Peer Groups and Academic Achievement: Panel Evidence from Administrative Data

February 2004

Julian R. Betts[a],
University of California, San Diego and Public Policy Institute of California
and
Andrew Zau,
Public Policy Institute of California

The authors thank the Public Policy Institute of California for financial support of the multi-year project underlying this research, to participants at the fall APPAM meeting, the University of Texas-Dallas and Jens Ludwig for helpful comments, and to Lorien Rice for assistance with the test score data.

[a] Corresponding author. jbetts@ucsd.edu.

Abstract

We examine the impact of classroom and grade level peer achievement on individual elementary students' rate of achievement gain in math and reading using a detailed panel data-set from San Diego Unified School District. We find that controls for unobserved student heterogeneity through addition of student fixed effects to be important, as are controls for potential simultaneity and regression to the mean. The preferred specifications suggest positive and significant effects of peer achievement. Tests for independent effects of peer achievement at the classroom and grade level suggest that the former are more important but for reading gains both types of peer achievement are significant. We also find some evidence for asymmetry in peer effects.

1. Introduction

Social scientists have for many decades conjectured that social interactions between people can directly alter a wide variety of outcomes for individuals. With externalities of this sort, it becomes unlikely that a market-based system will produce first-best outcomes. Perhaps nowhere is there greater potential for social interactions than in public schools, where children and youth spend a good portion of their young lives interacting in the classroom and in the school corridors.

Early attempts to test for the impact of the characteristics of a student's peers on his or her achievement in the 1970's produced strongly suggestive evidence that peer groups affect individual students' achievement. (See for instance, Summers and Wolfe 1977 and Henderson, Mieszkowski and Sauvageau 1976, 1978.)

Despite these early findings in favor of the notion that peer groups "matter", researchers have become increasingly aware of the difficulties inherent in identifying the true impact of peers. The foremost problem is quite straightforward to explain. Many researchers have documented that in American schools students tend to be assigned to classes in accordance with ability. This practice is known as "ability grouping", which is perhaps more properly termed "achievement grouping" because it refers to the practice of assigning students to classrooms at least roughly in line with their achievement at the end of the previous school year. In the real-world case in which researchers have only imperfect measures of individual students' achievement, the apparent impact of a student's classroom peers on his or her own achievement could in reality be proxying for the student's own imperfectly measured initial achievement. After all, if teachers observe a student over several grades, they should have better knowledge of the student's true

achievement than can be obtained any of the standardized tests that have become the focus of most research on school quality. It follows that the average achievement of the class will quite accurately reflect the student's own initial achievement, which is is imperfectly measured by researchers. A positive coefficient on the peer group's initial achievement could thus simply indicate that students with high initial achievement are more likely to increase their test scores faster than average in future grades.

Compounding this measurement error problem is what Manski (1993) has labeled the "reflection problem". The essential idea here, in the context of schools, is that a student's own achievement and that of his peers evolve in a dependent manner over time. If Johnny's achievement is affected by the achievement of his 24 classroom peers, then his 24 peers' achievement levels should also be affected by Johnny's own rate of achievement. This leads us to a classic case of simultaneity bias.

The key to unlocking both of these dilemmas is to find a way of controlling for the individual student's own initial achievement, motivation and raw ability as fully as possible, reducing the chance that any "peer group effect" is merely proxying for the student's own potential through ability grouping. At the same time, a method must be found to reduce or eliminate the simultaneity between a student's own performance and that of his peer group so as to reduce the chance that the reflection problem arises.

The goal of this paper is to address the peer group question using a richly detailed panel dataset from the eighth largest district in the United States, San Diego Unified School District. We use two central methods to remove unobserved characteristics of the individual student that might, if left unaccounted for, bias the coefficient on the peer group variable. The first method is to model gains in achievement rather than levels of

achievement, to reduce the chance that past unobserved family and school experiences create a spurious correlation between initial peer group achievement and our measure of the student's achievement. The second and more fundamental technique we employ to remove unobserved student characteristics is to use a fixed effect for each student. This removes from the model any aspect of the student's background, ability or motivation that is fixed over time. In addition, we use numerous methods to account for the possibility of the reflection problem.

While we were undertaking this research two important discussion papers that deal with peer group effects were released. Hanushek, Kain and Rivkin (2001) use student-level data from Texas to test for the existence of peer group effects on individual student achievement. Hoxby (2000) uses the same data source, using slightly different methods. This paper is written in much the same spirit as these other papers, with a focus on attempting to control for unobservable characteristics of students that might be confounded with peer group effects.

Two important distinctions exist between these papers and the present paper. First, the Texas data-set contains far more observations than are available in the San Diego data-set. However, it needs to be pointed out that with tens of thousands of observations, the present data-set is several times larger than national data-sets used by some of the earlier researchers.

A second distinction between the present paper and the two papers using Texas data is that the latter papers do not observe students at the classroom level. Thus they cannot observe the classmates of each student, or the characteristics of the teacher of each student. Thus, the present paper provides not only additional insights about the impact of

a students' peers at the grade level, but additionally tests for a separate impact of the classmates in the student's actual classroom. Attempting to identify peer effects at the classroom level does pose econometric challenges related to omitted variable bias and the reflection problem, but our approach enables a direct look at the impact of the peers in each student's classroom, with whom the student spends the bulk of the day. Intuitively, it is here, at the classroom level, that peer effects should be strongest. Past attempts to address this question have not observed the actual peers of each student and so have not been able to test for the impact of peers in the classroom.

The next section describes the method and the econometric challenges posed by peer group issues, while section 3 describes the data. Section 4 presents the results.

2. Method and Econometric Issues

To provide context, we begin with our definitions of peer group achievement at the level of the student's grade and his or her classroom. Suppose student i is in a class of n students. Define $\overline{Score}_{g,t-1}$ as the average score in grade g in period t-1 for all students in the district, with $\sigma_{g,t-1}$ representing the standard deviation across all students in the district of the score in grade g in period t-1. Then in period t, we define for student i in class c in grade g in school s:

$$(1) \qquad Peer_{icgs,t} = \frac{\dfrac{\sum_{j \neq i} Score_{j,g-1,t-1}}{n-1} - \overline{Score}_{g-1,t-1}}{\sigma_{g-1,t-1}}$$

In other words, the average classroom peer achievement variable is set to the average test score in the previous year for all of the other (n-1) students in the classroom, minus the

district average test score last year in the previous grade, and all of this divided by the standard deviation of test scores last year in the previous grade district-wide. So, a value of 1.0 for this variable means that the student's classroom peers this year on average last year scored one standard deviation above the district mean. A value of -2.5 means that the student's classroom peers last year scored 2.5 standard deviations below the district average.

The other measure of a student's peers' achievement is analogous to the above, but is defined as the average test scores last year of all the other students who this year are in that student's grade g at school s. Again, we subtract the district average and divide by the district standard deviation to standardize the measure.

A simplified model serves to highlight the key estimation issues when attempting to estimate the impact of these peer variables on student outcomes. Arguably the greatest potential bias has to do with unobserved factors that are highly correlated across students within a classroom and indeed, among schools. Here we will focus on the issue of correlations among students within a classroom, for ease of exposition, although similar issues arise concerning correlations among students within a grade at the school. Ability grouping or tracking have been much studied in the education literature. It is quite clear that on the whole U.S. schools often assign students to classrooms at least partially based on their previous achievement. See for example Oakes (1990), Argys Rees and Brewer (1996) and Betts and Shkolnik (2000).

In a very simple example suppose that gains in test scores, or $\Delta Score_{icgst}$ for student i in classroom c in grade g in school s in year t, equals the district-wide average rate of growth in achievement, ω, and that in addition test score gains for this student

depend on an error component α_i that captures the student's innate motivation, ability and related factors, plus a white noise error term:

(2) $$\Delta Score_{icgst} = \omega + (\alpha_i + \varepsilon_{icgst})$$

In this case, we are assuming that peer group achievement has no impact upon individual student learning. However, if we measured initial peer group achievement in the student's classroom, $PEER_{icgst}$, and added it to the constant in the model, the use of ability grouping in the classroom suggests that there will be a positive correlation between this peer variable and the fixed student ability/motivation component of the error term. We will obtain upwardly biased (positive) coefficients on the peer group variable, the true value of which is zero. Even in large samples we will obtain inconsistent estimates of the peer group variable. To see this, note that if we estimate:

(3) $$\Delta Score_{icgst} = \omega + \rho PEER_{icgst} + error_{icgst}$$

without taking explicit account of the random effect α_i in the error term, then applying the Frisch-Waugh-Lovell theorem, this model can be estimated by subtracting the mean of each variable on both sides of the equation and then estimating the revised model with all variables having zero means. The probability limit for the coefficient on the peer group in this model is:

(4) $$\text{plim}\hat{\rho} = 0 + \frac{\sigma_{\left((PEER_{icgst} - \overline{PEER})(\alpha_i - \overline{\alpha})\right)}}{\sigma^2_{(PEER_{icgst} - \overline{PEER})}} > 0$$

(This holds under the assumption of no correlation between the white noise error component ε_{icgst} and the PEER variable but a positive correlation between α_i and the PEER variable.) Thus the peer coefficient estimate is inconsistent and is biased upwards.

Similarly, the peer effect at the grade level might be capturing something unobserved about the student, or the entire grade or school.

Our solution to this impasse is to account for the student error component α_i explicitly by adding fixed effects for each student to sweep the α_i term out of the error term, removing both bias and inconsistency in the above example. Our basic regression model adds dummy variables to account for observed and unobserved variations across not only students, but their schools and their home neighborhoods, as proxied by home zip codes, as well as dummies for calendar year and grade. This approach of using fixed effects can remove unobserved heterogeneity related to schools, grade levels, and even the year in which the test is given, the latter of which controls for variations in test familiarization. We model gains in test scores, or $\Delta\text{Score}_{icgst}$ for student i in classroom c in grade g in school s in year t as a function of school, family and personal, and classroom characteristics. (Classroom characteristics include teacher characteristics, class size and classroom peer test scores.) Our regression model is

(5)

$$\Delta Score_{icgst} = \alpha_i + \beta_s + \chi_{Zipcode_{it}} + \delta_t + \phi_g + SCORE_{icgs,t-1}\pi + \mathbf{FAMILY}_{it}\mathrm{E} + \mathbf{PERSONAL}_{it}\Phi$$
$$+ PEER_{igs,t}\rho_g + PEER_{icgs,t}\rho_c + \mathbf{CLASS}_{icgst}\Gamma + \mathbf{SCHOOL}_{ist}\Lambda + \varepsilon_{it}$$

where the first five variables on the right-hand-side represent fixed effects for the student, his or her school, the student's home zip code, the year in which the test was given and the grade in which the test was given, $\text{SCORE}_{icg,t-1}$ is the lagged score of the student, items in bold characters indicate vectors of time-varying family, personal and classroom and school characteristics, the Peer variables capture the initial achievement of peers in the grade ($\text{PEER}_{igs,t}$) and the classroom ($\text{PEER}_{icgs,t}$), the corresponding Greek letters are

vectors of coefficients, and ε_{it} is an error term. The parameters of interest are ρ_g and ρ_c that measure the impact of grade-level and classroom-level peers on gains in individual achievement. The inclusion of the student fixed effects removes the problem of heterogeneous student rates of improvement.

Aspects of the above regression specification, including the use of the lagged test score as a regressor, are designed to contend with two other potential issues – the reflection problem and regression to the mean. We discuss these in turn.

Manski (1993) refers to simultaneity bias in the context of social interaction models as the reflection problem. This simultaneity becomes clear through a simple example. If we wanted to test whether the probability that individual A is more likely to smoke if the proportion of his peers who smoke rises, we must allow for the possibility that individual A himself influences his peers' probability of smoking. Thus we cannot simply model the probability that person A smokes as a function of whether his peers in the same period smoke as these variables are simultaneously determined.

To reduce or minimize this potential problem we need to find ways of decoupling individual A's outcome in period t from that of his peers in period t. The model of achievement above already implements two innovations that should reduce the reflection problem. First, unlike the example of modeling the contemporaneous correlations between individuals' smoking behavior, we are not modeling the level of student A's test score as a function of the contemporaneous scores of his classroom peers; rather, we are modeling the *gains* in the individual's scores as a function of the *past* mean score levels of his current peer group. The two variables, gains in own test scores and lagged peer scores, are related but do differ. Moreover, we use last year's peer group achievement

rather than the current year peer group to lower the risk of simultaneity between it and *current year* gains for the individual student.

Neither of these innovations is a panacea. In particular, by construction, our dependent variable $\Delta Score_{icgst}=Score_{icgst} - Score_{icgs,t-1}$ is negatively correlated with the student's own lagged score, which will be jointly determined with the test scores of those of the student's current year peers who were in the same class last year. This would tend to bias the peer group coefficients downward.

Another problem that, to the best of our knowledge, has received little previous attention in the literature on peer group effects on test scores, is regression to the mean. Suppose that for random reasons a certain class does poorly in a given year, and that in the next grade most of the same students stay together in a new classroom. This suggests that at the end of that first school year, peer scores will be unusually low, as will the specific student's score. But if the next year's classroom is better, for random reasons, we would expect large gains in the student's own test score as he or she "regresses to the mean". For example, having a bad teacher in grade two followed by an average teacher in grade 3 would likely generate this pattern. The mirror image story will occur if a certain class does particularly well for random reasons. The upshot in either case is that regression to the mean will induce a negative correlation between the student's gain in test scores and the prior year's scores of his peer group.

This problem is likely to be more than theoretical. Betts and Danenberg (2002) find evidence of some regression to the mean in test scores reported at the school level in California. Because test scores that are averaged over many students at a school will

have less noise than individual test scores, we would expect this problem to be more important at the level of the individual student.

For both of the reasons -- the reflection problem and the related problem of regression to the mean -- we expect the coefficient on peer scores to be biased down. To reduce the negative correlation between the student's own gain and peers' lagged scores, we include the student's own lagged score on the right hand side to purge other regressors including the peer group variables of any correlation with the student's own lagged score. This transforms the peer group variables so that their effect is now identified by inter-year variations for a student that are not linked to the students's own lagged score. In addition, the lagged student score variable along with the student fixed effects controls for unusual rises or declines in the student's lagged test score, in so doing reducing the likelihood of a negative bias on the classroom peer coefficient caused by regression to the mean. To see why this works, note that if after taking account of the student effects by de-meaning variables on both sides of the equation, a negative (de-meaned) value of the lagged student test score indicates that the student scored below his or her average in other years. If this shock also affected classmates, and the student regresses to the mean during the current year through a larger than average gain in achievement, we induce a spurious negative correlation between the peer score and the dependent variable. But including the lagged student test score will capture such types of regression to the mean, preventing a downward bias on the peer test score that would otherwise result.

3. Data

Our data come from the administrative data systems of the San Diego Unified School District. The dependent variable is the annual gain in scores on the Stanford 9 test in math and reading. Beginning in spring 1998, the Stanford 9 test has been given annually to all students in California in grades 2 through 11 inclusive. Although the test results are provided in a number of formats, included national percentile rankings, we use the scaled scores. Based on psychometrically based norming, the scaled scores are designed to compare what a student knows between one year and the next. [1] Our data comprise data from the first four waves of this testing (spring 1998 through spring 2001). We focus on elementary school students because unlike middle and high school students, elementary students spend substantial parts of their school day in the classroom with the same peers.

Our models condition upon measures of individual student background that are fixed over time, such as race, by including a fixed effect for each student. In addition we control for background variables that potentially can change over time including parental education and home zip code. Table 1 details the controls we include for student, family and neighborhood.

Table 2 lists regressors that we include at the school, grade and classroom levels. In brief, at the school level, we control for all observed and unobserved characteristics of schools by including dummy variables for each school. In addition we control for school characteristics that could change over time, including student body characteristics. At the classroom level, we control for class size and an unusually rich depiction of the individual teacher's background in terms of credentials, experience, education, language

[1] In addition, the scaling procedure ensures, insofar as possible, that "a difference of 5 points between two students' scores represents the same amount of difference in performance wherever it occurs on the scale".

certifications and race and gender. The latter is the product of a six-month collaborative project with the district to convert its human resources databases to a format that captures teacher characteristics longitudinally.

The lagged peer group test scores have already been defined in the previous section of the paper.

4. Results

Based on the econometric discussion in section 2, we expect that adding student fixed effects to the model will lower the coefficients on the peer group variables, because the student fixed effects will remove spurious positive correlation between the student's own unobserved ability and motivation and the lagged peer group score. This prediction is largely borne out by the results. [2]

Table 3 shows models of gains in reading and math achievement for all students in elementary schools in the district. Columns 1 and 2 show results for reading. Column 1 uses the specification in equation (5), while column 2 removes the student fixed effects to test the prediction that the student fixed effects will remove omitted ability bias.

In column 1 lagged peer achievement at *both* the classroom and grade levels are positive and significant. Column 2, which is included only to examine the contention that failure to control for student fixed effects could bias the peer group coefficients up, shows that indeed the coefficient on classroom peer scores rises considerably without

---

(Harcourt Brace Educational Measurement, 1997, p. 17)

[2] In addition, because of simultaneity bias between the student's lagged score and peers' lagged scores as well as regression to the mean, we would expect a downward bias on the peer group coefficients in a model that did not attempt to control for these potential problems. Although we do not show these results, this is exactly what happens, with the peer group variables typically *negative* and significant. This pattern is entirely consistent with both regression to the mean and simultaneity bias. In contrast, in the models

student fixed effects. However, the grade level peer coefficient is actually lower and negative in this model. It could be that omitted ability bias manifests itself mainly by the way students are assigned to the classroom more than the way students are assigned to schools.

We view model 1 as the preferred model, given that it alone controls for both omitted student ability or motivation and simultaneity/regression to the mean, while at the same time our predictions of the directions of change in coefficients when we correct for these two issues are largely borne out by the data.

The table excludes coefficients on other regressors to conserve space, but as a point of comparison it does include the coefficient on class size. This variable is always negative and significant in the reading models.

Columns 3 and 4 show corresponding results for gains in math scores for all students. The patterns in this table are very similar to what we have just seen for reading scores. The classroom peer variable is highly significant and positive in the preferred model with student fixed effects. The grade level peer score is positive but not quite significant at 5%. This model has lower coefficients on the classroom peer variable than does model 4 that omits student fixed effects, suggesting that as we expected in models without student fixed effects we overstate the impact of peers due to omitted ability/motivation bias. As before we do not find this pattern with respect to the grade level peer measures.

At the bottom of the table we perform some simulations of the predicted impact of changing peer group achievement on the student's own test score gains. We present the

---

presented here, peer effects are generally positive. Moreover, the lagged student test score is highly significant and negative, which is exactly what we would expect in the presence of regression to the mean.

predicted changes as a percentage of the average annual gain in scaled scores in our sample, which, for example is a gain of 27.6 points in reading. Thus, if changing the peer achievement by one unit led to a 2.8 point predicted increase in the student's annual test score gain, we would report this as a 10% gain in the rate of learning. Although we present predicted gains for all the models, we focus our discussion on the predictions for models 1 and 3, because they are the models for reading and math that simultaneously control for omitted ability/motivation and for simultaneity between the student's lagged score and his peers' lagged scores.

The first two rows in the simulations show the predicted impact of increasing classroom and grade level peers' achievement from the 25th to 75th percentile district-wide. The predicted impacts on gains in reading achievement are sizable, with the classroom and grade-level peer effects predicted to increase the annual gain in test scores by about 5.0% and 5.6% respectively. The next pair of rows show a more conservative simulation which is to increase either type of peer score by the median of the absolute value of the actual observed changes for individual students. (For reading, this median real-world change is 0.26 for the classroom peer scores and 0.11 for grade level peer scores. For the math results that we will discuss later, the variations are slightly bigger, at 0.16 and 0.30 respectively.) The predicted gains in the annual change in reading scores from these more modest increases in peers' lagged test scores are 1.3 and 0.7% for classroom and grade level peer effects respectively. By way of comparison, a reduction in class size of 5 students in model 1 is predicted to increase the annual test score gain in reading by about 1.6%.

Interquartile changes in math peer scores at the classroom and grade levels are predicted to increase the annual gain in math scores by about 6.5% and 3.6% respectively (although the second estimate is not quite statistically significant at conventional levels). The more conservative estimates using the median absolute value of actual changes in peer group scores, at the bottom of the table, suggest increases in the rate of learning of 2.1% and 0.7% at the classroom and grade levels, although the latter is not significant. In comparison, a reduction in class size by 5 students has no statistically significant effect, and the point estimate is only 0.7%.

Robustness

Tables A-1 and A-2 in the appendix show the results of various robustness checks for reading and math models respectively. There, we enter the classroom and grade level peer groups separately instead of together. Column 1 replicates the results from Table 3, while columns 2 and 3 drop first grade level peer scores and then the classroom level peer scores. The results were not markedly different, except that when we entered the grade level peer group variable alone, similar to what Hanushek, Kain, Markman and Rivkin (2001) were able to do with their grade-level data, the grade level peer coefficient becomes bigger, essentially because it is capturing the sum effect of the classroom and grade level peer effects. This also explains why the significance of the grade level peer score is much higher in the models that drop the classroom peer effect.

The final column in appendix Tables A-1 and A-2 repeats model 1 but replaces the gain in the test score as the dependent variable with the level of the score, S_{icgst}. This is one of the standard formats for the value-added approach to modeling school quality,

although it lacks the intuitive flavor of our preferred specification in which the metric is gains in test scores. The results are very close to those in our preferred model.

English Learners

A key policy issue in California is how well English Learners (EL students) are learning. Table 4 replicates the specifications in table 3 except that the sample is the EL population. The subsample is much smaller, but we find similar patterns. For reading, in the preferred specification in column 1, the classroom peer effect is highly significant and the coefficient is roughly double the corresponding coefficient in the model for all students. The simulations at the bottom of the table suggest bigger effects for changes in the classroom peer score for EL students than for all students. On the other hand, the grade level peer coefficient is not significant. Notably, the impact of class size variations, like the classroom peer effect, is bigger in the EL sample than in the sample of all students.

The math results for EL students in table 4 shows the same patterns relative to the math results for all students, with the peer classroom effect being both significant and bigger than for all students in the EL sample. As for the reading results for EL students, the grade level peer effect is not statistically significant.

Implications for the Aggregate Effects of Ability Grouping vs. Heterogeneous Grouping

The impact of peer groups is of crucial policy importance with regards to whether the school uses ability grouping or heterogeneous grouping. Who would win and lose from ending ability grouping? An interesting question here is whether the peer effects

are symmetric regardless of whether a student is placed in a relatively high or low peer group. If the effects are symmetric, it suggests a zero-sum game. That is, if we went from completely heterogeneous classes with the same mean peer score in each class to a case of complete ability grouping in which class assignments were made in strict order based on student test scores, the gains to stronger students would be exactly offset by losses to the weaker students. In theory, though, the gains could be either greater than or less than the losses, suggesting a non-zero-sum game.

To address this issue, we re-ran the models in Table 3 after adding indicator variables aboveclass and abovegrade, which indicate whether the test scores of a student's peer group at the class and grade levels in a given year were above the average for that student over all of the years we observe. In addition we add interactions between these two dummy variables and the corresponding peer group test scores.

Table 5 shows the results for reading in the first two columns and math in the second two columns. Columns 1 and 3 show the preferred models for reading and math, which include student fixed effects as well as a lagged test score. (Columns 2 and 4 show the same models but without student fixed effects.) Models 1 and 3 show that neither of the dummies indicating that the student's peer group was above that student's average are significant, but the interactions of these dummies with the peer scores themselves are negative and significant.

What does this imply? Imagine a student who has a school year in which the lagged scores of her classroom and grade level peers is the average she experiences over the period under study. If next year her peer group improves, then the aboveclass and abovegrade dummies will equal one. A few calculations show that *an increase in the*

student's peer group does less to increase her rate of learning than a decrease in her peer group does to decrease her rate of learning. The simulations at the bottom of the table provide documentation. For instance, model 1 shows that an interquartile increase in the classroom peer score in reading is predicted to increase achievement gains by about 4.0% while a similar decrease in peer scores is predicted to lead to a decline of 8.6%. For math, the results in column 3 are even more dramatic with predicted changes of +5.1% and -9.9% respectively.

While preliminary, these results suggest that ending ability grouping might not be a zero-sum game: it may hurt the achievement gains of the top-performing students more than it helps the achievement gains of the lower-achieving students.

Sources of Identification: Learning from Subsamples

While evidence that peers matter seems strong, especially at the classroom level, it is natural to ask whether the source of variation that is identifying these effects is students who switch between elementary schools or students who stay in the same school but experience slightly different peers each year. A related but not identical question is whether students with abnormally large changes in peer groups are driving the results. Tables 6 and 7 address these issues for reading and math respectively.

We divided students into those who did not switch between elementary schools during the sample period and those who did. The first two columns of the tables show the results when we re-estimated our preferred student fixed-effect model on these two subsamples. We find evidence that classroom peers matter for non-switchers but we find no such evidence for school switchers. The standard errors on the peer group coefficients

for the relatively small sample of school switchers are fairly big, meaning that we cannot say for certain that the impact of peer groups is different between the two groups. However, in three out of four cases, the coefficients on the peer variables are smaller for school switchers than the non-switchers. This and the lack of precision in the estimates for school switchers suggest that the main source of identification is small changes in peer groups from one year to the next among students who do not switch schools. The implication is that our results come closer to finding that ability grouping within a school matters than to finding that school switching matters. Of course, a larger sample might have revealed different results, especially for the school switchers.

Column 3 of Tables 6 and 7 remove students whose one-year changes in either type of peer group score were in top 10%, when ranked in terms of the absolute value of the changes. In other words, we are deleting outliers. Interestingly, the grade level peer coefficients rise considerably and become highly significant for both reading and math. The classroom peer effects are little changed. What we conclude is that students who experience the largest changes in peer scores are not responsible for our finding that peers matter, and in fact these outliers make the impact of grade level peers look smaller. Again, this hints at the idea that within-school changes in peers may matter more than dramatic changes in peers that can accompany school switching. [3]

5. Conclusion

This paper presents evidence that the initial achievement of classroom peers is a highly significant predictor of student's gains in both math and reading achievement at

the elementary level. This result arises once we control for possibly endogenous changes in student's recent achievement, and is robust to controls for a wide variety of personal, school and classroom characteristics including detailed teacher qualifications, and student fixed effects. We view the inclusion of student fixed effects as essential for controlling for unobserved student ability, motivation and environmental factors. In essence, we have shown small year-to-year changes in students' peers appear to have important effects on students' rate of learning.

Some of the most convincing work done to date on peer effects, by Hoxby (2000) and Hanushek, Kain, Markman and Rivkin (2001), has measured peers at the grade level. Although the overall results of the present study support those in these other papers, we have been able to observe students at the classroom level rather than by grade. Unlike most earlier papers our unusually rich panel data-set has allowed us to test whether peer effects work solely through the classroom, solely through one's peers in a given grade, or whether both mechanisms are at work.

For math achievement, the peer effects seem to work mainly through the classroom. For reading, we find evidence that both mechanisms are at work, although the classroom evidence is more significant in a statistical sense.

The simulated effect of an interquartile change in peer scores is quite large. But the more conservative and convincing simulation is the predicted effect of changing a student's classroom and grade-level peer scores by the median of the absolute value of the actual year-to-year changes. The predicted increases in the rate of student gain in reading and math from these latter changes in classroom peer scores are 1.3% and 2.1%

[3] We also tried replicating the models with only classroom or grade level peer scores, after deleting students whose absolute value of one-year gains for that specific measure of peer group in any year identified them

respectively. These changes are of a plausible size, and also compare favorably with the predicted impact of fairly large changes in class size.

We estimated regressions on subsamples to test whether our results depend on students with large changes in peer groups, such as those that could result from switching schools. We found that our identification comes more strongly from small changes in peer groups, and from students who do not switch schools.

We also find evidence that the effects of changing a student's peer group are not symmetric around zero. That is, an increase in a student's peers' achievement is predicted to increase that student's rate of achievement gain less than an identically sized decrease in peer achievement is predicted to lower the student's rate of achievement gain. The implication is that ending ability grouping in favor of heterogeneously grouped classes might hurt the top-achieving students more than it would help the bottom-achieving students.

---

as being in the top 10% of changes. The results were similar to what we report above.

Table 1 A List of Student, Family and Neighborhood Controls Used in the Statistical Models

Student Characteristics
Fixed effects for each student to control for all characteristics of a student which are fixed over time, such as race. Fixed effects for grade and year. In addition the student's own lagged test score, controls for students who changed schools that year, switched schools unexpectedly, were new to district, age, grade level, English Learner (EL) status, Fluent English Proficient (FEP), non-Spanish EL, non-Spanish FEP, special education, skipped a grade that year, retained a grade that year, % of days absent, and indicator variables for students who skipped a grade that year, were retained a grade that year, changed schools that year, switched schools unexpectedly, or were new to the district.
Family Characteristics
Controls for the level of education of the more highly educated parent.
Neighborhood Characteristics
Fixed effects for student's home zip code.

Table 2 School, Classroom and Student Body Controls Used in the Statistical Models

School Characteristics
Fixed effects for each school to control for all fixed characteristics of the school. In addition controls for whether the school ends in grade 5 or higher, and for whether it was a year-round school, a charter school or an atypical school. (In a small number of cases these latter characteristics change over time.)
Student Body Characteristics at the School Level
% eligible for free/reduced-price lunch; separate controls for % of students who are Hispanic, black, Asian, Pacific Islander, native American; % of students who are EL, % FEP; controls for student mobility: % who changed schools that year, who switched schools unexpectedly, and who were new to the district.
Student Body Characteristics at the Grade Level
Mean test scores in previous spring's test of all students in the student's current grade, standardized to district average.
Classroom and Teacher Characteristics
Class size, controls for teacher characteristics: interactions of credentials (intern, emergency credential, full credential) with indicators of years of teaching experience (e.g. 0-1, 2-5, 6-9), master's degree, Ph.D., bachelor's in math, English, social science, science, language, other major (except education) (separate variables for each major), corresponding controls for minors by field except that the omitted group is teachers with a minor in education or other, CLAD, (Spanish) BCLAD, CLAD alternative credential, BCLAD alternative credential, interactions for the last four dummy variables with two student indicators for EL status and for FEP status, controls for teachers who are black, Asian, Hispanic, other non-white, and female.
Average Student Characteristics in the Classroom
Mean test scores of students in the class in previous spring's test, standardized to district averages.

Note: CLAD stands for Crosscultural, Language and Academic Development (CLAD) certificate, which prepares teachers to teach students who are English Learners. A closely related credential is the Bilingual CLAD (BCLAD) certificate, which certifies that a teacher is equipped to teacher English Learners in a language other than English.

Table 3 Results for Reading and Math Achievement Gains for All Students

Model	1	2	3	4
Dependent variable is gains in achievement in:	Reading	Reading	Math	Math
Student fixed effects	Yes	No	Yes	No
Peer scores in classroom	1.3973	5.6114	1.9295	6.9902
	(0.3023)**	(0.2134)**	(0.3265)**	(0.2273)**
Peer scores in grade	1.7105	-11.0385	1.2539	-15.7064
	(0.7304)*	(0.5390)**	(0.6577)	(0.4802)**
Class size	-0.0840	-0.1222	-0.0378	-0.0888
	(0.0331)*	(0.0271)**	(0.0357)	(0.0291)**
Predicted effect of changing from 25th to 75th percentile of classroom peer variable	4.97%	19.97%	6.45%	23.37%
Predicted effect of changing from 25th to 75th percentile of grade-level peer variable	5.64%	-36.44%	3.56%	-44.64%
Predicted effect of changing of classroom peer variable by median absolute value of actual change	1.34%	5.37%	2.12%	7.67%
Predicted effect of changing of grade peer variable by median absolute value of actual change	0.69%	-4.44%	0.72%	-9.02%
Number of observations	74557	74557	77897	77897
R-Squared	0.73	0.25	0.73	0.23

Note: All models contain the list of regressors in Tables 1 and 2.

** Significant at the p<0.05 level
** Significant at the p<0.01 level

Table 4 Table 5 Results for Reading and Math Achievement Gains for English Learner (EL) Students

Model	1	2	3	4
Dependent variable is gains in achievement in:	Reading	Reading	Math	Math
Student fixed effects	Yes	No	Yes	No
Peer scores in classroom	3.1764	7.3921	3.4795	6.7734
	(0.6016)**	(0.4166)**	(0.6361)**	(0.4505)**
Peer scores in grade	0.9294	-6.9712	1.5296	-9.1796
	(1.4436)	(0.9631)**	(1.2849)	(0.8937)**
Class size	-0.1588	-0.0635	-0.1815	-0.1215
	(0.0646)*	(0.0512)	(0.0681)**	(0.0550)*
Predicted effect of changing from 25th to 75th percentile of classroom peer variable	9.17%	21.35%	8.44%	16.42%
Predicted effect of changing from 25th to 75th percentile of grade-level peer variable	2.32%	-17.41%	3.23%	-19.39%
Predicted effect of changing of classroom peer variable by median absolute value of actual change	3.04%	6.46%	3.40%	6.62%
Predicted effect of changing of grade peer variable by median absolute value of actual change	0.35%	-2.65%	0.73%	-4.35%
Number of observations	19843	19843	21453	21453
R-Squared	0.72	0.20	0.75	0.24

See notes to table 3.
** Significant at the p<0.05 level
** Significant at the p<0.01 level

Table 5 Tests for Asymmetric Effects of Changing Peer Achievement in Models of Reading and Math Achievement Gains for All Students

Model	**1**	**2**	**3**	**4**
Test Score Modeled:	**Reading**	**Reading**	**Math**	**Math**
Student fixed effects	**Yes**	**No**	**Yes**	**No**
Peer scores in classroom	2.4133	6.6186	2.9674	8.1867
	(0.4605)**	(0.2418)**	(0.5042)**	(0.2615)**
Peer scores in grade	2.2338	-9.5182	3.3468	-13.2927
	(0.9678)*	(0.5752)**	(0.8899)**	(0.5245)**
Above average peer scores in class (aboveclass)	-0.2245	-2.6799	-0.2392	-4.3539
	(0.2202)	(0.2021)**	(0.2482)	(0.2199)**
Above average peer scores in grade (abovegrade)	0.0363	-0.8221	-0.5812	-1.4889
	(0.2146)	(0.2047)**	(0.2391)*	(0.2264)**
Aboveclass*peer score	-1.0652	-0.0473	-1.1687	1.0033
	(0.2277)**	(0.2681)	(0.2736)**	(0.3139)**
Abovegrade*peer score	-1.1112	-1.4950	-1.0876	-1.2909
	(0.3050)**	(0.3467)**	(0.3717)**	(0.4220)**
Class size	-0.0917	-0.1064	-0.0495	-0.0680
	(0.0332)**	(0.0270)**	(0.0358)	(0.0290)*
Simulations assuming peer scores increase from average peer score for individual of zero				
Predicted effect of changing from 25th to 75th percentile of classroom peer variable	3.98%	13.66%	5.14%	25.28%
Predicted effect of changing from 25th to 75th percentile of grade-level peer variable	3.83%	-39.25%	4.29%	-46.89%
Predicted effect of increasing classroom peer variable by median absolute value of actual change	0.48%	-3.43%	1.10%	4.64%
Predicted effect of increasing grade	0.58%	-7.42%	-0.83%	-13.82%

peer variable by median absolute value of actual change				
Simulations assuming peer scores decrease from average peer score for individual of zero				
Predicted effect of changing from 75th to 25th percentile of classroom peer variable	-8.59%	-23.55%	-9.92%	-27.37%
Predicted effect of changing from 75th to 25th percentile of grade-level peer variable	-7.36%	31.35%	-9.51%	37.78%
Predicted effect of reducing classroom peer variable by mean absolute value of actual change	-2.31%	-6.34%	-3.26%	-8.98%
Predicted effect of reducing grade peer variable by mean absolute value of actual change	-0.90%	3.83%	-1.92%	7.63%
Number of observations	74557	74557	77897	77897
R-Squared	0.73	0.25	0.73	0.24

** Significant at the p<0.05 level
** Significant at the p<0.01 level

Table 6 Alternative Samples for Models of Reading Achievement Gain, All Students

Model	1	2	3
Subsample	Non-Switchers	Switchers	Remove students whose change in peer scores are outliers.[a]
Student fixed effects	Yes	Yes	Yes
Peer scores in classroom	1.2733	0.2939	1.2639
	(0.3475)**	(1.1023)	(0.3569)**
Peer scores in grade	1.6718	1.0915	3.8492
	(0.8836)	(2.5173)	(0.9140)**
Class size	-0.0989	-0.0016	-0.0879
	(0.0392)*	(0.1219)	(0.0369)*
Number of observations	63647	10910	67519
R-Squared	0.74	0.77	0.73

Note: The dependent variable is gains in test scores. All models contain the list of regressors in Tables 1 and 2 including a lagged student test score and student fixed effects.

[a] Column 3 removes students whose gains in peer scores of either type are in the top 10% of the absolute values of inter-year gains in any year.

** Significant at the p<0.05 level
** Significant at the p<0.01 level

Table 7 Alternative Samples for Models of Math Achievement Gain, All Students

Model	1	2	3
Subsample	Non-Switchers	Switchers	Remove students whose change in peer scores are outliers.[a]
Student fixed effects	Yes	Yes	Yes
Peer scores in classroom	1.8225	1.1175	1.8080
	(0.3754)**	(1.1635)	(0.3798)**
Peer scores in grade	0.6505	2.3000	2.2195
	(0.7883)	(2.1946)	(0.8061)**
Class size	-0.0523	0.0212	-0.0202
	(0.0423)	(0.1274)	(0.0390)
Number of observations	66290	11607	70349
R-Squared	0.73	0.79	0.74

Note: The dependent variable is gains in test scores. All models contain the list of regressors in Tables 1 and 2 including a lagged student test score and student fixed effects.

[a] Column 3 removes students whose gains in peer scores of either type are in the top 10% of the absolute values of inter-year gains in any year.

** Significant at the p<0.05 level
** Significant at the p<0.01 level

Table A-1 Alternative Specifications of Reading Achievement, All Students

Model	1	2	3	4
Dependent Variable	Gain in test score	Gain in test score	Gain in test score	Level of test score
Student fixed effects	Yes	Yes	Yes	Yes
Peer scores in classroom	1.3973	1.6220		1.5695
	(0.3023)**	(0.2867)**		(0.2972)**
Peer scores in grade	1.7105		2.7820	2.4210
	(0.7304)*		(0.6929)**	(0.6419)**
Class size	-0.0840	-0.0837	-0.0674	-0.0770
	(0.0331)*	(0.0331)*	(0.0329)*	(0.0324)*
Number of observations	74557	74557	74557	74557
R-Squared	0.73	0.73	0.73	0.63

Note: All models contain the list of regressors in Tables 1 and 2 including a lagged student test score and student fixed effects.

** Significant at the p<0.05 level
** Significant at the p<0.01 level

Table A-2 Alternative Specifications of Math Achievement, All Students

Model	1	2	3	4
Dependent Variable	Gain in test score	Gain in test score	Gain in test score	Level of test score
Student fixed effects	Yes	Yes	Yes	Yes
Peer scores in classroom	1.9295	2.1910		1.9607
	(0.3265)**	(0.2963)**		(0.3259)**
Peer scores in grade	1.2539		2.8866	1.0673
	(0.6577)		(0.5972)**	(0.6514)
Class size	-0.0378	-0.0377	-0.0198	-0.0375
	(0.0357)	(0.0357)	(0.0356)	(0.0356)
Number of observations	77897	77897	77897	77987
R-Squared	0.73	0.73	0.73	0.61

Note: All models contain the list of regressors in Tables 1 and 2 including a lagged student test score and student fixed effects.

** Significant at the p<0.05 level
** Significant at the p<0.01 level

References

Argys, L. M., Rees, D. I., Brewer, D. J., 1996. *Detracking America's Schools: Equity at Zero Cost?* Journal of Policy Analysis and Management 15, (4), 623-645.

Betts, Julian R. and Anne Danenberg (2002, "School Accountability in California: An Early Evaluation", in Diane Ravitch (Ed.), **Brookings Papers on Education Policy 2002**, (Washington, D.C.: The Brookings Institution), pp. 123-197

Betts, Julian R. and Jamie L. Shkolnik, (2000), "The Effects of Ability Grouping on Student Math Achievement and Resource Allocation in Secondary Schools", *Economics of Education Review*, (**19**:1), pp. 1-15.

Harcourt Brace Educational Measurement (1997), Stanford Achievement Test Series Spring Norms Book, Ninth Edition, San Antonio, TX: Harcourt Brace and Company.

Hanushek, Eric A., John F. Kain, Jacob M. Markman and Steven G. Rivkin (2001), "Does Peer Ability Affect Student Achievement?" National Bureau of Economic Research Working Paper 8502.

Henderson, Vernon, Peter Mieszkowski and Yvon Sauvageau (1976), *Peer Group Effects and Educational Production Functions*, Ottawa, Ontario, Canada: Economic Council of Canada.

Henderson, Vernon, Peter Mieszkowski and Yvon Sauvageau (1978), “Peer Group Effects and Educational Production Functions”, *Journal of Public Economics*, (10), August, pp. 97-106.

Hoxby, Caroline (2000), “Peer Effects in the Classroom: Learning from Gender and Race Variation”, National Bureau of Economic Research Working Paper 7867.

Manski, Charles F. (1993), “Identification and Endogenous Social Effects: The Reflection Problem”, *Review of Economic Studies* (60), July, pp. 531-542.

Oakes, J., (1990), *Multiplying Inequalities - The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science*. RAND, Santa Monica.

Summers, Anita and Barbara Wolfe (1977), “Do Schools Make a Difference?”, *American Economic Review* (67), September, pp. 639-652.