Classroom Peer Effects and Academic Achievement: Evidence from a Chinese Middle School

Katherine Carman and Lei Zhang[1]

April 2008

Abstract

This paper estimates peer effects on student achievement using a panel data set obtained from a middle school in China. Two unique features of the organization of Chinese middle schools (Grades 7 to 9) and the panel data allow us to identify peer effects at classroom level; in particular, we are able to overcome difficulties that have hindered the separation of peer effects from omitted individual factors due to self-selection and from common teacher effects. First, students are assigned to a class at entry of middle school (Grade 7) and stay with their classmates together for all subjects and for all grades in middle school. In other words, any self-selection into a class occurs before the interaction with classmates. Thus, individual fixed effects can capture all omitted student and family characteristics relevant for selection. Second, each teacher of Math, English, and Chinese teaches two classes and stay with the same two classes from Grade 7 to Grade 9. This panel nature allows us to use teacher by test fixed effects to capture the time-varying common teacher effect. We estimate peer effects for Math, English, and Chinese test scores separately. In a linear-in-means model controlling for both individual and teacher-by-test fixed effects, peers are found to have a positive and significant effect on math test scores and a positive but insignificant effect on Chinese test scores, but no effect on English test scores. Additionally, students at the middle of the ability distribution tend to benefit from better peers, whereas students at both ends do not.

[1] Carman: k.g.carman@uvt.nl, Tilburg University. Zhang: Corresponding author, email: zlei@clemson.edu, Clemson University and Hoover Institution, Stanford University. We are grateful to Raj Chetty, Giacomo De Giorgi, Eric Hanushek, Han Hong, Caroline Hoxby, Luigi Pistaferri, and seminar participants at the Stanford University, University of Oregon, and AEFA for helpful discussion and comments.

1. Introduction

Peer effects in schools have figured importantly in recent research on school production and in educational policy debates. The existence and structure of peer effects may have important implications for policies regarding ability tracking, classroom organization, and school choice, to name just a few. Empirical studies of peer effects in the school context face the same three challenges as in all other peer effects analysis: proper definition of a peer group, omitted variable bias due to self-selection into a group and common teacher effects that affect all members of a group (correlated effects), and the reflection problem (Manski 1993). Careful identification of peer effects requires detailed data that can address these identification problems.

This paper takes advantage of the unique features of the organization of Chinese middle schools (Grades 7 to 9) and a panel data set to identify peer effects at classroom level; in particular, we are able to overcome difficulties that have hindered the separation of peer effects from omitted individual factors due to self-selection and from common teacher effects. The two unique features are: (1) Students are assigned to a class at entry into middle school (Grade 7) and stay with the same classmates for all subjects and for all grades in middle school. In other words, any self-selection into a class occurs before the interaction with classmates. Thus, individual fixed effects can capture all omitted student and family characteristics relevant for selection. (2) Each teacher of Math, English, and Chinese teaches two classes and stay with the same two classes from Grade 7 to Grade 9. This panel nature allows us to use teacher by test fixed effects to capture time-varying common teacher effects. An additional advantage is that the data set follows all of the students in a single cohort, and hence the peer group includes all of the students one interacts with in a classroom.

We address the reflection problem, i.e., that student and peer achievement are determined simultaneously, through the use of lagged peer achievement. Specifications based on lagged peer achievement eliminate the problem of simultaneous equations bias and capture the systematic predetermined aspects of peer interactions. Because individual fixed effects pick up both the individual fixed characteristics and the fixed aspects of their environment, including their peers' fixed characteristics such as ability, the variation that remains is due to small changes over time in peer characteristics and behavior. Therefore, the coefficient on peer achievement will pick up the effect of transient changes in the behavior or underlying characteristics of peers.

We estimate peer effects for Math, English, and Chinese test scores separately. In a linear-in-means model controlling for both individual and teacher by test fixed effects, we find that peers have a positive and significant effect on math test scores, a positive but insignificant effect on Chinese test scores, and no effect on English test scores. When peer effects are allowed to vary by individuals' ability, we find that students in the two mid quartiles of ability distribution tend to benefit from better peers in Math and Chinese, and that the lack of average peer effect in English is largely due to the negative and significant influence of peers on the bottom quartile – students in the second quartile indeed benefit significantly from better peers. Because the estimates are identified from transient changes in peer achievement, we conjecture that one major source of the effects of peers is the pressure in a highly competitive environment.

There is a growing literature studying the effects of peers in schools, most of which considers schools in the United States. These studies focus on peer interactions either at the classroom level (Hoxby and Weingarth 2005, Burke and Sass 2006, Cooley 2007) or, due to data limitation, at the grade level (Hanushek et al. 2003, Hoxby 2000). Since students tend to interact more with other students in the same class than with students in the same grade, research based on the former tend to find larger peer effects than the latter. All of these studies attempt to deal with the concern that measures of peer achievement may be proxies for omitted or mismeasured student, family, teacher and school characteristics that affect individual achievement. A common identification strategy is a fixed effect model – using individual fixed effects to address omitted variable bias due to self-selection into a school or a classroom (all of the above papers except for Cooley 2007 where some individual characteristics are controlled for) and using teacher or grade fixed effects to address common teacher influences. However, fixed effect models do not adequately address the omitted variable problems in the U.S. school context. First, using individual fixed effects to address the self-selection problem assumes that selection is based on pre-determined achievement and other fixed characteristics. In U.S. schools grade and class compositions change every year due to school transfer and class reassignment, and students self-select multiple times with each selection based on time-varying individual or family characteristics. Second, using fixed effects to address common teacher influences assumes that teacher effect does not change over time. Studies of U.S. schools usually use longitudinal data where multiple cohorts are taught by the same teacher in different calendar years. Peer effect estimated in these models will still be contaminated by time-varying teacher influences. Our data allow us to address both issues because classroom composition and teacher assignment change less frequently.

A number of papers have considered peer effects in schools of other countries. Zimmer and Toma (2000) and Ammermueller and Pischke (2006) focus on primary schools in various European countries and Canada, but their identification strategy is not clear. The only other paper that studies peer effects making use of the special organizational feature of Chinese schools is Ding and Lehrer (2007). They study peer effects among high school students (Grades 10 to 12) using data from a number of high schools from one county. They argue that selection into high school is completely based on high school entrance test score, which is observed and controlled for. They also control for various teacher characteristics averaged over teachers of the entire school. The downsides of their paper are that the peer effects are measured at the grade level instead of the classroom level and that their sample includes high school graduates who have been accepted by a college so that their peer group is a much smaller group than the actual peer group that students interact with in high schools.[2]

2. Background of China's Middle School and Data

There are two periods in China's secondary education: middle school (Grades 7 to 9) and high school (Grades 10 to 12). This research uses data obtained from a typical middle school in the capital city of a North China province. This middle school is also highly representative of middle schools in urban China.[3]

Starting in the early 1990s, middle schools in China were required to abandon ability tracking and to accept any students finishing elementary school in their districts. There are usually several middle schools and a number of elementary schools in each district. Each elementary school is responsible for randomly assigning its students to the middle schools in the district. Once assigned to a middle school, students will be randomly assigned to a class. Classes stay together for all subjects and for all grades in middle school. Middle school principals work

[2] There are also studies focusing on peer effects in college. Sacerdote (2001) and Zimmerman (2003) consider peer effects among students living in the same dorm in Dartmouth College and Williams College respectively, and they argues that the dorm assignment is random. More recently, Carrell et al. (2007) and De Giorgi (2007) make use of special organization features of the US Air Force Academy and Bocconi University of Italy respectively to identify peer effects in better-defined peer groups. They find significant and larger effects of peers than Zimmerman.

[3] The city in question has an urban population of 2 million and covers an area of close to 7,000 square kilometers. It is divided into 6 districts, each of which covers a large geographical area. Unlike the United States, most of the best schools are located in big cities.

very hard to achieve randomness and comparable student quality across classes. This random assignment allows an objective evaluation of teacher performance. Teacher bonuses and promotions are to a large extent based on how well their students perform on midterms, finals, and eventually the high school entrance exam.[4] At the end of 9th grade, students take a city-wide high school entrance exam, the total score of which determines a student's eligibility for various tiers of high school. The quality or reputation of a middle school is ultimately measured by the percentage of its students eligible for admission into the top-ranked high schools in the city. Working hard in middle school to secure a seat in a top-ranked high school is crucial for college admission – in some urban areas, close to 100% of graduates from the best high schools are admitted to a college, whereas less than 10% from some lesser high schools are admitted. Additionally, academically successful students are well respected among students and friends.

Four core subjects are taught in all three years of middle school: Math, Chinese, English, and Social Science. Physics is taught in the 8th and 9th grades; chemistry in the 9th grade; history, geography and biology are taught in one or two years of the middle school. Students are given a midterm and a final exam every semester on all subjects taught in that semester. Tests are written collectively by all of the teachers for that subject and grade in the school; so tests are closely related to the materials taught in lectures. All students take the same tests, and hence the test scores are perfectly comparable across students. All schools in the city follow the same curriculum. A city-wide high school entrance exam is given at the end of the 9th grade in early July, which tests all subjects but geography.

A student remains in the same class with virtually the same classmates for all three years of middle school. Students cannot transfer between classrooms; there are very few transfers between schools. Each class has a unique head teacher, who is responsible for organizing sports events and other activities. The head teacher usually stays with the same class for all three years. The head teacher also teaches a subject. Subject teachers usually teach the same classes for the entire three-year period of middle school. Each teacher in Math, Chinese, and English teaches two classes at the same time. Other teachers may teach more classes at the same time.

[4] Selection at both the middle school assignment and class assignment in middle school is inevitable. Parents with connections who are not satisfied with their children's assignment can find ways to have it changed. We further address the possibility of non-random class assignment below.

We construct a panel of test scores for all 923 students who entered the sample middle school in the fall of 2003.[5] They are assigned to 16 classes with class size ranging from 51 to 65. We focus on Math, Chinese, and English test scores because these subjects are taught in all three years and allow the longest time series. For each subject we obtain test scores of 5 exams during middle school: midterm exams two months into middle school, final exams at the end of Grade 7, final exams at the end of Grade 8, final exams of the autumn semester of Grade 9, and the high school entrance exam. The data do not contain other student characteristics besides test scores. Class size varies slightly over time due to students transferring in or out of school. Appendix Table 1 reports the initial class size and the number of students that transfer in and out of each class – at most 3% of students in a class transfer in or out of school by each exam. The student test score data are matched with administrative data about teachers, which contain information about each teacher's education, experience, rank, and class assignment. At school entry, all but one head teacher teaches Math, Chinese or English; the remaining head teacher is a biology teacher who does not teacher her class in the 9th grade. Over the three years, only two head teacher are replaced (Class 5 in Grade 8 and Class 15 in Grade 9). The majority of the math, Chinese, and English teachers teach the same two classes all three years, but there are a few changes: one in Math (Classes 15 and 16 in Grade 8), one in English (Class 8 in Grade 8), and four in Chinese (Classes 5 and 6 in Grade 8, Classes 11 and 12 in Grade 8, Classes 9 and 10 in Grade 9, Class 15 in Grade 9).

Because we do not have an admission or diagnostic test score at entry to middle school, we cannot directly test the random class assignment assumption. As an approximation, we examine the test scores from the first exam in our sample, two months into the middle school. The mean test score is presented in Figure 1, along with means for later tests.[6] In the figure, classes taught by the same teachers are adjacent, for example group 1 consists of classes 1 and 2, and group 2 consists of classes 3 and 4. For both mean and variance in all three subjects, the pairs of classes taught by the same teachers are, in all but a few cases, statistically indistinguishable, whereas pairs of classes taught by different teachers are in general statistically different from each other. These results could mean two things: (a) class assignment is random, teacher and peer effects happen quickly, and a teacher has same effect on the two classes she

[5] We lose less than 2% of students who transfer out of the sample school at various points of the middle school. Students who transfer in are included in the data set.
[6] We present only the graphical representation in the paper. Tables are available from the authors.

taught, but different teachers have different effects; or (b) there is selection into classes, and two classes taught by the same teacher are considered as the same by parents. These two hypotheses are not distinguishable with the data, and non-random class assignment has to be dealt with.[7]

From the above description, two unique features of the organization of the Chinese middle school stand out. First, students are assigned to a class at entry of middle school (Grade 7) and stay with their classmates together for all subjects and for all grades in middle school. Second, virtually every teacher of Math, English, and Chinese teaches two classes and stay with the same two classes from Grade 7 to Grade 9. These combined with the panel nature of the data allow us to deal particularly well with the omitted variable bias in peer effect identification due to student self-selection into a class and teacher common effect, which will be discussed in detail in the next section.

3. Empirical Strategy

This research focuses on the reduced form relationship between a student's performance and measures of peer quality. We first consider the linear-in-means model of peer effects described in Manski (1993) and Brock and Durlauf (2001); models of other functional forms face the same identification issues. In this model, an individual *i* is a member of class *j* taught by teacher *k* and is observed at time *t*. An individual's test score, A_{ijkt}, is described by the following equation:

$$A_{ijkt} = \alpha_t + \beta\overline{A}_{-i,jkt} + \lambda_{kt} + \delta_i + \varepsilon_{ijkt}. \qquad (1)$$

Time specific fixed effects, capturing for example test characteristics, test conditions and school policy changes, are represented by α_t. The performance of a student's peers are expressed by the average score of the other students in class *j*, $\overline{A}_{-i,jkt}$, where –*i* indicates that the average is calculated excluding individual *i*. The coefficient on peer performance, β, capture the peer

[7] However, we can also see that the two classes in groups 3, 5, 8 have significantly different math score in test 1, and the difference becomes smaller for group 8 but not for groups 3 and 5.This suggests that something else is happening for these groups; i.e., there might be other omitted variables even after we have controlled for individual and teacher fixed effects. This may be related to teacher changes as described above. We discuss the issue of teacher changes further in section 4.

effects. Teacher fixed effects are captured by λ_{kt} and can vary over time.[8] Finally individuals are assumed to have some constant unobservable characteristics that are captured by individual fixed effects, δ_i . These individual fixed effects capture all permanent characteristics of the students, such as their gender, year of birth, family background characteristics, and subject specific talent.

The unique features of the Chinese middle school and the panel data set allow us to deal with one of the most difficult issues – the omitted variable bias – in peer effect identification. First, class formation may be subject to school policy and parents' choice. Even though the sample school works hard to randomly assign students to different classes, parents who care about their children's education and have connections can still succeed in getting their children reassigned to classes taught by teachers with a reputation as better teachers.[9] In the presence of self-selection, any measure of peer quality also captures, for example, the unobserved characteristics of parents such as the attention they paid to children's education and resources they have to obtain information about teachers and to change their children's assignment. These characteristics are likely to affect students' performance through family inputs. In the data, students stay in the same class for virtually all three years once they are assigned at the entry of middle school. While the characteristics associated with selection may change over time, what matters for selection are the initial characteristics. Because selection into classes is determined prior to the first exam, the inclusion of individual fixed effects will address the relevant omitted variable problem. Once individual fixed effects are included, all variables that do not vary over time will be captured in the individual fixed effect and will drop out of the model, as illustrated by the following equation:

$$\delta_i = x_i + \mu_{ij} + \varphi \cdot \bar{x}_{ij} .$$

This fixed effect will capture constant unobserved individual characteristics unrelated and related to selection, x_i and μ_{ij} , and constant peer characteristics $\bar{x}_{ij}$.

Second, similarity in outcomes of students in a class could be a result of common unobserved classroom inputs, most importantly, teacher quality. Common teacher effects complicate the peer effect estimation even in the absence of student self-selection into a class,

[8] Our ability to separate teacher-by-test effects and peer effects requires the assumption that a teacher has identical effects on the two classes she teaches.

[9] Parents' tend to select classes based on the quality of Math or English teacher, and quality of Chinese teacher is of less concern. Quality of Physics and Chemistry teachers are also important, but the two subjects are not taught in Grade 7.

and disentangling teacher effects from peer effects is usually hampered by data limitations. Either teachers are not matched with students, or their effects are assumed to be constant over time. In the data, virtually every teacher of Math, English, and Chinese teaches two classes of the grade and stays with the same two classes for the entire three years of middle school; therefore, for each teacher there are two separate peer groups. The panel data allow us to use teacher by test fixed effects to separately identify the time-varying teacher effects, λ_{jt}, common to all students in a class and peer effects, β.

The fixed effects model does not address one additional issue in peer influence estimation, the possible simultaneous determination of achievement for all classmates, with high achievement of one student directly improving the achievement of classmates and vice versa. This is the reflection problem as discussed in detail by Manski (1993). As Moffit (2001) shows, this situation can be thought of as a standard simultaneous equation problem, where the derived correlation between $\overline{A}_{-i,jkt}$ and ε_{ijkt} leads to inconsistent estimate of peer effect. We address the reflection problem through the use of lagged peer achievement. By using the lagged value of peer achievement to explain current individual achievement, we remove the mechanical simultaneity between individual and peer behavior. After including fixed effects, the transient variation in peer behavior and characteristics remains. Therefore, the coefficient on lagged peer achievement measures the effects of changes in peer characteristics and behavior.

One concern is that the reflection problem may remain even after using lagged values of peer behavior, because the individual in question affected their peers in the previous period. Including individual fixed effects may help to lessen this concern, because it allows us to control for the time invariant impact that classmates have on each other. Fixed effects will not adequately address this problem when there are significant trends in individual performance; in this case an individual’s past performance is a good predictor of the performance in the next period.

To address this problem we ran regressions for each individual that calculate individual trends. If for most people there is no trend this weakens the correlation between an individual’s lagged achievement and their current achievement which is one of the main concerns with regards to the reflection problem. The results suggest that between 10 and 20% of students have trends that are significant at the 10% level, this is slightly more than would be expected if there are actually no trends, but does not reflect a large proportion of our sample. Therefore we conclude that while there are some trends they are unlikely to overwhelm our results.

Table 1 reports the summary statistics of lagged peer achievement, both average and standard deviation. Test scores on each subject are normalized to have mean 0 and standard deviation 1 within each test. The standard deviations of average lagged peer achievement are 0.155, 0.167, and 0.214 for English, Chinese and Math respectively; therefore, there is sufficient variation in peer achievement in the sample. Similarly, there is also reasonable variation in standard deviation of peer achievement.

Table 2 reports the means and standard deviations of peer achievement in the first test of Math, English, and Chinese for each class separately. The group numbers indicate that classes 1 and 2 are taught by the same set of Math, English, and Chinese teachers, and so forth. Within each group, there is generally a significant difference between the two classes in both the mean and variance of the average peer achievement in all three subjects. This difference suggests that the two classes taught by the same teacher are far from identical in peer achievement; therefore, each teacher does have two separate and distinct peer groups, and fixed effects will be effective in separating teacher influences and peer influences. Significant differences in the standard deviations of peer achievement in all three subjects are also found between the two classes within each group.[10]

One complexity of peer effect estimation in the education context is that cumulative historical inputs affect current achievement. Historical inputs bias peer effect estimates only when they are correlated with lagged values of peer achievement. The uniqueness of the data implies that the use of the fixed effects eliminates most of the historical input variables. *First*, students are randomly assigned to the sample middle school from a number of elementary schools; thus, any historical inputs before entering the middle school are uncorrelated with peer achievement. *Second*, because student transfer is rare and teachers stay with the same classes for the entire middle school period, student and teacher-by-test fixed effects are likely to account for most of the unmeasured student and teacher inputs that systematically enter the education

[10] Table 2 presents the average and standard deviation of the peer achievement variable, while Figure 1 presents the average of individual performance. The results in Figure 1 and Table 2 suggest that while classroom composition is similar, there are significant differences in the peer environment that students face. This is possible because while the average of individual scores and the average of peer achievement are identical, the standard deviation for peer achievement is much smaller.

production process. Any remaining time-varying inputs are unlikely to be much correlated with the lagged peer achievement.

4. Effects of Peers

We present the effects of peers for Math, Chinese, and English tests separately. We first focus on the linear-in-means model; then we will allow peers to have different effects based on an individual's ability.

4.1 Average Peer Effects

Table 3A presents the average peer effects for different subjects estimated from a linear-in-means model, where peer quality is measured by the average of lagged peer achievement in each subject. All the specifications include test fixed effects. The first column is the baseline model without individual or teacher fixed effects. There is a positive and significant relationship between average peer achievement and one's math test score, a positive and marginally significant relationship for one's Chinese test score, but no effect for one's English test score. Notice that the standard deviation of the coefficient estimates are large. There are two reasons. First, the sample size is small – there are only 16 peer groups. Second, each peer group is big – the class size ranges from 51 to 65. Therefore, the variances of the average of peer achievement are quite small, as seen in Table 1. Thus, one should be cautious in interpreting the results, especially in concluding the lack of peer effects.

Column 2 adds individual fixed effects. Coefficient estimates on all the peer quality measures diminish. This is expected and suggests that part of the peer effects estimated in Column 1 is due to the unmeasured time-invariant individual, family or teacher characteristics that are correlated with both peer quality and own achievement. Constant teacher influence is taken away because individual fixed effects consist of all constant factors for students over all three years including teachers that stay with them. In particular, the magnitude of the effect of average peer achievement for Math drops significantly, but the effect is still significant at the 5% level. There is no significant effect for peer's Chinese or English achievement.

The specification in Column 3 controls additionally for teacher-by-test fixed effects to remove the time-varying teacher effects. This is our preferred specification. Average peer achievement in Math has a positive and significant (at 5% level) effect on one's Math test score. A 0.10 standard increase in average Math achievement of peers increases one's Math test score by 0.04 standard deviations. [11] Peers' English and Chinese achievement still do not appear to influence own performance in English and Chinese respectively.

The magnitude of the estimate for Math is larger than that for Math from Hoxby (2000) and Hanushek et al. (2003). This is in part due to the fact that their peer groups are defined at grade level, whereas here it is defined at the classroom level. It is reasonable to believe that students interact more with their classmates than with students outside their classroom.[12] The magnitude of our estimate for Chinese is lower and less significant than findings of Cooley (2007), where, controlling for teacher fixed effects, a 0.1 standard deviation increase in average reading achievement of peers in the class increases one's reading test score by 0.048 standard deviations based on a sample of 3rd to 8th graders. Because she uses contemporaneous peer achievement as a measure of peer behavior, her estimates may capture some important classroom interaction in learning that is not captured by the lagged peer achievement as in the present paper, and this may explain her slightly higher estimate.

Comparing Columns 2 and 3, the estimates on peer achievement measures increase after teacher-by-test fixed effects are controlled for, and the increase is considerable for Math and Chinese. This increase, however, should not be interpreted to mean that teachers have a negative effect on student achievement. With individual fixed effect, estimates in column 2 are already purged of constant teacher effects. Nevertheless, in column 2 coefficients on peer achievement still capture not only peer effects but also time varying environmental factors such as time-varying teacher effects that affect all students in a class. Larger estimates in Column 3 could suggest that teacher effects on student achievement vary around its mean, and on average the deviation is negative.

Table 3B reports results from a robustness test. In this panel, we restrict our sample to classes that remain with the same teacher for all 3 years.[13] Although most teachers stay with their classes over all three years, there are occasional reassignments of teachers. These could be due to health condition such as maternity leave or to unsatisfactory performance. Teacher reassignment

[11] We discuss these effects in terms of a 0.1 standard deviation increase because a 1 standard deviation increase in peer average behavior is essentially impossible.
[12] Indeed, Hanushek et al. (2003) find that peer effects are stronger for the subsample of students that stay in a school than for the entire sample and the subsample of students that transfer schools, suggesting that more interaction leads to larger peer influence.
[13] Classes 5, 6, 15, and 16 are excluded for all subjects due to changes in head teacher assignments. For English classes 7 and 8 are also dropped due to a change in the English teacher. For Chinese, classes 9, 10, 11, and 12 are also dropped due to a change in the Chinese teachers. For Math no additional classes are dropped.

may affect the estimates in Column 2 because we rely on the assumption that the same teacher stays with a class for all three years to remove constant teacher effect. With Table 3B we address the possibility that new teachers may influence students' achievement in a different way, which may contaminate the peer effect estimate in Column 2. By focusing on classes that keep the same teachers for all 3 years, we are able to satisfy all of the identifying assumptions.[14] In Table 3B, for Math the coefficient in column 2 is much larger than the corresponding estimate in Table 3A, while the coefficient in Column 3 is quite similar to that in Table 3A. Additionally, in Table 3B, estimate in Column 2 is larger than that in Column 3, suggesting that time-varying teacher effects are positive. Thus, peer effects for math are robust to changes in teacher assignment. We still find no evidence of peer effects for English or Chinese.

As another robustness test, we perform the following thought experiment: assume that the effects we estimate have nothing to do with interaction between classmates and are generated by mere sample variation (or measurement error or anything else). We then construct placebo peer groups by randomly assigning half class of students to another class taught by the same teacher – this allows us to continue to control for the common teacher effect. We expect to find no significant effect when the classes are formed through this randomization. We focus on Math and Chinese. In fact, in a fixed effect model controlling for both individual and teacher fixed effects, the coefficient estimate on lagged peer achievement is only half the magnitude of the estimates in Column 3 of Table 3A, and neither is significantly different from 0.

Our preferred specification controls for individual and teacher-by-test fixed effects; thus the peer effect is identified off of the variation in peer achievement over time. One potential concern is that removing all between individual variation reduces the ratio of signal to noise by leaving too little actual variation in peer attributes. Table 4 reports variances of peer achievement (average and standard deviation) before and after removing the individual and teacher by test fixed effects. The ratio between the original variance and the variances after removing all fixed effects ranges between 4% in average Math achievement to 15% in the standard deviation of English achievement. Take average Math achievement as an example, this means that a one standard deviation change in the residual roughly equals 0.04 standard deviations of the original test score distribution. Therefore, removing individual and teacher fixed effects still leaves

[14] The estimates of the coefficient on peer behavior in Column 3 are unlikely to be affected by the exclusion of classes with teacher changes because the teacher-by-test fixed effects should capture any change in teacher assignment.

adequate variation in peer group quality, and the pattern of estimates do not support a simple measurement error explanation.

Our data do not allow us to pinpoint the precise structure of underlying causal relationship. There are many possible channels through which peers affect one's performance: Peers are sources of motivation (for both students themselves and parents), aspiration, and direct interactions in learning. Peers may affect the classroom process – adding learning through questions and answers, contributing to the pace of instruction, or hindering learning through disruptive behavior á la Lazear (2001). Because our estimates are identified from transient changes in peer achievement, we conjecture that one major source of the effects of peers is the pressure or motivation from having high-achieving classmates. In the sample school, a teacher-parent meeting is held after each final exam in each class. During the meeting, a spread sheet of test scores and ranks of all students (with student names) of the class is posted on the wall of the classroom. Teachers also emphasize the importance of making good progress in every stage of school. The pressure due to this public knowledge is enormous for both parents and students themselves, which could motivate some students to work hard to improve their performance in the next test.[15]

Table 5 considers several different specifications as robustness tests. First, in Columns 1 through 3, following Hanushek et al. (2003), we allow student test scores to be affected by the heterogeneity of her peers' achievement, measured by the standard deviation of lagged peer achievement. Heterogeneity of peer achievement does not appear to affect one's test score in any subject. Controlling for the standard deviation does not qualitatively change the coefficients on the average of peer achievement and the coefficients on the standard deviation are never significant. We still find significant peer effects for math in all three specifications and no significant peer effects for English. In our preferred specification, Column 3, we again find that a 0.10 standard increase in average Math achievement of peers increases one's Math test score by 0.037 standard deviations (this is 0.04 in the specification without controlling for the standard deviation of peer achievement). The only qualitatively different result is that we now find a significant relationship between the peers' average achievement in Chinese in the specification that includes both individual and teacher-by-test fixed effects, here a 0.10 standard increase in

[15] Similar teacher-parent meetings are held in virtually all schools in China, so it is likely that the peer pressure is strong in this highly-competitive school environment and society at large.

average Chinese achievement of peers increases one's Chinese score by 0.042 standard deviations.

In the last column of Table 5, we report results for a different dependent variable. The dependent variable for each subject is an indicator variable equal to 1 if a student's test score is above the median score over all students taking the test in that subject and equal to 0 otherwise. The reason to consider this indicator is that overall between one third and half of students in the sample middle school are able to do well enough in the high school entrance exam to be eligible for the top-tier high schools.[16] Therefore, an important signal of students' potential to both students and teachers is whether they can stay at or improve to the top half of the test score distribution. We estimate peer effects in a linear probability model controlling for all fixed effects. Again, better peers increases one's probability of being in the top half in Math and Chinese tests, but has no effect on English test. Additionally, larger dispersion of peers' Math and Chinese test score increase one's probability of being in the top half in Math and Chinese tests, whereas larger dispersion of peers' English test score decrease one's probability of being in the top half in English test. The sharp contrast between peer influence in English and peer influence in Math and Chinese, both here and from Column 3, suggests that there may be a critical difference in student interaction in learning a foreign language as opposed to learning other subjects.

4.2 Differences by Quartile

The results in Tables 3 and 5 reveal significant influence of peer average achievement on all students in Math, less in Chinese and none in English. However, peers may affect some students more than others depending on students' own initial ability. In view of the peer pressure as discussed above, top students may not feel as much pressure as other students in the class. To examine this possibility, we interact the mean and standard deviation of peer achievement with indicators for the student's position in the achievement distribution based on test score in the first test in Grade 7.

Table 6A reports peer effects estimated for students at different quartiles of initial ability distribution over the entire grade, while Table 6B reports estimates for students at different

---

[16] This is based on conversation with the school principal. In 2006, the cutoff score for admission to a high-ranked high school is 550, and 36% of students in the sample school scored higher than the cutoff score.

quartiles of initial ability distribution relative to classmates only. All fixed effects are included. In both specifications, standard deviation of peer achievement has no effect on students at any of the four points on ability distribution. The average of peer achievement however shows different patterns of influence for different subjects in both specifications. For Math and Chinese, students at the middle two quartiles benefit significantly from having better peers; students at the bottom quartile (quartile 1) also benefit, but the estimate is not precise; whereas students at the top quartile (quartile 4) are not affected by peer achievement at all. For English, students at the bottom of the distribution are significantly hurt by better peer achievement; students in the other quartiles benefit from better peers, but the estimates are not significant. The estimates for English also appear to explain the lack of peer effect in English from Tables 3 and 5 – the average conceals the heterogeneity of peer influence on different types of students. [17]

Although results in Tables 6A and 6B are broadly similar, they have different implications for assigning students across classes. For example, results in Table 6A suggest that if we move several top students and bottom students from a class with a lot of top students to one with few such that peer achievement of the former does not change, whereas peer achievement of the latter increases, then middle students in the first class will be not be hurt, but the middle students in the second class will benefit from the improved peer quality. In the mean time, however, this reallocation changes the relative ranking of students in the first class; results in Table 6B suggest that students that are moved from quartile 2 to quartile 1 or from quartile 3 to quartile 4 will no longer benefit from the high achievement classmates. More generally, results in Table 6A provide an argument against ability tracking within school, whereas results in Table 6B suggest that any reallocation of students across classes leads to gains of some students at the cost of others.

One hypothesis is that a student is more likely to be influenced by peers whose achievement is closer to her own, or more specifically, peers in her own quartile, as opposed to all her classmates. Under this hypothesis, it would not be surprising to find that if peer performance is measured by the average of the entire class then there are no significant peer effects for students in the top and bottom quartiles, as found in Table 5. To illustrate this,

[17] Although we cannot reject the hypothesis that the estimates in Tables 6A and 6B are identical for each subject, comparing results in the two tables suggest that better peers have a more positive effects on middle students relative to classmates than relative to all students in a grade. This is consistent with our interpretation that the estimated peer effects come from peer pressure – when one's ability is measured more accurately relative to the peer group (as in Table 6B), one may feel stronger peer pressure.

consider the effect of peer performance on the best student in the class. This student will get higher scores on the exam if the top quartile also does well, but will not be much affected by the performance of peers in the second and third quartile. But the average of peer performance for the whole class is more likely to reflect the average of peer performance in the second and third quartile than the average of peer performance in the top quartile.[18]

To address this hypothesis, we create peer achievement measures for classmates in each quartile of the grade-wise distribution in each subject. We find first that achievement of peers in the entire class is significantly different from achievement of peers from each quartile in a class, and the latter has considerably smaller variance than the former. Second, in a model where we examine whether a student is differentially affected by classmates in her own achievement quartile and in other quartiles, the coefficient estimates are all statistically insignificant, which may in part be due to the small variation in achievement of peers in each quartile. To sum up, the results do not provide support to the hypothesis that a student is more affected by peers in her own ability quartile than those in other quartiles, and the results in Table 5 do indicate that students in the middle are more likely to be influenced by peers.

5. Conclusions

This paper estimates peer effects on student achievement using a panel data set obtained from a middle school in China. The unique features of the organization of Chinese middle school (Grades 7 to 9) and the panel data allow us to identify peer effects at classroom level and to deal with the omitted variable bias in estimating peer effects due to student self-selection into classrooms and common teacher effects.

The two unique features employed in the analysis are: (1) Students are assigned to a class at entry of middle school (Grade 7) and stay with the same classmates for all subjects and for all grades in middle school. In other words, any self-selection into a class occurs before the interaction with classmates. (2) Each teacher of Math, English, and Chinese teaches two classes and stay with the same two classes from Grade 7 to Grade 9. Thus, individual fixed effects can

[18] Another way to approach this hypothesis would be to use the method proposed by Hoxby and Weingarth (2005). They propose to look at the effect of the proportion of students in each decile on individual performance, and to allow the effect to vary by the individual student's own initial decile. We modify this test by using quartiles rather than deciles. In this case, we are testing how students are affected by the makeup of their class not the performance of their peers in each quartile. These results found no relationship between the makeup of the class and student performance.

capture all omitted student and family characteristics relevant for selection, and teacher-by-test fixed effects can capture the time-varying common teacher effect.

Controlling for both individual and teacher fixed effects, we find positive and significant average peer effects in Math and sometimes for Chinese, but no effect in English. However, peer effects vary with one's position on the ability distribution based on the initial test score. In Math and Chinese, students in the middle of ability distribution benefit significantly from better peers, whereas students at the top and bottom do not. In English, students at the bottom are hurt by better peers, while other students all benefit from better peers but with different degrees of significance. There is no evidence that the heterogeneity of peer achievement affects individual test scores. We hypothesize that the estimated peer effects are likely to come from the peer pressure in China's highly competitive school environment.

References

Ammermueller, Andreas and Jorn-Steffen Pischke. Peer Effects in European Primary Schools: Evidence from PIRLS. NBER Working Paper No. 12180, April 2006.

Burke, Mary A. and Tim R. Sass. Classroom Peer Effects and Student Achievement. Unpublished manuscript, 2006.

Carrell, Scott E., Robert N. Gilchrist, Richard L. Fullerton, and James E. West. Peer and Leadership Effects in Academic and Athletic Performance. Unpublished manuscript, 2007.

Cooley, Jane. Desegregation and the Achievement Gap: Do Diverse Peers Help? Unpublished manuscript, University of Wisconsin-Madison, March 2007.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. Be as Careful of the Books You Read as of the Company You Keep: Evidence on Peer Effects in Educational Choices. IZA DP No. 2833, June 2007.

Ding, Weili and Steven F. Lehrer. Do Peers Affect Student Achievement in China's Secondary Schools? *Review of Economics and Statistics*, 89(2): 300-312, May 2007.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics*, 18: 527-544, 2003.

Hoxby, Caroline M. Peer Effects in the Classroom: Learning from Gender and Race Variation, NBER Working Paper No. 7867, 2000.

Hoxby, Caroline M. and Gretchen Weingarth. Taking Race out of the Equation: School Reassignment and the Structure of Peer Effects. Unpublished manuscript, 2005.

Lazear, Edward P. Educational Production. *Quarterly Journal of Economics*, 116(3): 777-803, 2001

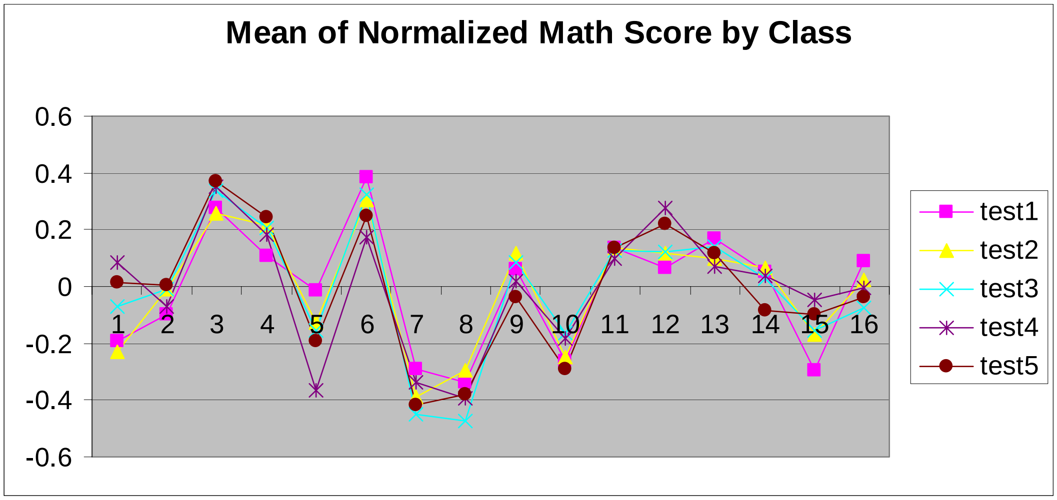
Manski, Charles. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economics Studies*. 60(3): 531-542, 1993.

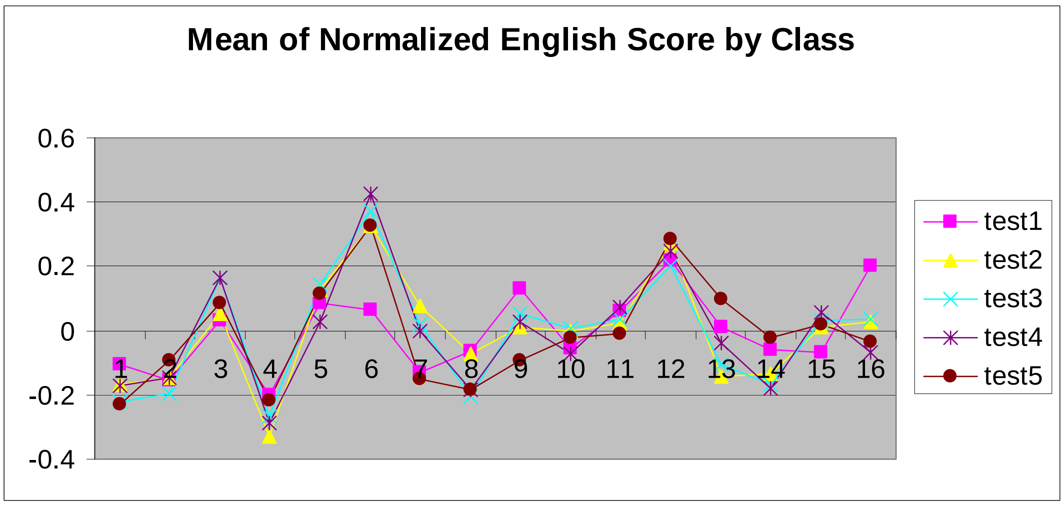
Sacerdote, Bruce. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics*, 116(2): 681-704, May 2001.

Zimmer, Ron W. and Eugenia F. Toma. Peer Effects in Private and Public Schools across Countries. *Journal of Policy Analysis and Management*, 19(1): 75-92, 2000.

Zimmerman, David J. Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics*. 85(1): 9-23, 2003

Figure 1: Comparison of Average Achievement across Classes





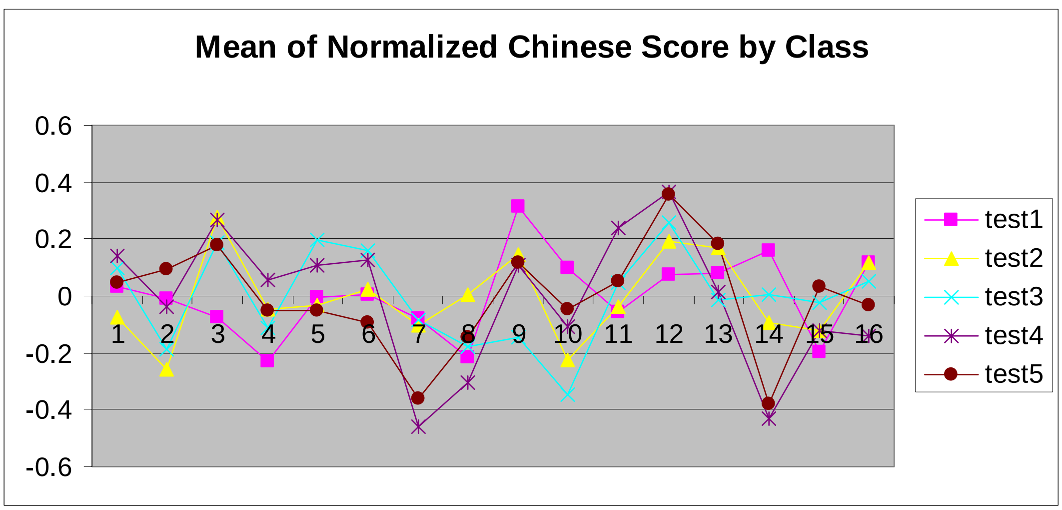


Table 1: Summary Statistics of Lagged Peer Achievement

Variable	Obs	Mean	Std. Dev.	Min	Max
Average of lagged peer Math score	3626	0.000	0.214	-0.497	0.427
St.D. of lagged peer Math score	3626	0.968	0.128	0.595	1.308
Average of lagged peer English score	3626	0.000	0.155	-0.325	0.480
St.D. of lagged peer English score	3626	0.976	0.153	0.470	1.351
Average of lagged peer Chinese score	3622	0.000	0.167	-0.491	0.452
St.D. of lagged peer Chinese score	3622	0.978	0.115	0.642	1.266

Table 2: Mean and Standard Deviation of Peer Achievement for Initial Test by Class

		Math		English		Chinese	
Group	Class	Mean	St.D.	Mean	St.D.	Mean	St.D.
1	1	-0.1926	0.0226	-0.1038	0.0213	0.0330	0.0188
	2	-0.0964	0.0170	-0.1562	0.0196	-0.0096	0.0162
2	3	0.2789	0.0144	0.0316	0.0169	-0.0744	0.0170
	4	0.1085	0.0188	-0.2027	0.0232	-0.2289	0.0230
3	5	-0.0125	0.0179	0.0841	0.0138	-0.0030	0.0176
	6	0.3845	0.0127	0.0633	0.0139	0.0031	0.0144
4	7	-0.2901	0.0212	-0.1295	0.0191	-0.0818	0.0199
	8	-0.3365	0.0217	-0.0635	0.0186	-0.2171	0.0169
5	9	0.0589	0.0164	0.1307	0.0159	0.3162	0.0149
	10	-0.2645	0.0186	-0.0541	0.0161	0.0990	0.0179
6	11	0.1356	0.0149	0.0598	0.0153	-0.0553	0.0145
	12	0.0641	0.0145	0.2169	0.0156	0.0747	0.0164
7	13	0.1697	0.0146	0.0120	0.0184	0.0808	0.0202
	14	0.0529	0.0193	-0.0617	0.0215	0.1580	0.0146
8	15	-0.2941	0.0212	-0.0678	0.0215	-0.1990	0.0241
	16	0.0893	0.0191	0.2013	0.0161	0.1174	0.0182

Table 3: Average Effects of Peers on Math, English, and Chinese

Table 3A: Entire Sample

	1	2	3
Dependent Variable: normalized Math score			
Average of lagged peer Math score	0.41	0.259	0.4
	[0.091]**	[0.112]*	[0.177]*
Dependent Variable: normalized English score			
Average of lagged peer English score	0.041	-0.045	-0.029
	[0.100]	[0.107]	[0.153]
Dependent Variable: normalized Chinese score			
Average of lagged peer Chinese score	0.127	0.104	0.259
	[0.074]+	[0.077]	[0.225]
Individual FE	No	Yes	Yes
Subject Teacher FE	No	No	Yes
Sample	all	all	all

Note: Robust standard errors clustered at individual level in brackets. + significant at 10% level; * significant at 5% level; ** significant at 1% level.

Table 3B: Restricted Sample

	1	2	3
Dependent Variable: normalized Math score [a]			
Average of lagged peer Math score	0.542	0.401	0.373
	[0.115]**	[0.148]**	[0.189]*
Dependent Variable: normalized English score [b]			
Average of lagged peer English score	-0.098	-0.284	-0.258
	[0.174]	[0.217]	[0.300]
Dependent Variable: normalized Chinese score [c]			
Average of lagged peer Chinese score	0.124	0.097	0.079
	[0.085]	[0.088]	[0.318]
Individual FE	No	Yes	Yes
Subject Teacher FE	No	No	Yes

Note: Robust standard errors clustered at individual level in brackets. + significant at 10% level; * significant at 5% level; ** significant at 1% level.

[a.] Classes 5, 6, 15 and 16 are dropped for the Math regressions.
[b.] Classes 5, 6, 7, 8, 15 and 16 are dropped for the English regressions.
[c.] Classes 5, 6, 9, 10, 11, 12, 15 and 16 are dropped for Chinese regressions.

Table 4: Residual Variance of Lagged Peer Performance Before and After Removing Individual and Teacher Fixed Effects

Variable	Variance before removing individual FE (1)	Variance after removing individual FE (2)	Variance after removing individual & teacher FE (3)	Ratio (2)/(1)	Ratio (3)/(1)
Average of lagged peer Math score	0.0457	0.0049	0.0021	0.11	0.05
St.D. of lagged peer Math score	0.0164	0.0042	0.0016	0.26	0.10
Average of lagged peer English score	0.0241	0.0039	0.0020	0.16	0.08
St.D. of lagged peer English score	0.0234	0.0073	0.0036	0.31	0.15
Average of lagged peer Chinese score	0.0279	0.0143	0.0024	0.51	0.09
St.D. of lagged peer Chinese score	0.0133	0.0081	0.0019	0.61	0.14

Table 5: Effects of the Average and Standard Deviation of Peer Achievement on Math, English, and Chinese

	1	2	3	4[a]
Dependent Variable:	Normalized Score	Normalized Score	Normalized Score	Above Median
Math:				
Average of lagged peer Math score	0.362	0.219	0.366	0.356
	[0.102]**	[0.120]+	[0.211]+	[0.161]*
St.D. of lagged peer Math score	-0.132	-0.129	-0.071	0.349
	[0.110]	[0.112]	[0.292]	[0.201]+
English:				
Average of lagged peer English score	-0.008	-0.086	0.083	-0.185
	[0.115]	[0.121]	[0.210]	[0.153]
St.D. of lagged peer English score	-0.073	-0.064	0.112	-0.228
	[0.079]	[0.080]	[0.150]	[0.105]*
Chinese:				
Average of lagged peer Chinese score	0.171	0.147	0.416	0.452
	[0.093]+	[0.097]	[0.246]+	[0.162]**
St.D. of lagged peer Chinese score	0.119	0.117	0.282	0.36
	[0.112]	[0.115]	[0.271]	[0.157]*
Individual FE	No	Yes	Yes	Yes
Subject Teacher FE	No	No	Yes	Yes
Sample	all	all	all	all

Note: Robust standard errors clustered at individual level in brackets. + significant at 10% level; * significant at 5% level; ** significant at 1% level.

a. In Column 4, the dependent variable is an indicator variable equal to 1 if a student's test score is above the median score over all students taking the test in each subject and equal to 0 otherwise.

Table 6: Effects of Peers on Math, English and Chinese by Quartile of Initial Performance

Table 6A: Ability Quartile over Entire Grade

	Math	English	Chinese
Average of lagged peer performance * Quartile 1	0.565	-0.784	0.28
	[0.387]	[0.386]*	[0.328]
Average of lagged peer performance * Quartile 2	0.407	0.472	0.759
	[0.267]	[0.282]+	[0.285]**
Average of lagged peer performance * Quartile 3	0.492	0.16	0.433
	[0.257]+	[0.268]	[0.270]
Average of lagged peer performance * Quartile 4	0.069	0.318	0.144
	[0.252]	[0.231]	[0.328]
St.D. of lagged peer performance * Quartile 1	-0.207	-0.053	0.092
	[0.395]	[0.277]	[0.360]
St.D. of lagged peer performance * Quartile 2	0.164	0.151	0.627
	[0.354]	[0.196]	[0.349]+
St.D. of lagged peer performance * Quartile 3	0.053	0.143	0.282
	[0.327]	[0.181]	[0.343]
St.D. of lagged peer performance * Quartile 4	-0.118	0.227	0.142
	[0.325]	[0.151]	[0.307]

Table 6B: Ability Quartile over Classmates

	Math	English	Chinese
Average of lagged peer performance * Quartile 1	0.41	-0.689	0.293
	[0.407]	[0.367]+	[0.319]
Average of lagged peer performance * Quartile 2	0.6	0.535	0.685
	[0.290]*	[0.294]+	[0.280]*
Average of lagged peer performance * Quartile 3	0.548	0.194	0.562
	[0.241]*	[0.246]	[0.272]*
Average of lagged peer performance * Quartile 4	-0.034	0.323	-0.03
	[0.247]	[0.251]	[0.324]
St.D. of lagged peer performance * Quartile 1	-0.093	-0.099	0.039
	[0.408]	[0.275]	[0.358]
St.D. of lagged peer performance * Quartile 2	-0.169	0.214	0.551
	[0.357]	[0.191]	[0.340]
St.D. of lagged peer performance * Quartile 3	0.235	0.181	0.323
	[0.327]	[0.172]	[0.335]
St.D. of lagged peer performance * Quartile 4	-0.233	0.189	0.049
	[0.319]	[0.159]	[0.309]

Robust standard errors clustered at individual level in brackets. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Individual and teacher by test fixed effects are included. Quartile 1 is the bottom of distribution; Quartile 4 is the top of distribution.

Appendix Table 1: Class Size and Number of Transfer Students

		1. G7, Fall, Midterm	2. G7, Spring, Final		3. G8, Spring, Final		4. G9, Fall, Final		5. G9, Spring, HS Entrance Exam	
class	group	initial classize	in	out	in	out	in	out	in	out
1	1	51	1	0	1	4	1	0	2	1
2	1	58	0	0	3	1	0	0	1	3
3	2	64	0	1	1	2	1	0	0	1
4	2	55	1	0	3	0	1	1	0	3
5	3	57	1	0	4	1	4	1	0	3
6	3	65	2	2	2	2	1	0	0	0
7	4	55	0	0	0	0	0	2	0	0
8	4	58	0	0	0	1	0	2	0	1
9	5	58	2	1	2	0	2	3	2	2
10	5	56	0	1	0	4	0	1	0	0
11	6	62	0	0	1	0	0	2	0	1
12	6	62	1	0	6	2	1	5	0	1
13	7	62	1	0	3	1	0	2	0	1
14	7	53	0	0	2	1	1	1	0	2
15	8	54	4	2	0	2	1	1	0	0
16	8	52	1	2	1	0	1	0	1	2