China Economic Review 23 (2012) 223–237

Contents lists available at SciVerse ScienceDirect

China Economic Review

Classroom peer effects and academic achievement: Evidence from a Chinese middle school[☆]

Katherine Grace CARMAN [a,*], Lei ZHANG [b,1]

[a] Department of Economics, Tilburg University, Netherlands
[b] National Institute for Fiscal Studies, Tsinghua University, PR China

ARTICLE INFO

Article history:
Received 12 April 2011
Received in revised form 26 August 2011
Accepted 31 October 2011
Available online 9 November 2011

JEL classification:
H40
I21
O15

Keywords:
Peer effects
Classroom
Fixed effects
Chinese schools

ABSTRACT

This paper estimates peer effects on student achievement using a panel data set from a middle school in China. Unique features of the organization of Chinese middle schools (Grades 7 to 9) and panel data allow us to overcome difficulties that have hindered the separation of peer effects from omitted individual factors due to self-selection and from common teacher effects and to identify peer effects at the classroom level. 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, a positive but insignificant effect on Chinese test scores, but no effect on English test scores. Importantly, in Math and Chinese students at the middle of the ability distribution tend to benefit from better peers, whereas students at the ends of the ability distribution do not, suggesting that policy makers who want to exploit positive peer effects face difficult tradeoffs in classroom and school assignment.

1. Introduction

Understanding the influence of peers is important for improving the design of socioeconomic policies. In education particularly, the existence and structure of peer effects may have substantial implications for policies regarding ability tracking, classroom organization, and school choice, to name just a few. This may be of particular importance for developing countries, where the limited financial resources call for more efficient allocation of educational inputs — with students' own ability and effort being among the most important of these inputs.

This paper estimates the influences of classroom peers on students' academic achievement in the context of a Chinese middle school (Grades 7 to 9), where several unique features of class organization and panel data allow us to overcome difficulties that have hindered the separation of peer effects from omitted individual factors due to self-selection and from common teacher effects and to identify peer effects at the classroom level. The unique ways classes and teaching are organized are as follows. First, students are assigned to a class at entry into middle school (Grade 7) according to a "balanced assignment" rule, such that the average initial quality of students is comparable across classes. Second, there is no re-assignment afterwards, and hence students

☆ We thank Raj Chetty, Giacomo De Giorgi, Eric Hanushek, Han Hong, Caroline Hoxby, Luigi Pistaferri, Petra Todd, and seminar participants at Beijing University, Stanford University, Tilburg University, the University of Oregon, Tsinghua University, and AEFA for helpful discussion and comments, and Michael Feng for excellent research assistance. We are grateful to Xiaofen Wang and Junyi Zhang for providing us the data. Lei Zhang thanks Hoover Institution of Stanford University for financial support in the form of W. Glenn Campbell and Rita Ricardo-Campbell National Fellowship. Part of the research was completed when Lei Zhang was an Assistant Professor of Economics at Clemson University.

* Corresponding author. Tel.: +31 13 466 8758.
E-mail addresses: k.g.carman@uvt.nl (K.G. Carman), zlei89@gmail.com (L. Zhang).
[1] Tel.: +86 10 6277313.

1043-951X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.chieco.2011.10.004

stay with the same classmates for all subjects and for all grades in middle school. Third, each teacher of Math, English, and Chinese teaches two classes and stays with the same two classes from Grade 7 to Grade 9. These characteristics combined with panel data allow us to use individual fixed effects to capture all time-invariant student, family, and classroom characteristics and teacher-by-test fixed effects to capture time-varying common teacher effects. 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 transient changes over time in peer characteristics and behavior.

Empirical studies of peer effects face three challenges: 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. We address the reflection problem, i.e., that student and peer achievements are determined simultaneously, through the use of lagged peer achievement. Specifications based on lagged peer achievement eliminate the problem of simultaneous equations. The coefficient estimate on the lagged peer achievement reflects a combination of contextual (peer characteristics) and endogenous (peer behavior) peer effects as defined by Manski (1993).[2]

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 middle quintiles of the ability distribution tend to benefit from better peers in Math and Chinese, while students at both ends do not; and that the lack of average peer effect in English is largely due to the negative and significant influence of peers on students in the bottom quintile. The fact that peer influence varies with students' own ability suggests that alternative classroom assignment rules may benefit some students at the cost of others. Because the estimates are identified from transient changes in peer achievement, we conjecture that one major mechanism of the effects of peers is the pressure felt by students in China's highly competitive school environment.

Most of the existing literature on school peer effects considers schools in the United States.[3] These studies focus on peer interactions either at the classroom level (Burke & Sass, 2011; Cooley, 2010; Hoxby & Weingarth, 2005) or, due to data limitations, at the grade level (Hanushek, Kain, Markman, & Rivkin, 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 tends to find larger peer effects than the latter. One common strategy 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 is a fixed effect model — using individual fixed effects to address omitted variable bias due to self-selection into a school or a classroom and using teacher or grade fixed effects to address common teacher influences. However, fixed effect models of this type 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,[4] and student self-selection occurs multiple times, with each selection potentially based on time-varying individual or family characteristics. Second, using fixed effects to address common teacher influences assumes that teacher effects do 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 effects estimated in these models will still be contaminated by time-varying teacher influences.[5]

The literature on peer effects in developed countries can help to inform the study of peer effects in less developed countries. However, cultural and institutional differences are likely to impact the importance of peer effects. Students in developing countries may react very differently to the performance of their peers than students in the United States. This is particularly likely to be true in China where relative performance is discussed openly and where there is significant social and familial pressure to be a top student. In addition, peer effects may operate differently due to the significantly larger class sizes common in Chinese middle schools. Class sizes typically are between 50 and 60.

This paper adds to a growing literature on the theoretical and empirical analysis of peer effects on student performance in a developing country context, for example, Glewwe (1997) of the Philippines, McEwan (2003) of Chile, Selod and Zenou (2003) of South Africa, Kang (2007) of South Korea, Duflo, Dupas, and Kremer (2011) of Kenya, and Ding and Lehrer (2007) and Lai (2008) of China.[6] In particular, Ding and Lehrer (2007) estimate grade-level peer effects among high school students (Grades 10–12) using data from a county in Jiangsu Province and find positive peer effects with the middle students benefiting the most from having high-quality peers. Lai (2008) finds positive classroom peer effects using panel data from middle schools of a Beijing school district; the drawback

[2] Contextual effects are also referred to as exogenous effects.

[3] Zimmer and Toma (2000) and Ammermueller and Pischke (2009) study primary schools in various European countries and Canada. They argue that classrooms are formed roughly randomly; the latter also shows the importance of correcting measurement errors.

[4] For example, according to Hanushek et al. (2003), in Texas public schools, annual mobility of students into or out of schools averages greater than 20% per year.

[5] 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 argue that the dorm assignment is random. More recently, Carrell, Gilchrist, Fullerton, and West (2009) and De Giorgi, Pellizzari, and Redaelli (2010) 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 Sacerdote and Zimmerman.

[6] Zhang (2010) uses data from Chinese middle schools to study the estimation bias in a specific type of model where own and peers' past achievements are included simultaneously. He shows that a positive correlation between measurement errors in own and peer achievement leads to a negative bias in the estimate of peer coefficient.

is that her data do not include matched teacher–classroom information. The similarity between their results and findings in the present paper suggests that the policy implications discussed here may be applicable to middle schools in urban China in general.

Methodologically, our paper differs from the previous literature because we have multiple observations within a given peer group over time; this is in sharp contrast to most data used in this literature that have one test score per year coupled with reassignment of classes each year. Our data allow us to better address both student self-selection and common teacher effects because classroom composition and teacher assignment change less frequently. Previous estimates of peer effects have assumed that teacher effects do not vary over time or with classroom composition; we relax this assumption by controlling for teacher effects that vary over time and for any constant effects related to classroom composition. Stated differently, previous research has relied on one observation of student achievement for each classroom and teacher assignment, we, on the other hand, have multiple observations of student achievement for each classroom and teacher assignment.

Section 2 describes the institutional background of the middle school that the data come from — the rule of the class assignment and the organization of teaching. Although middle schools in China can differ much in selectivity, the school organization described here is highly representative. Section 3 lays out the empirical strategy. We also provide summary statistics of the data in this section. Section 4 describes the regression results including sensitivity analysis. Section 5 concludes.

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 from a middle school in the capital city of a North China province. The organization of teaching assignments described below is quite typical of middle schools in urban China.[7] Our main analysis uses a panel of test scores for all 923 students who entered the sample middle school in the fall of 2003; they were assigned into 16 classes.

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. Since middle school education in China is compulsory, students attending their assigned schools do so free of charge. However, the practice of school choice is widespread, even though it is openly disapproved of by the government. With school choice, students can bypass the standard school assignment process and enroll in a different and usually better public school by paying a hefty tuition.[8] Thus, many high-quality middle schools are *de facto* hybrid public–private schools. Indeed, not all students willing to pay can enroll in the desired schools. These schools are usually over-subscribed, and they conduct their own admission tests. In general, only students passing a threshold score are admitted. The middle school studied here has approximately one-third tuition paying students and two-thirds public students, who are assigned to separate classes as described below.

At the beginning of middle school, diagnostic tests in Math and Chinese are administered to all students who have been admitted to the middle school, both public and self-pay. The total of these tests is used as the basis for class assignment. Students are assigned to classes via a procedure that limits differences in average test score across classes and balances sex ratios across classes. Because teachers' performance and hence bonuses and promotions are based on class performance, deviations from this assignment procedure are rare.

When the school assigns students to classes, self-pay and public students are segregated, but the same assignment procedure is used. Out of 16 classes, 5 (classes 3, 6, 11, 12, 13) are reserved for self-pay students. The self-pay classes are larger than the public classes, perhaps due to school capacity and the tendency to have even number of classes. Students are separated into 4 categories for class assignment based on gender and self-pay or public status. Students are then ranked within each category based on their diagnostic test scores. Within a category, students are assigned to balance average achievement across classes. The assignment process can be split into rounds. The class assigned the best student in any given round will also be assigned the worst student in that round. This ensures that the average rank for each class is the same after each round of assignment. To understand the process it is easiest to consider a simplified case with only 3 classes.

1. Starting with the top 3 students, the 1st student is assigned to class one, the 2nd to class two, and the 3rd to class three. With the next 3 students the order of class assignment is reversed: the 4th student is assigned to class three, the 5th to class two, and the 6th to class one.
2. With the next 3 students, the order of class assignment starts with the second class and proceeds sequentially: the 7th student is assigned to class two, the 8th to class three, and the 9th to class one. With the next 3 students the order is reversed: the 10th student is assigned to class one, the 11th to class three, and the 12th to class two.
3. With the next 3 students, the order of class assignment starts with the third class and proceeds sequentially: the 13th student is assigned to class three, the 14th to class one, and the 15th to class two. With the next 3 students the order is reversed: the 16th student is assigned to class two, the 17th to class one, and the 18th to class three.
4. Steps 1 through 3 repeat until all students are assigned

[7] The city in question has an urban population of 2 million and covers an area of close to 7000 km^2. 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 urban areas. During the sample period, urban public schools in general do not admit children of migrant workers from the rural area.

[8] The practice is considered by school administrators as necessary to supplement the meager government appropriation — teacher bonus and other benefits come usually from this source. In the mean time, the practice is condemned by the media for perpetuating inequality. Lai (2008) discusses school choice in Beijing.

Table 1
Summary statistics of lagged peer achievement.

Variable	Obs	Mean	St.D.	Min	Max
Mean of peer Math score	3594	0.000	0.215	−0.505	0.437
St.D. of peer Math score	3594	0.967	0.131	0.596	1.307
Mean of peer English score	3593	0.000	0.161	−0.348	0.477
St.D. of peer English score	3593	0.974	0.158	0.456	1.362
Mean of peer Chinese score	3590	0.000	0.173	−0.498	0.453
St.D. of peer Chinese score	3590	0.977	0.119	0.650	1.277

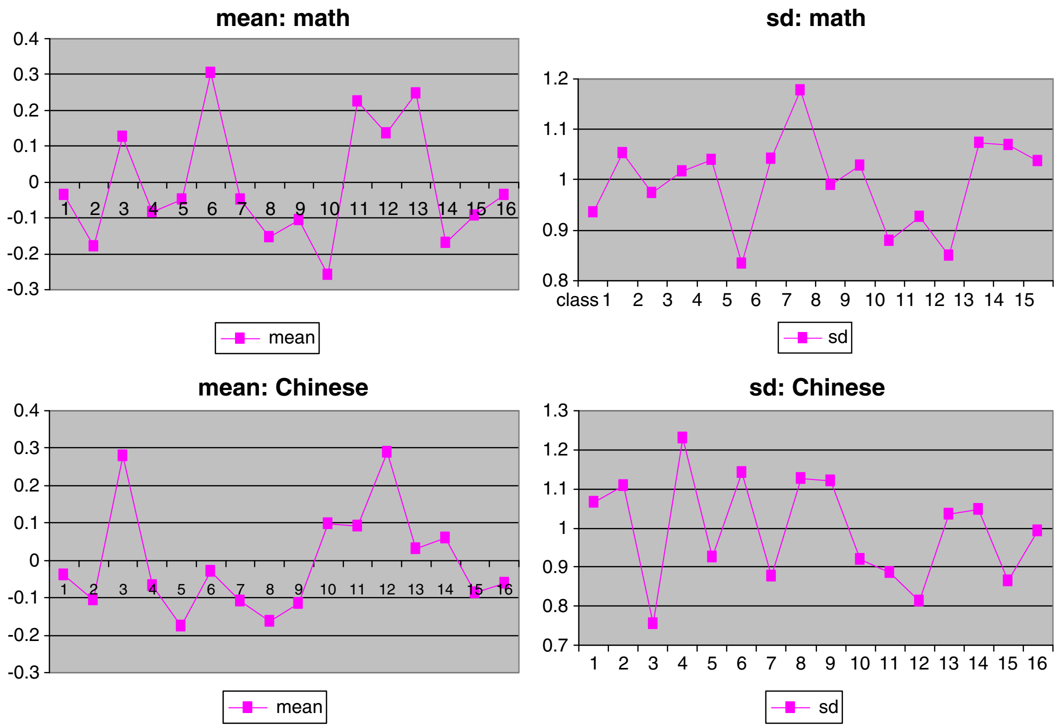

Fig. 1. Mean and standard deviation of diagnostic test score by class.

Appendix Table 1 illustrates the order of assignment for the case of the self-pay classes. The same procedure is used for the public classes, but there are more classes and therefore more students assigned per round.

After this assignment process takes place, it is possible that a very small number of students request to switch classes. The headmaster of the school reports that fewer than 5 students out of over 900 per year are allowed to do this. If this occurs, the school will try to switch that student with an equivalent student (based on diagnostic test score and gender) in the desired class. Finally, 4% of students do not have the diagnostic test score because they either are admitted after the tests or simply have missed the tests. These students are allocated to classes in a final round with equal numbers of boys and girls allocated to each class. Again a distinction is made between self-pay and public students.

Because we do not know the gender of students or whether the school changed the class number after assignment, we cannot replicate the assignment process. Nevertheless, we confirm that classes are indeed formed to achieve comparable distribution across classes. Within the self-pay and public categories respectively, Mann–Whitney tests confirm that the diagnostic test scores are all statistically equal between every pair of classes; however, the self-pay classes are significantly different from the public classes.[9] Fig. 1 illustrates that diagnostic test scores within the categories of public classes and self pay classes are similar. It also shows that self-pay classes have in general higher average diagnostic test scores and smaller variation than the public classes, reflecting the fact that self-pay students have to pass the admission test to be accepted. In essence, the school practices some degree of ability streaming (Glewwe, 1997) by separating self-pay and public students, but within each category, it mixes students of different abilities.

[9] These tables are available upon request.

One other middle school in the same school district is also able to admit tuition-paying students due to its good reputation; the remaining half dozen schools are pure public schools. Other school districts in the city are similar in this respect. Thus, students in hybrid schools have higher average ability than those in pure public schools. This hierarchical structure resembles the private–public school difference discussed in Epple and Romano (1998).

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 also very few transfers between schools — at most 3% of students in a class transfer in or out of school in each year (Appendix Table 2). Each class has a unique head teacher, who is responsible for organizing sport 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.

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. A city-wide high school entrance exam, which is closely related to the city-wide common curriculum, is given at the end of the 9th grade in early July and tests all subjects but geography, and its total score determines a student's eligibility for various tiers of high school. 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 students from the best high schools are admitted to a college, whereas less than 10% from some lesser high schools are admitted. Additionally, because academically successful students are well respected among students, friends, and relatives, students treat all exams seriously. As a result, test scores on “low-stake” exams such as the final exams at the end of school years, as well as test scores on the “high-stake” high school entrance exams, are equally good measures of what students can achieve at different stages of middle 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.

Our main analysis focuses 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 have 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. The student test score data are matched with data about each teacher's 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 teach her class in the 9th grade. Over the three years, only two head teachers 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).

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) according to a “balanced assignment” rule 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 stays with the same two classes from Grade 7 to Grade 9. These combined with the panel 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 common teacher effects, which will be discussed in detail in the next section.

Table 2
Mean and standard deviation of lagged peer achievement by class.

Group	Class	Math		English		Chinese	
		Mean	St.D.	Mean	St.D.	Mean	St.D.
1	1	−0.103	0.124	−0.165	0.046	0.051	0.083
	2	−0.045	0.042	−0.162	0.027	−0.124	0.103
2	3	0.305	0.041	0.103	0.064	0.163	0.144
	4	0.179	0.045	−0.273	0.051	−0.083	0.105
3	5	−0.169	0.131	0.095	0.048	0.069	0.092
	6	0.296	0.079	0.298	0.140	0.079	0.068
4	7	−0.365	0.063	−0.009	0.078	−0.180	0.160
	8	−0.374	0.070	−0.130	0.065	−0.173	0.113
5	9	0.069	0.040	0.054	0.049	0.105	0.166
	10	−0.217	0.047	−0.031	0.037	−0.140	0.169
6	11	0.124	0.021	0.048	0.026	0.047	0.117
	12	0.144	0.082	0.234	0.030	0.223	0.108
7	13	0.120	0.041	−0.068	0.062	0.061	0.073
	14	0.046	0.025	−0.137	0.052	−0.090	0.217
8	15	−0.165	0.090	0.009	0.050	−0.118	0.065
	16	0.007	0.063	0.048	0.098	0.035	0.108

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 similar 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:

$$A_{ijkt} = \alpha_t + \beta \bar{A}_{-i,jkt} + \lambda_{kt} + \delta_i + \varepsilon_{ijkt}. \tag{1}$$

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, $\bar{A}_{-i,jkt}$, where $-i$ indicates that the average is calculated excluding individual i. The coefficient on peer performance, β, captures the peer effects. Teacher fixed effects are captured by λ_{kt} and can vary over time. Finally individuals are assumed to have some constant unobservable characteristics that are captured by individual fixed effects, δ_i. ε_{ijkt} captures unobserved shocks that vary across time, class, and individual.

The unique features of the Chinese middle school and the panel data set allow us to deal with one of the most difficult issues in peer effect identification: omitted variable bias. Because students at entry are assigned into a class according to a "balanced assignment" rule based on a diagnostic test score with minimal deviations from the rule allowed, this process precludes students' self-selection into a class based on measured diagnostic test score and implies that selection based on unmeasured factors is minimal. Since students stay in the same class with virtually the same classmates for all three years, with the panel data, we have multiple observations within a given peer group and are able to include individual fixed effects in Eq. (1) to capture all permanent characteristics of the students and also any permanent characteristics of their class. An additional benefit of the individual fixed effects is that they can capture the remaining, very small amount of self-selection based on unobservables after the initial assignment process. For example, the few reassignments may be because these parents have information about certain teachers and have connections.[10] In this case, a measure of peer quality also captures the unobserved characteristics of parents, such as the attention they paid to children's education or the resources they have to obtain information about teachers and to change their children's assignment. Because selection into classes is determined prior to the first exam, what matters for this selection are the initial parental characteristics. Including 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 and will drop out of the model, as illustrated by the following equation:

$$\delta_i = x_i + \mu_{ij} + \phi \cdot \bar{x}_{ij}. \tag{2}$$

This fixed effect will capture constant unobserved individual characteristics unrelated and related to selection, x_i and μ_{ij} respectively, 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, and disentangling teacher effects from peer effects is usually hampered by data limitations. In previous papers, either teachers are not matched with students, or their effects are assumed to be constant both over time and across classes. In our 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 peer effects, β, and the time-varying teacher effects, λ_{jt}, that are common to all students in a class . We are able to relax the assumption of previous literature and only assume that time-varying teacher effects are common across classes.[11]

By including individual and teacher-by-time fixed effects, any remaining variation in peer performance is attributed to transient changes. Our identification strategy thus measures whether small changes in peer achievement in the previous period affect individual achievement. Controlling for selection and individual ability, if one's peers perform better in the previous period, how does that affect individual achievement? There are three issues that require further explanation: first, what causes these transient changes in achievement, second, can transient changes be used to identify peer effects, and third, what are the implications of focusing on the effect of past peer performance.

First, transient changes in achievement could be the result of changes that occur outside of the classroom during middle school, such as divorce, moving to a different neighborhood, or differences in the timing of puberty and growth spurts. Because student seating is rearranged each semester by height with the shortest students in the front, students that undergo a growth spurt will be moved to the back of the classroom. It is a common belief that, *ceteris paribus*, students in the front rows tend to receive more attention from teachers and tend to be more attentive during class, and hence may do better academically. Students at the front of the classroom may have a different impact on their peers, particularly if teachers focus more attention on these

[10] Middle school principals also try to balance the quality of teachers of a class; for example, if a class has a particularly good Math teacher, its English teacher may be a little weaker. This practice is largely due to the fact that the criterion to be admitted to a good high school is the total score of the high school entrance exam on all subjects. A student with superb math skill but poor English will not be admitted. Parents' tend to select classes based on the quality of Math or English teachers, and the quality of Chinese teachers is of less concern. The quality of Physics and Chemistry teachers is also important, but the two subjects are not taught in Grade 7.

[11] Our ability to separate teacher-by-test effects and peer effects requires the assumption that the time varying teacher effects on the two classes she teaches are identical. Our identification allows teacher effects to vary across the two classes as long as the differences are constant over time.

students. While in the long run height is determined by factors such as nutrition and genetics, the precise timing of growth spurs of children during puberty tends to be random. The periodical change in seat assignment due to changes in height can be considered as random and thus may be an important source of transient changes in achievement.

Second, we consider whether transient changes are likely to influence peers. For example, consider what happens if an above average peer undergoes a transient change that reduces their performance on exams, such as a growth spurt that moves them to the back of the class. Because the student was above average, they are likely to be replaced in the front of the classroom with a lower ability student. Students in the front of the classroom may be more likely to influence their peers because they may be called on more often and can be seen by a larger proportion of students in the classroom. Regardless of the source of changes, if the average understanding of the material presented in class declines, this is likely to influence the whole class. Ultimately, our identification strategy rests on the assumption that transient changes in peer understanding, not just permanent characteristics, are likely to influence individual achievement.

Third, we focus on the effect of past peer performance. The fixed effects model does not address the reflection problem (Manski, 1993). This refers to the possible simultaneous determination of one student's achievement and classmates' achievement such that the average achievement of classmates is correlated with one's unobservables ε_{ijkt} (Moffitt, 2001), leading to inconsistent estimation of peer effects. We address the reflection problem through the use of lagged peer achievement, which removes the mechanical simultaneity between individual and peer behavior. Given that learning is cumulative, achievement in the past may directly influence achievement in the future. Intuitively, if one's peers fail to grasp the concepts of seventh grade math, their understanding of eighth grade math may be limited, thus limiting the individual's test scores as well. Thus our use of past achievement assumes that past peer achievement contributes to current individual achievement. This does not eliminate the possibility that current peer achievement also makes a contribution, but we cannot test this in our data.

One concern is that the reflection problem may remain even after using lagged values of peer behavior, because the individual in question affected his peers in the previous period. Including 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 and the time-varying impact that a teacher has on all her students. 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 gauge the extent of 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 the main concern 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.

It is important to state that by using the lagged peer achievement we do not try to distinguish between contextual and endogenous peer effects; rather, we interpret the estimated peer effects as a combination of the two. In fact, Manski (1993) shows that it is not possible to separately identify these two effects. In the education context, the distinction between the two is less obvious given the available peer attribute measures. For example, average peer achievement may capture peer innate ability, a characteristic. But ability is naturally closely related to effort into school work, answering questions or disrupting learning in lectures, and helping or distracting fellow students in out-of-classroom study. Thus, the effect of our measure of lagged peer achievement on individual achievement can be interpreted as the effect of time varying peer behavior and characteristics.

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 0.161, 0.173, and 0.215 for English, Chinese and Math respectively; therefore, there appears to be 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 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. The Mann–Whitney test shows that there is generally a significant difference between the two classes within each group in both the average and the standard deviation of peer achievement in all three subjects. These differences suggest 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.

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 assigned to middle schools either via random assignment or by taking an admission test; 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 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, English, and Chinese 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.

Table 3
Average effects of peers on Math, English, Chinese, and total.

3A: all classes

	1	2	3	4
Dependent variable: normalized Math score				
Average of lagged peer Math score	0.41	0.259	0.4	0.515
	[0.091]**	[0.112]*	[0.177]*	[0.234]*
Dependent Variable: normalized English score				
Average of lagged peer English score	0.041	−0.045	−0.029	0.031
	[0.100]	[0.107]	[0.153]	[0.245]
Dependent variable: normalized Chinese score				
Average of lagged peer Chinese score	0.127	0.104	0.259	0.194
	[0.074]+	[0.077]	[0.225]	[0.320]
Number of observations	3594	3594	3594	1708
Test FE	Yes	Yes	Yes	Yes
Individual FE	No	Yes	Yes	Yes
Subject teacher-by-test FE	No	No	Yes	Yes
Sample	All	All	All	Public students

Note: Robust standard errors clustered at classroom level in brackets. + significant at 10% level; * significant at 5% level; ** significant at 1% level. For total score Group fixed effects are substituted for Subject Teacher fixed effects. Sample in Column 4 includes classes for which both classes in a group consist of public students (Classes 1, 2, 7, 8, 9, 10, 15, and 16).

3B: classes without teacher changes

	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]
Test FE	Yes	Yes	Yes
Individual FE	No	Yes	Yes
Subject teacher-by-test FE	No	No	Yes

Note: Robust standard errors clustered at classroom 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; 2693 observations.
[b] Classes 5, 6, 7, 8, 15 and 16 are dropped for the English regressions; 2260 observations.
[c] Classes 5, 6, 9, 10, 11, 12, 15 and 16 are dropped for Chinese regressions; 1783 observations.

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. [12] All the specifications include test fixed effects and report robust standard errors clustered at class level. 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 individual achievement for 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 deviations 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 relationship 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 influences are 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.

[12] The lack of variation in the diagnostic test score across classes makes it impossible to use it as a measure of pre-interaction peer characteristics and to estimate contextual peer effects at different points of the middle school period. Such regressions show very large standard errors for the peer effect estimates.

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 deviation increase in average Math achievement of peers increases one's Math test score by 0.04 standard deviations.[13] Peers' English achievement still does not influence own performance in English. For Chinese test scores, while the estimate is still statistically insignificant, the magnitude is significantly larger. One possible explanation for the differences across subjects may be the use of lagged peer achievement. If the learning of some subjects is more cumulative than others we would expect to see differences across subjects. Alternatively, there could be differences in the peer pressure across subjects. These mechanisms are investigated in more detail below.

Regressions in Columns 1–3 employ the entire sample, including both public students and self-pay students. Although the individual fixed effects capture the systematic differences between the two types of students, there still might be the concern that the patterns of interaction among students and between students and teachers may vary over time in different manners for the two types of students. As a robustness check, we estimate peer effects on the sample of 8 classes for which both classes in a group consist of public students. The results for the model with all fixed effects are reported in Column 4 of Table 3A. Again, the peer effects for math are positive and significant, and are insignificant for English and Chinese. The estimates are not statistically different from those of Column 3. For the remainder of the analysis, we do not distinguish between public and self-pay students.

Comparing Columns 2 and 3 of Table 3A, the coefficients on peer achievement 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 effects, 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 have the same teacher for all 3 years.[14] 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 may affect the estimates in Column 2 because it assumes 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.[15] 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. 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 of the 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. For both subjects, in our preferred specification (controlling for both individual and teacher-by-test fixed effects), the coefficient estimates on lagged peer achievement are only half the size of the estimates in Column 3 of Table 3A, and neither is significantly different from 0. This suggests that the relationship identified in Table 3 indeed measures effects due to interactions between classmates, as opposed to measurement errors.

Our preferred specification controls for individual and teacher-by-test fixed effects; thus the peer effect is identified off of the transient 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. For Math, the test with the smallest amount of remaining variance, this means that a one standard deviation change in the transient component of peer achievement roughly equals 0.04 standard deviations of the overall distribution of peer achievement. While small, this transient variation allows us to estimate the peer effect for Math precisely. The robustness of the estimates over various specifications suggests that the Math test is likely to measure quite accurately students' actual learning of the subject and refutes a simple measurement error explanation. For English and Chinese, while the remaining variance is slightly larger than Math, the peer effect estimates are more noisy, suggesting that the English and Chinese tests may not be a good measure of students' learning in these subjects. These results are consistent with the general observations that Math skills are more objective and easier to measure in tests, whereas language skills are more subjective and difficult to measure.

Overall, peers have a positive and significant effect on a student's math achievement, a positive but insignificant effect on Chinese achievement, and no effect on English achievement. Because our estimates are identified from transient changes in

[13] 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.

[14] 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.

[15] The coefficient estimates on peer achievement 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.

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 and teacher FE (3)	Ratio (2)/(1)	Ratio (3)/(1)
Average of lagged peer Math score	0.0464	0.0044	0.0018	0.09	0.04
St.D. of lagged peer Math score	0.0172	0.0042	0.0013	0.24	0.08
Average of lagged peer English score	0.0260	0.0038	0.0019	0.15	0.07
St.D. of lagged peer English score	0.0251	0.0081	0.0038	0.32	0.15
Average of lagged peer Chinese score	0.0299	0.0147	0.0023	0.49	0.08
St.D. of lagged peer Chinese score	0.0141	0.0084	0.0018	0.60	0.13

Table 5
Once- and twice-lagged peer achievement on Math, English, and Chinese.

	Dependent variable					
	Normalized Math score		Normalized English score		Normalized Chinese score	
Average of once-lagged peer Math score		0.445 [0.100]**				
Average of twice-lagged peer Math score	0.058 [0.141]	−0.079 [0.200]				
Average of once-lagged peer English score				−0.395 [0.184]*		
Average of twice-lagged peer English score			0.25 [0.057]**	0.351 [0.113]**		
Average of once-lagged peer Chinese score						1.019 [0.254]**
Average of twice-lagged peer Chinese score					0.206 [0.241]	0.67 [0.443]
Constant	0.018 [0.035]	0.034 [0.019]	0.106 [0.004]**	0.115 [0.009]**	−0.13 [0.067]+	−0.052 [0.059]
Number of observations	2645	2645	2644	2644	2642	2640
Number of students	910	910	910	910	910	910
R-squared	0.02	0.02	0.02	0.02	0.05	0.06

Note: Robust standard errors clustered at classroom level in brackets.
+ Significant at 10% level.
* Significant at 5% level.
** Significant at 1% level.

peer achievement, we conjecture that there are plausibly two major channels of the effects of peers, and if they operate differently for different subjects, we would expect to see differences in peer effect estimates across subjects. First, learning is cumulative so that transient declines in peers' learning in the past can hinder instruction in the present as current instruction must compensate for the past deficiencies. Second, students feel pressure to perform well relative to their peers and hence respond to prior outcomes with their current effort choices. In the sample school, a teacher–parent meeting is held in each class after each final exam. 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.[16] 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.

To distinguish between these two explanations, we estimate our preferred specification (Column 3 of Table 3) controlling instead for twice-lagged peer achievement. Intuitively, if the peer effects result from the cumulative nature of learning, then twice-lagged peer achievement should affect current achievement in a fashion similar to the once-lagged peer achievement. If peer effects result from students responding to transitory changes in peer achievement and the resulting social pressure, any social pressure resulting from transitory changes in a student's relative position two exams prior is not likely to persist after a "new" ranking of peers is released with a new set of exam results. Thus social pressure from changes in peers' achievement two periods prior may not influence current scores given that the change in peers' performance two periods prior was transitory.[17]

These results are presented in Table 5. For Math, the coefficient estimate on twice-lagged peer achievement (shown in column 1) is 0.058 with a standard error of 0.141; the estimate is also significantly different from that on the once-lagged peer achievement in

[16] 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, and this may also underlie Lai's (2008) finding of positive peer effects in a panel regression with student fixed effects.
[17] We thank an anonymous referee for this suggestion.

Column 3 of Table 3A. We further estimate a specification including both once- and twice-lagged peer achievements, shown in column 2. The estimate on twice-lagged peer achievement is −0.079 with a standard error of 0.2, while that on the once-lagged peer achievement is virtually identical to that in Table 3A. Both tests suggest that there is a strong "social pressure" explanation for the peer effects in Math test. Once lagged peer achievement does influence student performance, but twice lagged peer achievement does not. For Chinese, when twice-lagged peer achievement is included on its own, the estimate is 0.206 with standard error of 0.241 — it is not significantly different from zero, but also not significantly different from the estimate on the once-lagged peer achievement in Column 3 of Table 3A. When once- and twice-lagged peer achievements are both controlled for, again, the estimate on twice-lagged peer achievement is not significantly different from zero or from the estimate on the once-lagged peer achievement. Given the imprecise nature of the estimates for Chinese tests, both here and in Table 3, we cannot make a definitive statement about either the existence or the channel of peer effects in Chinese tests. For English test scores, the once- and twice-lagged effects cancel each other out; the unusual pattern of these results, particularly in comparison to Table 3, suggests a more complicated mechanism that we cannot identify with the data at hand.

Table 6 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 peer achievement, measured by the standard deviation of lagged peer achievement. While most of the literature on peer effects has focused on a linear-in-means model, it is possible that not only average achievement but also the heterogeneity of the class affects individual achievement. This issue is particularly relevant to the debate on tracking of students by ability. If students benefit from having a wide distribution of achievement levels, this suggests that tracking is not desirable. If on the other hand students benefit from more homogeneous classrooms then tracking is desirable. However, we find that 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 deviation 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). One qualitatively different result when the standard deviation of peer scores is included 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 deviation increase in average Chinese achievement of peers increases one's Chinese score by 0.042 standard deviations. Taken all together, this suggests that the effects of heterogeneity in the classroom are an argument neither for nor against tracking.

In the last column of Table 6, 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

Table 6
Effects of the average and standard deviation of peer achievement on Math, English, Chinese, and total.

Dependent variable:	1	2	3	4[a]
	Normalized score	Normalized score	Normalized score	Above median
Math:				
Average of lagged peer Math score	0.362	0.219	0.366	0.385
	[0.102]**	[0.120]+	[0.211]+	[0.160]*
St.D. of lagged peer Math score	−0.132	−0.129	−0.071	0.406
	[0.110]	[0.112]	[0.292]	[0.207]*
English:				
Average of lagged peer English score	−0.008	−0.086	0.083	−0.204
	[0.115]	[0.121]	[0.210]	[0.155]
St.D. of lagged peer English score	−0.073	−0.064	0.112	−0.225
	[0.079]	[0.080]	[0.150]	[0.101]*
Chinese:				
Average of lagged peer Chinese score	0.171	0.147	0.416	0.412
	[0.093]+	[0.097]	[0.246]+	[0.169]*
St.D. of lagged peer Chinese score	0.119	0.117	0.282	0.3
	[0.112]	[0.115]	[0.271]	[0.160]+
Number of observations	3594	3594	3594	3594
Test FE	Yes	Yes	Yes	Yes
Individual FE	No	Yes	Yes	Yes
Subject teacher-by-test FE	No	No	Yes	Yes
Sample	All	All	All	All

Note: Robust standard errors clustered at classroom level in brackets.

[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.

+ Significant at 10% level.

* Significant at 5% level.

** Significant at 1% level.

middle school are able to do well enough in the high school entrance exam to be eligible for the top-tier high schools.[18] 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 increase 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.[19]

4.2. Heterogeneity in peer effects

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 ability. In view of the peer pressure discussed above, top students may not feel as much pressure as other students in the class. To examine this possibility, we interact average peer achievement with indicators for the student's position in the achievement distribution based on test score in the first test in Grade 7.[20]

Table 7A reports peer effect estimates for students at different quintiles of initial performance distribution relative to their classmates. We only consider the average of lagged peer achievement. All fixed effects are included. Average peer achievement shows different patterns of influence for different subjects. For Math and Chinese, students at the middle quintiles generally benefit significantly from having better peers, whereas students' test scores in the top (Quintile 5) and bottom quintiles are not statistically significantly affected by peer achievement. In both cases, the coefficients for the top quintile are statistically significantly different from the coefficients for the middle quintiles (p-values for the joint test are 0.09 and 0.007 respectively). For Chinese, the coefficient for the bottom quintile is also statistically significantly different from those for the middle quintiles (p-value for the joint test is 0.007). For English, students in the bottom quintile of the distribution are significantly hurt by better peer achievement; students in the other quintiles largely benefit from better peers, but the estimates are not significant. The differences between the bottom quintile and the other quintiles are statistically significant (the p-value for the joint test is 0.001). The estimates for English also appear to explain the lack of peer effect in English from Tables 3 and 5 — the average masks the heterogeneity of peer influence on different types of students.

Table 8 shows the probability of falling in each quintile of the grade distribution conditional on the position in the class distribution of the initial achievement for public and self-pay students separately. The difference between the two types of students shown in Table 8 combined with results in Table 7A suggests that an alternative class assignment rule mixing public and self-pay students may benefit some students at the cost of others. Had classes been assigned pooling public and self-pay students, 25% of public class students in the second lowest quintile of class distribution would fall in the lowest quintile in a new, “mixed” class, whereas 20% of self-pay class students in the lowest quintile of class distribution would fall into the second quintile of a “mixed” class. Similarly, 10% of the top quintile students in the public classes would fall into the second highest quintile in a “mixed” class, while more than 30% of 4th quintile student in self-pay classes would fall into the top quintile in a “mixed” class. From Table 7A we see that students moving from the top or bottom to the middle of the class distribution would start to benefit from high-quality peers, whereas students moving the other way round would lose the benefit from high-quality peers.[21]

We also conduct an analysis similar to Table 7A, but estimate peer effects for students at different quintiles of initial distribution over the entire *grade*; the results are reported in Table 7B and are broadly similar to Table 7A. To the extent that the results in Table 7B are valid for all middle schools in the school district, the above implication would also be true if students were assigned to schools by a rule pooling pure public schools and public-private hybrid schools. Again, a more mixed school would benefit certain 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 quintile, 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 quintiles, as found in Table 7. To illustrate this, 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 quintile also does well, but will not be much affected by the performance of peers in the third quintile. But the average of peer performance for the whole class is more likely to reflect the average of peer performance in the third quintile than the average of peer performance in the top quintile.

[18] 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.

[19] This result may be unique to schools that focus their foreign language education on reading and writing (as in most Chinese schools) as opposed to listening and speaking.

[20] For Math and Chinese, we also use the diagnostic test scores to measure a student's initial ability, and the results are very similar.

[21] Duflo et al. (Forthcoming) examine the role of tracking for first graders in Kenya. While their results suggest that students at both the top and bottom benefit from tracking, their findings may result from the very large heterogeneity in terms of age, school preparedness, and support at home. In more developed countries or later grades, less heterogeneity may reduce the benefits of tracking.

Table 7
Effects of peers on Math, English, Chinese, and total by quintile of first test performance.

7A: quintile over classmates	Math	English	Chinese
Average of lagged peer performance * Quintile 1	0.449	−0.746	0.185
	[0.327]	[0.323]*	[0.204]
Average of lagged peer performance * Quintile 2	0.693	0.233	0.515
	[0.362]+	[0.336]	[0.190]*
Average of lagged peer performance * Quintile 3	0.472	0.191	0.214
	[0.149]**	[0.138]	[0.184]
Average of lagged peer performance * Quintile 4	0.397	−0.048	0.467
	[0.155]*	[0.194]	[0.162]*
Average of lagged peer performance * Quintile 5	0.073	0.201	−0.108
	[0.138]	[0.343]	[0.249]

7B: quintile over entire grade	Math	English	Chinese
Average of lagged peer performance * Quintile 1	0.515	−0.82	0.231
	[0.423]	[0.221]**	[0.188]
Average of lagged peer performance * Quintile 2	0.556	0.102	0.379
	[0.285]+	[0.337]	[0.212]+
Average of lagged peer performance * Quintile 3	0.401	0.136	0.268
	[0.238]	[0.250]	[0.174]
Average of lagged peer performance * Quintile 4	0.439	0.091	0.303
	[0.211]+	[0.219]	[0.143]+
Average of lagged peer performance * Quintile 5	0.059	0.143	0.029
	[0.140]	[0.356]	[0.324]

Robust standard errors clustered at classroom level in brackets. Test, individual, and teacher-by-test fixed effects are included. Sample includes all students. Quintile 1 is the bottom of distribution; Quintile 5 is the top of distribution.
+ Significant at 10% level.
* Significant at 5% level
** Significant at 1% level.

To test this hypothesis, we create peer achievement measures for classmates in each quintile 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 quintile 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 quintile and in other quintiles, the coefficient estimates are all statistically insignificant, which may in part be due to the small variation in achievement of peers in each quintile.

Table 8
Probability of first Math test performance in each grade quintile conditional on class quintile: public classes and self-pay classes.

Panel A: public classes

Class quintile	Grade quintile				
	1	2	3	4	5
1	99.15	0.85	0	0	0
2	25.83	69.17	5	0	0
3	0	35.71	56.35	7.94	0
4	0	0	35.25	62.3	2.46
5	0	0	0	10.43	89.57
Total	24.5	21.5	20	16.33	17.67

Panel B: self-pay classes

Class quintile	Grade quintile				
	1	2	3	4	5
1	78.33	21.67	0	0	0
2	0	50.77	49.23	0	0
3	0	0	57.75	42.25	0
4	0	0	8.77	57.89	33.33
5	0	0	0	5.26	94.74
Total	15.16	14.84	25.16	21.29	23.55

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 quintiles 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 quintile. We found no relationship between the makeup of the class and student performance.

To sum up, the results do not provide support to the hypothesis that a student is more affected by peers in her own ability quintile than those in other quintiles, and the results in Table 7 do indicate that students in the middle are more likely to be influenced by peers.[22]

5. Conclusions

Estimates of peer effects in a school setting in the US and other developed countries abound, and they emphasize the importance of peer effects in education. These findings however may not apply immediately to a developing country setting, given the large differences in cultural and educational institution background between the two types of countries. In addition, tighter budgetary constraints in the developing world make it important to carefully consider the allocation of all inputs to education, including other students.

This paper estimates peer effects on student achievement using a panel data set from a middle school in China. The unique organizational features of Chinese middle schools (Grades 7 to 9) and the panel data allow us to identify peer effects at the classroom level and to deal with the omitted variable bias in estimating peer effects due to unobserved fixed characteristics of the student or classroom 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) according to a “balanced assignment” rule and stay with the same classmates for all subjects and for all grades in middle school. (2) Each teacher of Math, English, and Chinese teaches two classes and stays with the same two classes from Grade 7 to Grade 9. These combined with the panel data allow us to use student- and teacher-by-test fixed effects to capture unmeasured constant student, class, and teacher characteristics and the time-varying common teacher effect.

Controlling for both sets of fixed effects, our preferred specification, we find positive and significant average peer effects in Math and sometimes for Chinese, but no effect in English. We did not find evidence that the heterogeneity of peer achievement affects individual test scores. Our results indicate that peer effects differ across subjects. Given that we identify peer effects off of the transient changes in lagged peer achievement, we conjecture that this difference comes from either the cumulative nature of learning or social pressure to excel that differ across subjects. We show that in Math the effects of peers stem from the social pressure to do well in the subject, a likely influence of China's highly competitive school environment. The small sample size however does not allow us to definitively test the possible mechanisms for other subjects. Another way to explain our results is as follows: at the time of assignment to classrooms, classes are characterized by within class heterogeneity but similarity across classes. After students are exposed to their classmates, the classes become more heterogeneous *across* classrooms, but more similar *within* classrooms.

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 initial achievement distribution benefit significantly from better peers, whereas students at the top and perhaps at the 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. The fact that students in the middle of a class distribution are more likely to benefit from high-quality peers suggests that altering classroom composition to take advantage of peer effects may not come without costs. A plausible, albeit speculative, extension of our results suggests that higher degree of ability mixing in school assignment may face the same tradeoff.

Appendix A

Appendix Table 1 class assignment by student rank and round.

	Student rank in round									
	1	2	3	4	5	6	7	8	9	10
Round 1	1	2	3	4	5	5	4	3	2	1
Round 2	2	3	4	5	1	1	5	4	3	2
Round 3	3	4	5	1	2	2	1	5	4	3
Round 4	4	5	1	2	3	3	2	1	5	4
Round 5	5	1	2	3	4	4	3	2	1	5

Note: Order of assignment for the case of the self-pay classes. Each number in the middle of the table is a class number. For example, the number 2 in the first column and second row indicates that the top student in round 2 of the assignment is assigned to class 2.

[22] The results are robust to using student quartile or decile distribution.

Appendix Table 2 class size and number of transfer students.

Class	Group	1. G7, Fall, Midterm	2. G7, Spring, Final		3. G8, Spring, Final		4. G9, Fall, Final		5. G9, Spring, HS Entrance exam	
		Initial class size	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

References

Ammermueller, Andreas, & Pischke, Jorn-Steffen (2009). Peer effects in European primary schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*, *27*(3), 315–348.

Brock, William A., & Durlauf, Steven N. (2001). Discrete choice with social interactions. *Review of Economic Studies*, *68*(2), 235–260.

Burke, Mary A., & Sass, Tim R. (2011). *Classroom Peer Effects and Student Achievement.* Federal Reserve Bank of Boston Public Policy Discussion Paper, No. 11-5.

Carrell, Scott E., Gilchrist, Robert N., Fullerton, Richard L., & West, James E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, *27*(3), 439–464.

Cooley, Jane. Desegregation and the Achievement Gap: Do Diverse Peers Help? Unpublished manuscript, University of Wisconsin-Madison, 2010.

De Giorgi, Giacomo, Pellizzari, Michele, & Redaelli, Silvia (2010 April). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, *2*(2), 241–275.

Ding, Weili, & Lehrer, Steven F. (2007 May). Do peers affect student achievement in China's secondary schools? *The Review of Economics and Statistics*, *89*(2), 300–312.

Duflo, Esther, Dupas, Pascaline, & Kremer, Michael (2011 August). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, *101*(5), 1739–1774.

Epple, Dennis, & Romano, Richard E. (1998 March). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, *88*(1), 33–62.

Glewwe, Paul (1997). Estimating the impact of peer group effects on socioeconomic outcomes: Does the distribution of peer group characteristics matter? *Economics of Education Review*, *16*(1), 39–43.

Hanushek, Eric A., Kain, John F., Markman, Jacob M., & Rivkin, Steven G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*, 527–544.

Hoxby, Caroline M. (2000). Peer effects in the classroom: Learning from gender and race variation. *NBER working paper no. 7867*.

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

Kang, Changhui (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics*, *61*(1), 458–495.

Lai, Fang. How Do Classroom Peers Affect Students Outcomes? Evidence from Natural Experiment in Beijing's Middle Schools. Unpublished manuscript, 2008.

Manski, Charles (1993). Identification of endogenous social effects: The reflection problem. *Review of Economics Studies*, *60*(3), 531–542.

McEwan, Patrick J. (2003). Peer effects on student achievement: Evidence from Chile. *Economics of Education Review*, *22*(1), 131–141.

Moffitt, Robert A. (2001). Policy interventions, low-level equilibria, and social interactions. In Steven Durlauf, & Peyton Young (Eds.), *Social dynamics* (pp. 45–82). Cambridge, MA: MIT Press.

Sacerdote, Bruce (2001 May). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, *116*(2), 681–704.

Selod, Harris, & Zenou, Yves (2003 August). Private versus public schools in post-Apartheid South African cities: Theory and policy implications. *Journal of Development Economics*, *71*(2), 351–394.

Zhang, Hongliang. Peer Effects on Student Achievement: An Instrumental Variable Approach Using School Transition Data. Unpublished manuscript, The Chinese University of Hong Kong, 2010.

Zimmer, Ron W., & Toma, Eugenia F. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management*, *19*(1), 75–92.

Zimmerman, David J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *The Review of Economics and Statistics*, *85*(1), 9–23.