggesting that class assignment is based almost entirely on initial ability as measured by second grade ITBS performance.

4. Baseline Results

⁹ The sharp increase in the fraction of Hispanic students between the third and sixth grades is due to the fact that children with limited English proficiency are not required to be tested. Many third grade students who have limited English proficiency progress sufficiently to be tested by the sixth grade.

¹⁰ I assign schools with just one classroom an r-squared value of zero.

¹¹ One might have a number of concerns regarding the use of this proxy. First, I am using the same data to measure tracking and control for performance—potentially inducing a mechanical relationship that biases my results. Second, if students stay together over time, the r-squared may reflect the history of past treatments or tracking policies. I show later that my results are robust to alternative tracking measures that address these concerns.

Having discussed my identification strategy and data source, it is now possible to examine the empirical findings. Tables 2a and 2b contain OLS estimates of peer effects using the specification described in equation (2). Columns 1-5 show how these estimates change with the inclusion of additional control variables. For both grades and subjects, the estimated effect of average classmate performance drops as covariates are added. Including initial ability measures and demographic characteristics reduces the estimated effect by up to 80 percent. Including another year of performance controls and a more flexible functional form further reduces estimated peer effects by 20 to 30 percent. If observables behave in the same way as unobservables, it seems likely that even the estimates of column 5 overstate the importance of peer effects. Taken at face value, the estimates suggest that increasing average peer ability by 1 grade equivalent (GE) raises the third grade reading gain by .13 GE and sixth grade reading performance by .11 GE.¹² The corresponding effect sizes for math are .17 GE for third grade and .07 GE for sixth. These estimates suggest that moving from the 10th percentile class to the 90th percentile class would increase achievement from .15 to .25 GE's.¹³ These OLS estimates suggest that a one standard deviation increase in peer ability raises reading performance by .05 to .06 standard deviations and math performance by .04 to .09 standard deviations. These OLS effect sizes are already smaller than those found by many previous researchers.

In order to overcome selection bias, I pursue the IV strategy discussed earlier. Table 3 shows results from the first stage estimation of equation (3). The coefficients indicate that students with high ability are more likely to have a high ability peer group when there is extensive tracking (as measured by the reading and math r-squared values).

¹² Recall that one grade equivalent is approximately equal to a year's worth of learning.

¹³ This represents about 1.5 to 2.5 months of learning.

My instruments act as expected and are highly significant with F-statistics that exceed 1,000 in all specifications.¹⁴

The second stage estimates are found in Table 4. The estimates suggest peer effects that are much smaller than those found using OLS with school*year fixed effects (see specifications 2 and 4). The point estimate suggests that raising classmate ability by 1 GE increases sixth grade reading performance by .03 GE. The effects for math are similar at .04 for third grade and .03 for sixth grade. The standardized effect sizes are all less than .024. To put this in perspective, going from the 10th percentile class of my sample to the 90th percentile class would increase the sixth grade reading gain by about .05 GE.¹⁵ The same move would have a similar effect for third and sixth grade math. These effects are about one third to one half the magnitude of the OLS estimates and represent about two weeks of learning.

5. Robustness Checks

These estimates are smaller than those found by a number of other researchers. It is worth discussing the factors that might bias my peer effects estimates. In this section, I will discuss potentially confounding factors and where possible, show that the concerns are unfounded.

5.1 Nonrandom Assignment of Students to Schools

¹⁴ This is not surprising given that my tracking measures are constructed using the actual class assignment of students. I show later that the results are robust to tracking measures constructed using the distribution of students across classes in non-contemporaneous years.

¹⁵ Going from the worst class in the average tracked school to the best class in the same school would yield a similar performance gain.

The first potential source of bias comes from the non-random assignment of students to schools. In particular, parents may choose a school based on its tracking policy. School-year fixed effects capture the mean observed and unobserved differences in the student body across schools. Thus if students in tracked schools come disproportionately from families who care deeply about education, my strategy will still yield unbiased estimates. A more subtle type of selection could be problematic, however. Motivated parents might want their children to be in classes with the best peers possible. These parents would want to send their high ability children to tracked schools and their low ability children to untracked schools. Under these circumstances, the correlation between initial *measured* ability and unobserved student motivation would be higher in tracked schools than in untracked schools (i.e. in tracked schools the high ability students have relatively good unobservables and the low ability students have relatively poor unobservables). The interaction between ability and tracking status would reflect this higher correlation and thus be an invalid instrument. It is possible to show that my estimates would be biased upwards in this case.

To examine whether school attendance decisions were based upon the quality of peers *within* the classroom, it is helpful to understand the effect of mean peer ability on the probability a student attends a different school the following year. To do so, I again instrumented mean peer achievement using the interaction of initial ability and schooling tracking measures. The sample and choice of covariates was identical to that used for the baseline estimates of peer effects. This procedure yields estimates that suggest that third graders in classes with better peers are somewhat less likely to change schools. In particular, a 1 G.E. increase in classmate achievement reduces the probability of

changing schools by 4.2 percentage points (compared to a baseline mobility rate of 21.6 percent for third graders in my sample). The corresponding t-statistic is 2.47. For sixth graders a 1 G.E. increase in classmate achievement raises the probability of changing schools by only .2 percentage points; the t-statistic is .18.

These estimates suggest that for third graders selection on the basis of classmate ability is a possible source of bias. Examining Table 3, it seems likely that the bias is minimal, however. In particular, my estimates appear to be robust to the inclusion of detailed student-level achievement and demographic covariates. In specifications 2 and 4, we include variables such as neighborhood characteristics and free lunch status, which are likely to be highly correlated to a family's willingness and ability to change schools on the basis of classmate ability. If my estimates severely biased due to unobservables (e.g. parent motivation), the inclusion of these covariates should reduce the bias and yield different (smaller in this case) estimates. Given that this doesn't occur, it would appear that my estimates are largely unaffected by this type of bias.

5.2 Cumulated Treatments

Another concern is that performance measures in tracked schools may reflect both the underlying ability of the student as well as the cumulated treatment effects of being in either the high track or the low track. Thus a high ability student in a tracked school might actually have worse unobservable characteristics than a student in an untracked school whose score doesn't reflect years of being in the high track. Similarly, a low performing child in an untracked school may possess worse unobservable characteristics than a low performing child in a tracked school whose performance is partially

attributable to time spent with low quality classmates. The fact that initial achievement scores may be driven by the past tracking status of the school could be problematic for two reasons.

First, the correlation between observed initial performance and unobserved characteristics may be affected by the tracking status of the school. In this case my instrument, the interaction of initial ability and tracking status, would predict unobserved ability and bias my estimated peer effects (downwards in this case). This problem should be less severe if instrument using ability measures from the more distant past. In row 1, I use only those instruments generated by the interaction of second (fifth) grade achievement and tracking status (my baseline r-squared measure). In row 2, I use instruments generated by the interaction of first (fourth) grade achievement and tracking status. These strategies yield very similar estimates—suggesting that cumulated treatments may not be an important source of bias.

A second related concern is that since my measure of tracking is based on initial ability, it may be reflect the amount of tracking in earlier grades. It is not obvious whether or how this might bias my estimated peer effects. Nevertheless, it is important to examine the sensitivity of my estimates to the choice tracking measure. In row 3 of Table 5, I show specifications in which the r-squared tracking measure is calculated using first (fourth) grade achievement measures. The instruments are two years lagged reading and math achievement interacted with this alternative r-squared measure. Again, the findings appear robust to this alternative tracking measure.

5.3 Empirical Tracking Measure

In addition to the concerns just raised, one may be worried that using a tracking measure generated by the same data used for analysis could generate a mechanical bias. To assuage these concerns, I can take advantage of data within the same school but from different years. In row 4, I use as my tracking measure, the average r-squared calculated from non-contemporaneous years. In other words, for the students in the 1995 cohort, my measure of tracking is the average of the school's 1993, 1994, and 1996 r-squared values. This procedure yields a tracking measure that is not mechanically related to the initial achievement of the students in the school during the year. Once again I find estimates that are substantively similar to my baseline coefficients.

5.5 Sensitivity to Sample Exclusions

In my choice of samples, I excluded classes with very few students, a large fraction of special education students, and with a large number of students with missing test scores. I did so to make sure that my results were not driven by a small number of atypical students—special education and limited English proficiency students in particular. This restriction on my sample might be problematic if schools that I classify as untracked are more likely to place low ability students in classes specifically designed for special education students or exclude the students from testing. In particular, the exclusion of such students from my sample could induce a sample selection bias.¹⁶

Examining the issue empirically, I find that for third graders, students from schools with

¹⁶ Recall that students who are classified as special education, have valid test scores, and are in regular classes are retained in the baseline sample. Thus my analysis should be unaffected by the number of such students in tracked and untracked schools, unless these students are pulled out of class for so much time as to substantively change the typical quality of a student's classmates. Unfortunately, I have no way to check whether tracked and untracked schools have different policies for pulling special education students out of class. If tracked and untracked school pull out special education students with a similar frequency, this

high r-squared measures are less likely to be in classrooms that are ultimately excluded from the analysis or have missing outcome measures—holding constant observable student and school characteristics.¹⁷ There is no discernible effect for sixth graders.

To address this issue, I examine a specification in which I include all students with valid test scores—even those in classes specifically designed for special needs students. My r-squared tracking measure in this specification is also calculated using the full sample of students with initial test scores. Row 5 of Table 5 contains the estimates from this specification. The third grade estimates are virtually identical to those I found in my baseline specification, suggesting that my estimates are not substantially affected by my sample restriction. For sixth grade students, the inclusion of these children increases the estimated peer effects substantially though they remain below the OLS estimates. This suggests either that peer effects operate strongly for these classes or that the students in these classes are fundamentally different from other students in my sample.

5.6 Differences between Tracked and Untracked Schools

Another potential criticism of my approach is that tracking policy is treated as exogenous though this is unlikely to be the case. In particular, tracking policy is likely to depend on the costs and benefits of tracking as well as administrator preferences. These factors, along with endogenous school choice on the part of students, are likely to lead to a situation in which tracked and untracked schools differ over both observable and

issue is unlikely to be problematic as the process for classifying children (in normal classes) as special education appears to be substantively similar in tracked and untracked schools.

¹⁷ Increasing the r-squared measure from 0 to .5 reduces the probability that a student is excluded from the sample by about 4 percentage points—the t-statistic is 4.1.

unobservable dimensions. Conceivably factors correlated to tracking status could differentially affect the achievement of high and low ability students. For example, tracked schools tend to be larger than untracked schools. If low ability students did poorly in large schools, it would appear that tracking increased the relative achievement of high ability students—falsely suggesting the presence of positive peer effects.

Before checking to see if my results are driven by differences between tracked and untracked schools, it is worth documenting that differences actually exist. In order to do so, I divide schools into thirds on the basis of my r-squared tracking measure. I then compute mean school level characteristics for each third. Table 5 shows these means.

There is very little difference in tracking between the bottom third and middle third of the tracking distribution. The bottom third is composed heavily of schools that have only one classroom included in the sample¹⁸—these schools have a tracking measure of zero by definition. The middle third is composed of schools with multiple classes that have r-squared values of just above zero, indicating a lack of substantial tracking. The top third is composed of schools with relatively high r-squared measures. Perhaps unsurprisingly, school size¹⁹ is increasing with the degree of tracking. Highly tracked schools also tend to have lower mean performance, higher ability variance, lower social status measures, and a larger fraction of black students.²⁰

5.7 Identifying Peer Effects Using Matched Schools

¹⁸ Only a small fraction of schools actually have only one classroom per grade (8 percent of schools for third grade and 6 percent for sixth grade). A somewhat larger fraction has multiple classrooms but only one included classroom (15 percent for third grade and 11 percent for sixth grade). Many of the excluded classes contain large numbers of bilingual students that weren't included for testing. I will later check to see how sensitive my results are to the exclusion of these students.

¹⁹ School size is defined as the number of third (or sixth) grade students that are in included classes.

Given that tracked and untracked schools differ over a number of dimensions, the apparent peer effects may be driven by factors that are correlated to tracking. It is worth reiterating that the school fixed effects account for school characteristics that affect high and low ability students in similar ways. Problems could only arise if school characteristics correlated to tracking affected the performance of high and low ability students differently. In order to address this concern, I estimate peer effects using schools that are matched on the basis of observable characteristics. This matching takes advantage of variation in school tracking policies that is attributable to administrator preferences or other unobserved factors. The strategy will still yield biased estimates if administrator preferences or unobserved factors have a differential effect on high and low ability students.

If my estimated peer effects simply reflect observable school-level differences, matching schools should yield estimates that are substantially different from the baseline estimates. These robustness checks also allow me to investigate whether peer effects operate more strongly in some schools than in others.

I divide schools into thirds on the basis of school size, initial ability variance, mean initial ability, average concentration of poverty, and fraction black. Though only the first two factors clearly reflect potential costs or benefits of tracking,²¹ the other factors are also correlated with tracking policy. I will estimate peer effects separately for

²⁰ These differences are not solely a function of which classes were excluded from the sample. The differences are similar even without the exclusions.

²¹ In large schools, administrators have more teachers per grade and thus have greater capacity to create remedial or accelerated classes. Thus school size is mechanically related to the cost of increasing classroom homogeneity. If the objective of tracking is to create homogeneous classroom environments, clearly the benefits of tracking are higher when the variance in student ability is also large.

schools in the bottom, middle, and upper thirds of the school size distribution. I will do likewise for the other school-level characteristics.²²

Table 6 reports the estimated effect of mean classmate ability on reading and math achievement for students in matched schools. This matching yields a pattern of results similar to my baseline estimates. In almost all of the specifications the point estimates are substantially lower than the OLS estimates and not significantly different than the baseline estimates. The coefficients generally have the same sign as the baseline estimates as well. It is tempting to divine a pattern in the differences. In third grade, peer effects appear to operate more strongly in schools with high mean ability. Closer inspection, however, reveals that the pattern isn't always monotonic and differs for the sixth grade. This lack of consistency across grades and subjects makes it difficult to draw inferences from some of the particularly large coefficients. Overall, the evidence from this section suggests that there is little reason to worry that the small positive peer effects I found are due to observable differences between tracked and untracked schools.

5.8 Differential Impacts of the Variance of Peer Ability

The variance of peer ability within the classroom could be an important determinant of the efficiency of instruction. Tracking might benefit students by reducing classroom heterogeneity. Any such benefits that accrue uniformly to all students within a school will be subsumed in the school fixed effects. In such a case, my estimate of the effect of mean peer ability would be unbiased, though the analysis would fail to address an interesting dimension of peer effects. If, however, high and low ability students

²² I also matched schools on the basis of propensity scores that were calculated using all of the school level variables. These results, though not shown, are broadly consistent with my baseline estimates.

benefit differentially from reductions in classroom variance, the estimated effect of mean peer ability will be biased. This is because I would attribute the relative difference in achievement between high and low ability students in tracked schools exclusively to mean peer ability—ignoring the differential importance of classroom heterogeneity to high and low ability students.

I would ideally have convincing instruments both for mean peer ability as well as for the variance of peer ability to address this concern. Unfortunately, this is not the case. I can, however, provide suggestive evidence regarding whether reductions in classroom heterogeneity affect the average student. In particular, I can regress student achievement on my tracking proxy directly, controlling for individual and school-level covariates. To the extent that students in tracked schools perform no better than students in untracked schools, *ceteris paribus*, it seems plausible that reductions classroom heterogeneity are not an important determinant of student achievement.²³ This would ease concerns that the estimated effects of mean peer ability were biased.²⁴

Table 8 shows results from a regression of student achievement on my r-squared measure of school tracking. I use the baseline sample of students, though the school-level averages are computed using all students for which the relevant data are available. My student-level covariates are the same from the prior analysis. Because I do not include school fixed effects (which would subsume the tracking measure), I control for the following school-level covariates: average performance, variance of initial

²³ Examining the importance of classroom variance directly would be complicated by the non-random allocation of students to classrooms and the relationship between the variance and other dimensions of peer quality (e.g. the extreme quantiles of the distribution of peer ability).

²⁴ Of course if reductions in heterogeneity were good for high ability students and bad for low ability students (or vice versa), there might still be a significant bias even though the average effect of tracking was zero. Administrators who use tracking, however, tend to justify it on the grounds that it increases instructional efficiency for *all* students.

performance, school size, fraction black, fraction Hispanic, average family social status, average concentration of poverty, and fraction of students in special education. The standard errors are corrected for clustering within school years.

The coefficients suggest that tracking has little overall effect on third grade achievement and sixth grade reading. Tracking appears to be associated with increased sixth grade math performance, *ceteris paribus*. While the non-random allocation of students to schools makes it difficult to be certain the tracking coefficients represent causal effects (even with the controls), it appears plausible that the effect of tracking on classroom heterogeneity are not biasing my findings for three of four specifications. It is worth recalling that even if reductions in classroom heterogeneity increase performance on average for sixth grade math students, the estimated effect of mean peer ability is biased only if classroom variance differentially affects high and low ability students.

Overall, it does not appear that my peer effects are driven by observable differences between tracked and untracked schools. My results are also generally robust to my choice of sample and instruments. Controlling for more covariates has no significant effect on my estimates. Additionally, there is suggestive evidence that reductions in classroom variance are not biasing my results for third grade reading and math as well as sixth grade reading. Tracking does appear to increase average sixth grade math achievement, hinting that reductions in classroom heterogeneity may have some benefits. Unless these benefits accrue differentially to high and low ability students, the effect of mean peer ability is still identified.

7. Conclusions

This paper is important in that it presents evidence that apparent class-level peer effects exist but are fairly modest. The IV estimates imply that going from the 10th percentile class to the 90th percentile class would increase achievement gains by between -.03 and .05 grade equivalents—less than the gain that occurs in a typical month. These effect sizes suggest that policies leading to moderate changes in the allocation of students across classes will have only a limited effect on the academic achievement of students.²⁵ The results stand in contrast to those found by many other researchers. In particular, my peer effect estimates are smaller than those found by Henderson et al. (1978), Zimmer and Toma (2000), Hoxby (2000), Boozer and Cacciola (2001), and Hanushek et al. (2001). My results are consistent with Cullen et al. (2001 and 2003) and Angrist and Lang (2002) who find a small or no effect of peer ability on academic achievement.

My preferred results do not appear to be driven by observable differences between tracked and untracked schools that are unrelated to class placement. The results are also robust to my measure of tracking, choice of sample, and choice of instruments. It seems likely that reductions in classroom heterogeneity associated with tracking do not bias the estimates.

It is important to be clear regarding the nature of peer effects identified in this study. They capture the importance of *within* classroom peer ability on math and reading achievement. They do not reflect possibly important effects operating at the school or neighborhood level. These classroom-level peer effects may operate through teacher quality, rigor of the curriculum, or direct interaction with classmates. As with all

²⁵ Understanding whether high or low ability students benefited differentially from peer ability could be important for evaluating the efficiency and distributional consequences of some policy interventions. Unfortunately, my identification strategy does not allow me to estimate differential effects that are sufficiently precise to draw meaningful inference—see Lefgren, 2001.

instrumental variables strategies, my estimate reflects a particular local average treatment effect (LATE).²⁶ In particular, the estimates reflect the impact of peer ability where one's peers are determined by the interaction of the student's own ability and the school's tracking policy.²⁷ My estimate does not necessarily capture the impact of a changing demographic or assignment to a different school. Fortunately, the LATE I identify is exactly the one that is most relevant for principals or other policy makers considering changing the allocation of students across classes within a school.

The relatively modest peer effects found in this and other studies will seem implausible to a number of readers. Many people (parents in particular) believe quite strongly that children will be better off if they attend class or school with more able peers. Researchers have taken as given the existence of such peer effects in the development of theoretical models of vouchers, school choice, and residential segregation. Even if the effect of average classmate ability on basic skills is limited, peoples' belief in the importance of peer effects may not be entirely misplaced for at least two reasons. First, peer effects may be more important at the school or neighborhood level than in the classroom settings. In particular, social interactions may not be limited to the classroom and decisions regarding pedagogy may be heavily influenced by the overall capability of the students within the *school*. Second, classroom peer effects may manifest themselves in ways not captured by basic skills achievement. For example the breadth of curriculum, college aspirations, self-image, and the amount of antisocial

²⁶ See Imbens and Angrist (1994).

²⁷ Recall that my instrument is the interaction of my tracking measure and initial ability. This instrument will be close to zero for all students in schools that engage in little tracking but will vary substantially in schools with a high tracking measure. Thus, while most schools engage in very little tracking, students in highly tracked schools induce the most variation in the instrument. The LATE disproportionately reflects the experience of such students.

behavior could all be heavily influenced by classmate ability and characteristics.²⁸ Peer effects operating outside of the classroom or in a manner unrelated to basic skills achievement will not be captured in my estimates.

Though I identify peer effects by examining the *relative* effects of tracking on high and low ability students, I present only suggestive evidence on whether high and low ability students benefit (or suffer) in an absolute sense from tracking.²⁹ Preliminary findings imply that tracking may on average raise the math performance of sixth grade students—perhaps due to reductions in classroom heterogeneity that increase instructional efficiency. For third graders and sixth grade reading students, tracking appears to have little effect on average achievement.

²⁸ Cullen et al. (2003) provide evidence that school quality positively affects non-cognitive outcomes.

²⁹ See Figlio and Page (2000) for a notable attempt to identify the absolute effect of tracking on the performance of students at different points of the initial ability distribution. Using an IV strategy, Figlio and Page find that tracking dramatically *increases* the performance of low ability students but has no significant effect on the performance of high ability students. The significance of the effect depends on the measure of tracking, however. Additionally, the OLS results show no effect of tracking on students at any point in the ability distribution. Unfortunately, the authors' data don't allow them to identify a student's classmates and thus relate the estimated effects of tracking to peer ability.

References

- Angrist, J. D. and K. Lang. "How Important Are Classroom Peer Effects? Evidence from Boston's METCO Program." Boston University Working Paper 2002.
- Benabou, R. "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance." *American Economic Review* 1996 vol. 86, pp. 584-609.
- Betts, J. R. and J. L. Shkolnik. "The Effects of Ability Grouping on Student Math Achievement and Resource Allocation in Secondary Schools." *Economics of Education Review* 2000 vol. 19, pp. 1-15.
- Black, S. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics* 1999 vol. 114, pp. 577-599.
- Boozer, M. and S. Cacciola. "Inside the Black Box of Project Star: Estimation of Peer Effects Using Experimental Data." *Center Discussion Paper* 832 2001.
- Coleman, J. S., E. Q. Campbell, C. J. Hobson, J. McPartlan, A. M. Mood, F. D. Weinfeld and R. L. York, 1966, *Equality of Educational Opportunity*. Government Printing Office, Washington, D. C. 1966.
- Cullen, J. B., B. A. Jacob, and S. D. Levitt. "The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools." *NBER Working Paper* 7888 2000.
- Cullen, J. B., B. A. Jacob, and S. D. Levitt. "The Effect of Magnet Schools on Student Outcomes: Evidence from Randomized Lotteries" Working Paper 2003.
- De Bartolome, C. A. M. "Equilibrium and Inefficiency in a Community Model with Peer Group Effects." *Journal of Political Economy* 1990 vol. 98, pp. 110-133.

- Epple, D., E. Newlon, and R. Romano. "Ability Tracking, School Competition, and the Distribution of Educational Benefits." *NBER Working Paper 7854* 2000.
- Epple, D. and R. E. Romano. "Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review* 1998 vol. 88, pp. 33-62.
- Epple, E. and R. E. Romano. "Neighborhood Schools, Choice, and the Distribution of Educational Benefits." *NBER Working Paper 7850* 2000.
- Figlio, D. N. and M. E. Page. "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Equality?" *NBER Working Paper 8055* 2000.
- Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin. "Does Peer Ability Affect Student Achievement?" *NBER Working Paper 8502* 2001.
- Henderson, V., P. Mieszkowski and Y. Sauvageau. "Peer Group Effects and Educational Production Functions." *Journal of Public Economics* 1978 vol. 10, pp. 97-106.
- Hoffer, T. B. "Middle School Ability Grouping and Student Achievement in Science and Mathematics." *Educational Evaluation and Policy Analysis* 1992 vol. 14, pp. 205-227.
- Hoxby, C. "Peer Effects in the Classroom: Learning from Gender and Race Variation." *NBER Working Paper 7867* 2000.
- Imbens, G. W. and J. D. Angrist. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 1994 vol. 62, pp. 467-475.
- Kulik, C. C. and J. A. Kulik. "Effects of Ability Grouping on Secondary School Students: A Meta-analysis of Evaluation Findings." *American Educational Research Journal* 1982 vol. 19, pp. 415-428.

- Lazear, E. P. "Education Production." Working Paper 1999.
- Lefgren, L. J. "Three Essays on Educational Policy and Peer Effects." University of Chicago Dissertation 2001.
- Manski, C. F. "Identification of Endogenous Social effects: The Reflection Problem." *Review of Economic Studies* 1993 vol. 60, pp. 531-542.
- Murnane, R., J. B. Willett, and F. Levy. "The Growing Importance of Cognitive Skills in Wage Determination." *Review of Economics and Statistics* 1995 vol. 77, pp. 251-266.
- Rothschild, M. and L. J. White. "The Analytics of the Pricing of Higher Education and Other Services in Which the Customers are the Inputs." *Journal of Political Economy* 1995 vol. 103, pp. 573-623.
- Slavin, R. E. "Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis." *Review of Educational Research* 1987 vol. 57, pp. 293-336.
- Slavin, R. E. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis." *Review of Educational Research* 1990 vol. 60, pp. 471-499.
- Summers, A. A. and B. Wolfe. "Do Schools Make a Difference?" *American Economic Review* 1977 vol. 67, pp. 639-652.
- Zimmer, R. W. and E. F. Toma. "Peer Effects in Private and Public Schools across Countries." *Journal of Policy Analysis and Management* 2000 vol. 19, pp. 75-92.

Table 1. Summary Statistics.

Variable	3rd Grade	6th Grade	Variable	3rd Grade	6th Grade
3rd Grade Math GE	3.390 (0.964)	-	Age	8.574 (0.433)	11.555 (0.478)
3rd Grade Reading GE	2.984 (1.138)	-	Male	0.496 (0.500)	0.491 (0.500)
2nd Grade Math GE	2.670 (0.786)	-	Black	0.689 (0.463)	0.589 (0.492)
2nd Grade Reading GE	2.271 (1.033)	-	Hispanic	0.178 (0.382)	0.269 (0.443)
1st Grade Math GE	1.803 (.631)	-	Special Education	0.074 (0.262)	0.086 (0.280)
1st Grade Reading GE	1.608 (.835)	-	Lives with Relative (Not Parent)	0.072 (0.259)	0.060 (0.237)
6th Grade Math GE	-	6.175 (1.314)	Lives with Foster Parent	0.042 (0.200)	0.031 (0.173)
6th Grade Reading GE	-	5.872 (1.521)	Concentration of Poverty in Census Tract	0.379 (0.757)	0.279 (0.723)
5th Grade Math GE	-	5.125 (1.132)	Average Social Status in Census Tract	-0.295 (0.689)	-0.285 (0.695)
5th Grade Reading GE	-	4.981 (1.291)	School Average ITBS	2.427 (.414)	4.946 (.568)
4th Grade Math GE	-	4.193 (1.006)	Class Average ITBS	2.462 (0.537)	5.047 (0.718)
4th Grade Reading GE	-	3.941 (1.185)	R-Squared Tracking Measure	.193 (.208)	0.191 (0.220)
			Observations	80,003	94,230

Notes for Table 1: Standard deviations are in parentheses.

Table 2a. Ordinary Least Squares Estimates of Class-Level Peer Effects for Third Graders.

Dependent Variable: 3rd Grade Reading	Specification				
	(1)	(2)	(3)	(4)	(5)
Average Peer ITBS GE	.904** (.015)	.847** (.014)	.171** (.014)	.139** (.014)	.128** (.013)
School-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes	Yes	Yes
Prior Year Math and Reading Controls	No	No	Yes	Yes	Yes
Two Years Prior Math and Reading Controls	No	No	No	Yes	Yes
Quadratic Reading and Math Controls	No	No	No	No	Yes
R-Squared	.315	.349	.604	.613	.617
Observations	79,794	79,027	79,027	79,027	79,027

Dependent Variable: 3rd Grade Math	Specification				
	(1)	(2)	(3)	(4)	(5)
Average Peer ITBS GE	.873** (.013)	.826** (.013)	.214** (.013)	.187** (.013)	.165** (.013)
School-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes	Yes	Yes
Prior Year Math and Reading Controls	No	No	Yes	Yes	Yes
Two Years Prior Math and Reading Controls	No	No	No	Yes	Yes
Quadratic Reading and Math Controls	No	No	No	No	Yes
R-Squared	.351	.386	.691	.705	.713
Observations	79,007	78,243	78,243	78,243	78,243

Notes for Table 2a: Only classes with less than one third of the students in special education or having missing test scores are included. I also exclude classes with fewer than ten students. Average peer ability is computed at the class level. Standard errors are corrected for within class clustering and are in parentheses.

* p<.10

** p<.05

Table 2b. Ordinary Least Squares Estimates of Class-Level Peer Effects for Sixth Graders.

Dependent Variable: 6th Grade Reading	Specification				
	(1)	(2)	(3)	(4)	(5)
Average Peer ITBS GE	.974** (.010)	.874** (.010)	.150** (.008)	.106** (.008)	.105** (.008)
School-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes	Yes	Yes
Prior Year Math and Reading Controls	No	No	Yes	Yes	Yes
Two Years Prior Math and Reading Controls	No	No	No	Yes	Yes
Quadratic Reading and Math Controls	No	No	No	No	Yes
R-Squared	.300	.365	.659	.685	.691
Observations	93,962	93,139	93,139	93,139	93,139

Dependent Variable: 6th Grade Math	Specification				
	(1)	(2)	(3)	(4)	(5)
Average Peer ITBS GE	.846** (.009)	.756** (.010)	.097** (.007)	.074** (.007)	.073** (.007)
School-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes	Yes	Yes
Prior Year Math and Reading Controls	No	No	Yes	Yes	Yes
Two Years Prior Math and Reading Controls	No	No	No	Yes	Yes
Quadratic Reading and Math Controls	No	No	No	No	Yes
R-Squared	.327	.392	.801	.816	.820
Observations	93,444	92,623	92,623	92,623	92,623

Notes for Table 2b: Only classes with less than one third of the students in special education or having missing test scores are included. I also exclude classes with fewer than ten students. Average peer ability is computed at the class level. Standard errors are corrected for within class clustering and are in parentheses.

* p<.10

** p<.05

Table 3. First Stage Results of Instrumental Variables Estimation of Class-Level Peer Effects.

Dependent Variable	Specification	
	Mean Peer 2 nd Grade ITBS GE	Mean Peer 5 th Grade ITBS GE
2 nd (5 th) Grade Reading GE *	.494**	.475**
R-Squared	(.012)	(.013)
1 st (4 th) Grade Reading GE *	.173**	.161**
R-Squared	(.014)	(.011)
1 st (4 th) Grade Reading Missing *	.326**	.625**
R-Squared	(.029)	(.080)
2 nd (5 th) Grade Math GE *	.392**	.352**
R-Squared	(.014)	(.015)
1 st (4 th) Grade Math GE *	.092**	.118**
R-Squared	(.017)	(.015)
1 st (4 th) Grade Math Missing *	.029	.404**
R-Squared	(.035)	(.088)
F-Statistic of Instrument	1,121	1,258
School-Year Fixed Effects	Yes	Yes
Demographic Controls	Yes	Yes
Prior Year Math and Reading Controls	Yes	Yes
Two Years Prior Math and Reading Controls	Yes	Yes
Quadratic Reading and Math Controls	Yes	Yes
Observations	79,027	93,193

Table 3 notes: The r-squared measure is computed by regressing ITBS test scores of students within a school on class fixed effects. Only classes with less than one third of the students in special education or having missing test scores are included. I also exclude classes with fewer than ten students. The grades in parentheses refer to the sixth grade specifications. Standard errors are corrected for within class clustering and are in parentheses.

* p<.10

** p<.05

Table 4. Second Stage Results of Instrumental Variables Estimation of Class-Level Peer Effects for Third and Sixth Grade Students.

Dependent Variable	Specification			
	3 rd Grade Reading		3 rd Grade Math	
	(1)	(2)	(3)	(4)
Average Peer ITBS GE	-.030 (.026)	-.024 (.025)	.046** (.021)	.041** (.020)
School-Year Fixed Effects	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	No	Yes
Prior Year Math and Reading Controls	Yes	Yes	Yes	Yes
Two Years Prior Math and Reading Controls	Yes	Yes	Yes	Yes
Quadratic Reading and Math Controls	No	Yes	No	Yes
Observations	79,027	79,027	78,243	78,243

Dependent Variable	Specification			
	6 th Grade Reading		6 th Grade Math	
	(1)	(2)	(3)	(4)
Average Peer ITBS GE	.014 (.014)	.027** (.014)	.020* (.011)	.032** (.011)
School-Year Fixed Effects	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	No	Yes
Prior Year Math and Reading Controls	Yes	Yes	Yes	Yes
Two Years Prior Math and Reading Controls	Yes	Yes	Yes	Yes
Quadratic Reading and Math Controls	No	Yes	No	Yes
Observations	93,193	93,193	92,623	92,623

Table 4 notes: Average ability measures are instrumented using own ability multiplied by a measure of within school tracking. Only classes with less than one third of the students in special education or having missing test scores are included. I also exclude classes with fewer than ten students. Standard errors are corrected for within class clustering and are in parentheses.

* p<.10

** p<.05

Table 5. Sensitivity of Instrumental Variables Estimates to Sample and Choice of Instruments.

Specification	Dependent Variable			
	3 rd Grade	3 rd Grade	6 th Grade	6 th Grade
	Reading	Math	Reading	Math
	(1)	(2)	(3)	(4)
2 nd (5 th) Grade Instruments Only	-.026 (.025)	.037* (.020)	.028** (.014)	.032** (.011)
1 st (4 th) Grade Instruments Only	-.014 (.028)	.055** (.022)	.020 (.016)	.029** (.012)
Tracking Measure Calculated Using 1 st (4 th) Grade Performance	.018 (.028)	.075** (.023)	.024* (.015)	.022** (.011)
Tracking Measure Calculated Using Non-Contemporaneous Years	-.020 (.033)	.054** (.025)	.006 (.018)	.022* (.013)
Full Sample (Including Special Ed. and Other Excluded Classes)	-.022 (.025)	.041** (.018)	.068** (.014)	.061** (.010)

Table 7 notes: Average ability measures are instrumented using own ability multiplied by a measure of within school tracking. Only classes with less than one third of the students in special education or having missing test scores are included. I also exclude classes with fewer than ten students. Standard errors are corrected for within class clustering and are in parentheses. In all cases, I control for student demographics, linear and quadratic terms of two years' prior test scores, and school-year fixed effects.

* p<.10

** p<.05

Variable	3rd Grade Tracking Policy			6th Grade Tracking Policy		
	Bottom Third	Middle Third	Top Third	Bottom Third	Middle Third	Top Third
R-Squared	.001	.053	.397	.002	.051	.414
Number of Children in Grade	36.037	63.409	71.552	43.337	67.862	74.014
Variance of Initial Ability	.507	.506	.533	.933	.922	.989
Mean Initial Ability	2.663	2.429	2.396	5.276	4.955	4.985
Mean Social Status	-.120	-.227	-.426	-.106	-.288	-.359
Mean Concentration of Poverty	.050	.340	.521	.105	.303	.404
Fraction Black	.472	.664	.744	.532	.572	.665
Fraction Hispanic	.282	.185	.166	.253	.281	.221
Number of Schools	543	543	542	566	565	565

Table 5 notes: Only classes with less than one third of the students in special education or having missing test scores are included. I also exclude classes with fewer than ten students.

Table 7. Instrumental Variables Estimates of Class-Level Peer Effects Using Matched Schools.

Specification	Dependent Variable			
	3 rd Grade Reading (1)	3 rd Grade Math (2)	6 th Grade Reading (3)	6 th Grade Math (4)
School Size				
Bottom Third	-.019 (.087)	.114** (.057)	.134** (.041)	.054* (.032)
Middle Third	-.023 (.035)	.046* (.027)	.007 (.025)	-.002 (.019)
Top Third	-.023 (.036)	.031 (.028)	.012 (.019)	.044** (.015)
<i>F-Statistic</i> (Prob>F)	0.001 (.999)	.840 (.432)	4.044** (.018)	2.089 (.124)
Variance of Initial Ability				
Bottom Third	-.084 (.054)	.026 (.045)	.066* (.035)	-.003 (.028)
Middle Third	.019 (.040)	.041 (.030)	.029 (.026)	.045** (.020)
Top Third	-.022 (.036)	.061** (.028)	.032* (.019)	.043** (.015)
<i>F-Statistic</i> (Prob>F)	1.175 (.309)	.249 (.780)	.421 (.657)	1.146 (.318)
Mean Initial Ability				
Bottom Third	-.050 (.040)	-.037 (.035)	.022 (.026)	.019 (.021)
Middle Third	-.058 (.041)	.024 (.037)	.051** (.025)	.063** (.019)
Top Third	.012 (.042)	.088** (.028)	-.009 (.020)	.005 (.016)
<i>F-Statistic</i> (Prob>F)	.830 (.436)	3.991** (.019)	1.756 (.173)	2.828* (.059)

Table 6 notes: Average ability measures are instrumented using own ability multiplied by a measure of within school tracking. Only classes with less than one third of the students in special education or having missing test scores are included. I also exclude classes with fewer than ten students. The F-statistic is a measure of whether the coefficients are statistically different from each other. Standard errors are corrected for within class clustering and are in parentheses. In all cases, I control for student demographics, linear and quadratic terms of two years' prior test scores, and school-year fixed effects.

Table 7 continued. Instrumental Variables Estimates of Class-Level Peer Effects Using Matched Schools.

Specification	Dependent Variable			
	3 rd Grade Reading (1)	3 rd Grade Math (2)	6 th Grade Reading (3)	6 th Grade Math (4)
Mean Concentration of Poverty				
Bottom Third	.063** (.031)	.086** (.028)	-.001 (.023)	.068** (.017)
Middle Third	.050 (.038)	.065** (.032)	.024 (.023)	-.001 (.019)
Top Third	.001 (.047)	.022 (.036)	.029 (.028)	.013 (.021)
<i>F-Statistic</i> (Prob>F)	.623 (.537)	.987 (.373)	.435 (.647)	4.112** (.016)
Fraction Black				
Bottom Third	-.014 (.031)	.041* (.025)	-.001 (.021)	.050** (.017)
Middle Third	.005 (.041)	.080** (.031)	.067** (.026)	.015 (.022)
Top Third	.026 (.049)	.038 (.039)	.029 (.027)	.016 (.020)
<i>F-Statistic</i> (Prob>F)	.246 (.782)	.582 (.559)	2.087 (.124)	1.207 (.299)

Table 6 notes: Average ability measures are instrumented using own ability multiplied by a measure of within school tracking. Only classes with less than one third of the students in special education or having missing test scores are included. I also exclude classes with fewer than ten students. The F-statistic is a measure of whether the coefficients are statistically different from each other. Standard errors are corrected for within class clustering and are in parentheses. In all cases, I control for student demographics, linear and quadratic terms of two years' prior test scores, and school-year fixed effects.

Table 8. The Direct Effect of Tracking on Student Achievement.

Specification	Dependent Variable			
	3 rd Grade Reading (1)	3 rd Grade Math (2)	6 th Grade Reading (3)	6 th Grade Math (4)
R-Squared Tracking Measure	0.022 (0.48)	0.047 (1.07)	0.013 (0.15)	0.173 (2.10)
Student-Level Covariates	Yes	Yes	Yes	Yes
School-Level Covariates	Yes	Yes	Yes	Yes

Table 8 notes: Only students in classes with less than one third of the students in special education or having missing test scores are included. I also exclude classes with fewer than ten students. Standard errors are corrected for school*year clustering and are in parentheses. I control for student demographics, and linear and quadratic terms of two years' prior test scores. I also control for the following school-level covariates: average performance, variance of initial performance, school size, fraction black, fraction Hispanic, average family social status, average concentration of poverty, and fraction of students in special education.

* p<.10

** p<.05

Figure 1. Histogram of R-Squared Values for the Third Grade

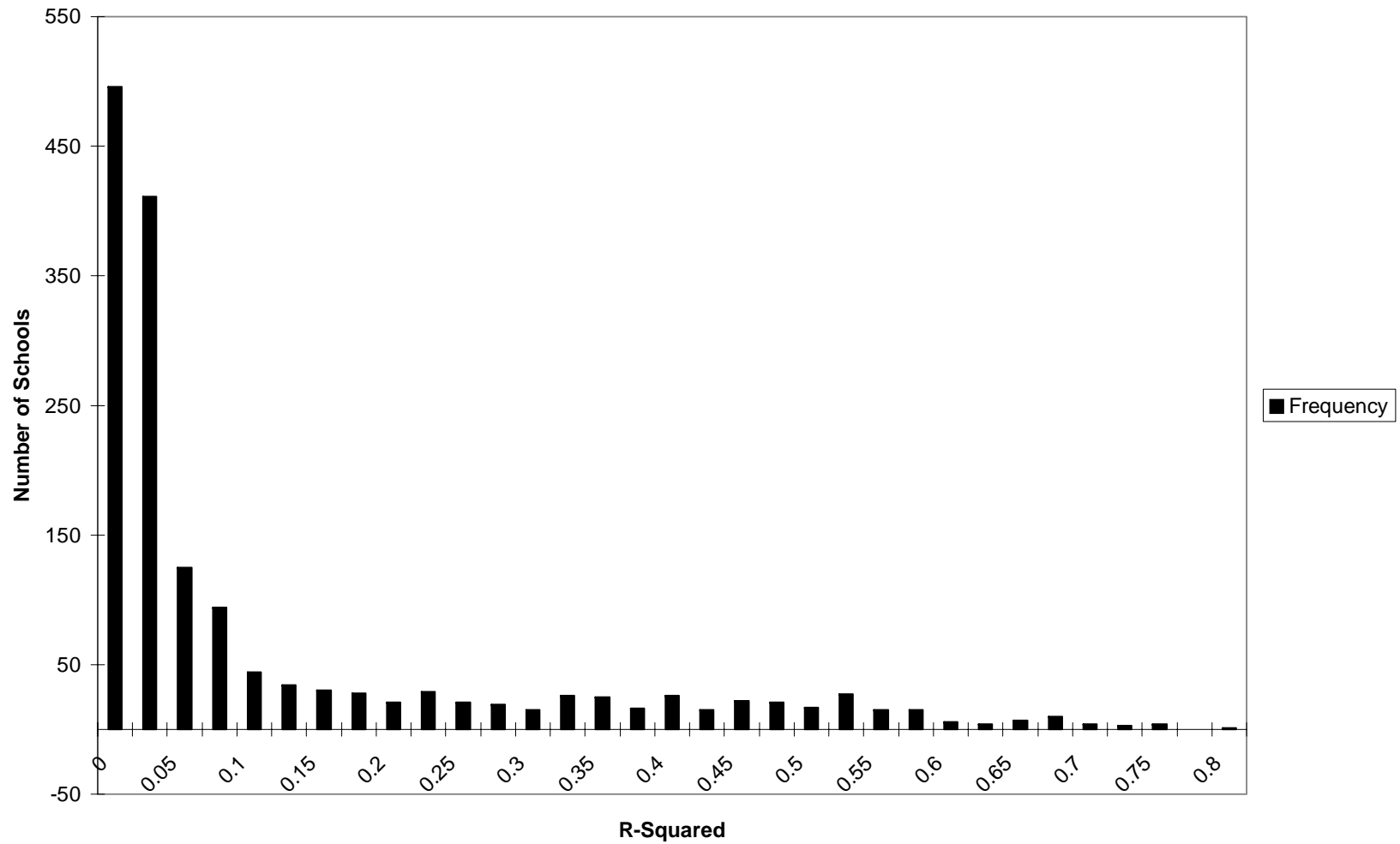


Figure 2. Histogram of Initial ITBS Scores of Third Grade Students in Two-Class School with R-Squared=0

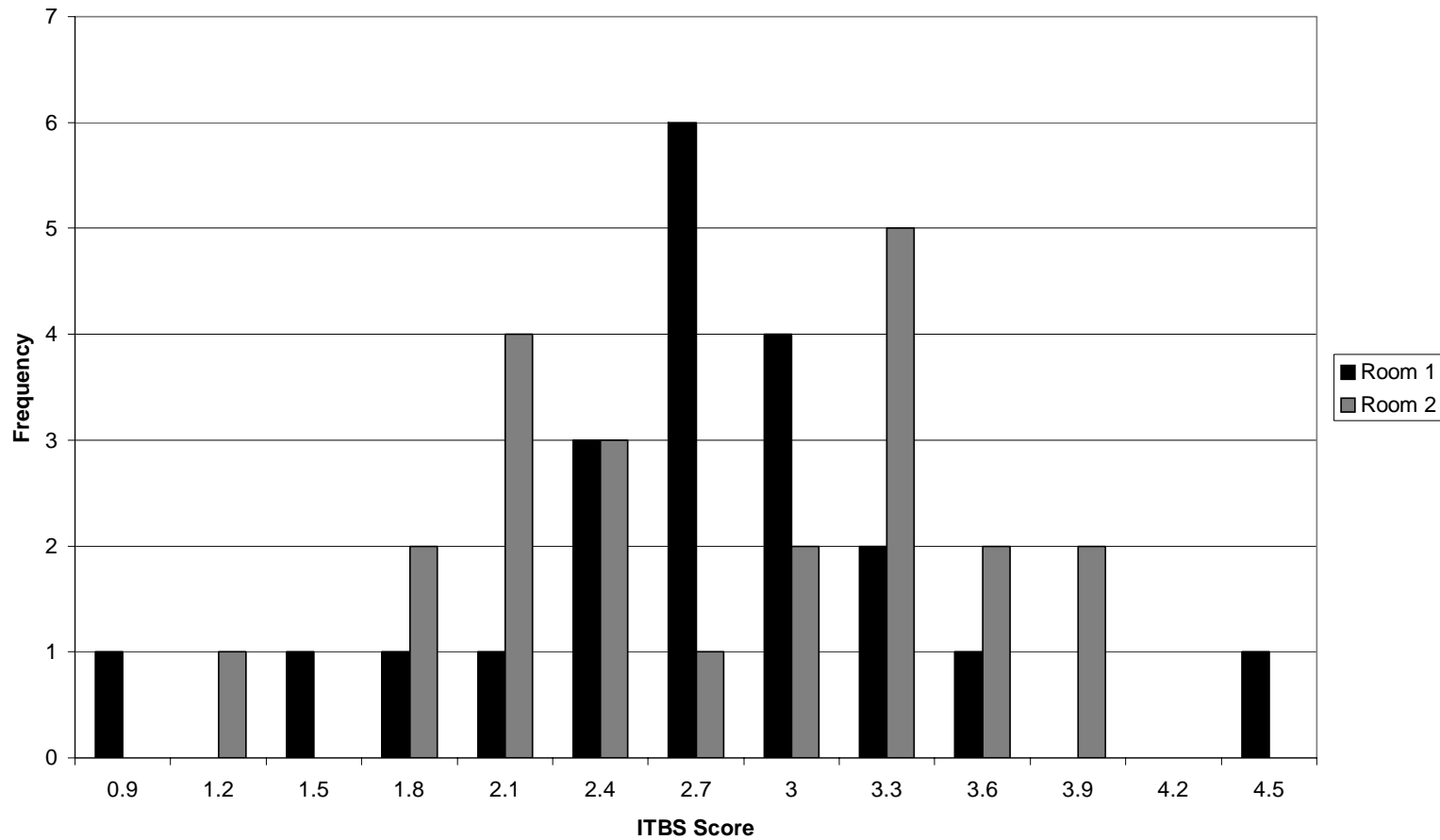


Figure 3. Histogram of Initial ITBS Scores of Third Grade Students in Two-Class School with R-Squared=.5

