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

Katherine Carman[1]
Lei Zhang[2]

February 2008
(Very Preliminary; Please Do Not Cite.)

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 school (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 have a positive and significant effect on math test score, a positive and marginally significant effect on Chinese test score, but no effect on English test score. Additionally, students at the middle of the ability distribution tend to benefit from better peers, whereas students at both ends do not.

[1] k.g.carman@uvt.nl Tilburg University
[2] Corresponding author, email: zlei@clemson.edu. Clemson University and Hoover Institution, Stanford University

1. Introduction

Peer effects in schools have figured very 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 desegregation, 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, 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.

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. The most commonly used 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 effects. 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 effects assumes that teacher effect does not change over time. In U.S. schools teachers usually teach only one grade; therefore using teacher fixed effects require 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 contributions.[3]

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 the 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 downside of their paper is that the peer effects are measured at the grade level not 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.[4]

The common approaches to address the reflection problem, i.e., that student and peer achievement are determined simultaneously, are either to use lagged peer achievement to measure peer attributes or to find instrumental variables for current peer achievement. The rationale for the former is that peer relationship can essentially be captured by the underlying characteristics including prior achievement – they are in part proxies for attitudes, behavioral patterns and learning related activities that systematically enter into the behavior and learning of each student (a combination of exogenous and endogenous peer effects). The latter approach

[3] If data include a panel of single cohort, teacher fixed effects and individual fixed effects cannot be simultaneously controlled for due to multicollinearity.

[4] There are also studies focusing on peer effects in college. Zimmerman (2003) considers peer effects among students living in the same dorm in Williams College, and he 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.

emphasizes the effects due to concurrent peer behaviors (endogenous peer effects). Which approach is used, however, largely depends on data availability.

In this paper, we estimate peer effects at classroom level taking advantage of two unique features of the organization of Chinese middle schools (Grades 7 to 9). We use a panel data set from a Chinese middle school. This allows us to deal particularly well with omitted variable bias in identifying peer effects due to both self-selection into class and 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.

The reflection problem is addressed through the use of lagged peer achievement. With fixed effects, specifications based on lagged peer achievement eliminate the problem of simultaneous equations bias and capture the systematic predetermined aspects of peer interactions. The coefficient estimate on the lagged peer achievement captures a combination of exogenous (peer characteristics) and endogenous (peer behavior) peer effects.[5] Individual fixed effects will pick up both the individuals 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 behavior 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

[5] Exogenous effects are also referred to as contextual effects.

that peers have a positive and significant effect on math test scores, a positive and marginally significant effect on Chinese test scores, but 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. Additionally, peer effects estimated from specifications with and without teacher fixed effects differ considerably, suggesting that teachers play an important role in the learning process.

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.[6]

Starting in the early 1990s, middle schools in China were required to abandon tracking and to take in 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 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.[7] 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

[6] 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.

[7] 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.

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 highest ranked high schools in the city. Working hard in middle school to secure a seat in a good high school is crucial for college admission – in some urban areas, close to 100% of graduates from the best high school 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 8^{th} and 9^{th} grades; chemistry in the 9^{th} 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 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 classes of a grade in the entire city follow the same curriculum. A city-wide high school entrance exam is given at the end of the 9^{th} 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.

We construct a panel of test scores for all 923 students who entered the sample middle school in the fall of 2003.[8] 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

[8] We loose 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.

middle school: midterm exam two months into middle school, final exam at the end of Grade 7, final exam at the end of Grade 8, final exam of the autumn semester of Grade 9, and the high school entrance exam. The data contain no other student characteristics besides test scores. Class size various slightly over time due to student transfer 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 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 9^{th} grade. Over the three year period, 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).

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 some measures of peer quality and ignores 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),[9] 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).

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:

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

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 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. 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 be successful in getting their children reassigned to classes taught by teachers with a reputation as better teachers.[10] In the

[9] The impact of peers as a source of motivation may be particularly strong for Chinese middle school students. During the teacher-parent meeting after each midterm or final exam, a spread sheet of test scores and ranks of all students (with student names) of a class is posted on the wall of the classroom. The competition pressure due to this public knowledge is enormous for both parents and students themselves.

[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.

presence of self-selection, any measure of peer quality also captures, for example, the unobserved characteristics of parents such as the attention they paid on 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 is 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 drop out of the model and will be captured in the individual fixed effect, 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 results of common unobserved classroom inputs, most importantly, teacher quality (correlated effects in Manski 1993). 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 limitation. Either teachers are not matched with students, or their effects are assumed to be constant over time. In the data, 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

---

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 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.

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 obvious simultaneity between individual and peer behavior. Specifications based on lagged peer achievement eliminate the problem of simultaneous equations bias and fixed effects capture the systematic predetermined and fixed aspects of peer interactions. 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 peer behavior.

It is important to state that by using the lagged peer achievement we do not try to distinguish between exogenous 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 behavior on individual behavior 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 0.155, 0.167, and 0.214 for English, Chinese and Math respectively; therefore, there is adequate variation in peer achievement in the sample. Similar, there is also reasonable variation in standard deviation of peer achievement. Table 2 reports the means and variances 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.

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 production process. Any remaining time-varying inputs are unlikely to be much correlated with the lagged peer achievement. Nevertheless, we directly test the concern of omitting time-varying historical inputs by explicitly controlling for own lagged achievement. We discuss the drawbacks of this approach in the next section when reporting the estimation results.

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 3 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. We also allow student test scores to be affected by the heterogeneity of her peers' achievement, measured by the standard deviation of lagged peer achievement. 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, but no effects for English and Chinese test scores. Heterogeneity of peer achievement does not appear to affect one's test score in any subject. 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 both the average and the standard deviation of peer achievement are quite small, as seen in Tables 1 and 2. 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 or family characteristics that are correlated with both peer quality and own achievement. In particular, the magnitude of the effect of average peer achievement for Math drops significantly, but the effect is still significant at the 10% level. A 0.10 standard deviation increase in the average of peer math scores is associated with a 0.02 standard deviation increase in individual math scores.[11]

Specification in Column 3 controls additionally for teacher-by-test fixed effects to remove the common and time-varying teacher effects. This is our preferred specification. Average peer achievement in both Math and Chinese has a positive and significant (at 10% level) effect on one's Math and Chinese test score respectively. A 0.10 standard increase in average Math achievement of peers increases one's Math test score by 0.037 standard deviations;

[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.

a 0.1 standard deviation increase in average Chinese achievement of peers increases one's Chinese test score by 0.041 standard deviations. Peers' English achievement still does not appear to influence own performance in English. Again, heterogeneity in peer achievement does not affect own performance in any subject.

One counter-intuitive finding from comparison of Columns 2 and 3 is that estimates on peer quality measures increase after teacher fixed effects are controlled for. One explanation is that teachers have different effects on different students in a class, and different students may also be affected by peers differentially. When we remove teacher influence on student performance, disproportionately more of the influence is removed from students that are less affected by peers and less from students that are more heavily affected by peers. The change in the coefficient estimate reflects the average of these interactions.

The magnitude of the estimates in Math and Chinese is larger than the estimates of Hoxby (2000) and Hanushek et al. (2003). This is in part due 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 estimates, however, is comparable to but slightly lower 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.

Column 4 reports results from a robustness test. 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 raise the issue of omitted unmeasured time-varying teacher characteristics. In the robustness

[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.

test, groups of classes that experienced teacher changes are dropped from the analysis for the entire period. This sensitivity analysis is conducted for Math, with the teacher of Classes 15 and 16 replaced in Grade 8, and for English, with the teacher of Class 8 replaced in Grade 8.[13] There are four reassignments in Chinese; the sharp reduction in sample size will render the sensitivity analysis unreliable, and hence we do not perform the analysis on Chinese. The estimates in Column 4 are very similar to the estimates in Column 3, suggesting that teacher reassignment may not be a critical issue for the estimate.

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-fixed effects. The ratio of the between the two variances ranges between 11% in average Math achievement to 61% in the standard deviation of Chinese achievement. Take average Math achievement as an example, this means that a one standard deviation change in the residual roughly equals 0.1 standard deviations of the original test score distribution. Therefore, removing individual fixed effects still leaves enough variation in peer group quality, and the pattern of estimates do not support a simple measurement error explanation.

4.2 Differences by Quartile

The results in Table 3 reveal significant influence of peer average achievement on all students in Math and Chinese but not in English. However, peers may affect some students more than others depending on their initial position in the ability distribution. 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 over the entire grade based on test score in the first test in Grade 7.

[13] Class 7 and Class 8 have different English teachers. The new English teacher of Class 8 teaches Grades 8 and 9.

Table 5 reports peer effects estimated for students at different quartiles of initial ability distribution. All fixed effects are included. Again, 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. For Math and Chinese, students at the middle two quartiles benefit significantly from having better peers; students at the bottom quartile also benefit, but the estimate is not precise; whereas students at the top quartile 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 at the second quartile benefit from better peers (significant at 10%); whereas students above median also benefit from better peers, but the estimates are not significant.

The above results suggest that reallocating students across classes will lead to significant efficiency gains, at least in Math and Chinese. For example, if we move several top students from a class with a lot of top students to one with few such that peer quality of the former only reduces slightly, whereas peer quality of the latter increases much more, then middle students in the first class will be hurt slightly due to decrease in peer quality, but the middle students in the second class will experience a much larger benefit from the much improved peer quality. More generally, the results provide an argument against ability tracking within school.

4.3 Specifications Controlling for Own Lag

The education production function given by Equation (1) assumes that education achievement is affected only by current inputs. One potential concern is that omitted cumulative time-varying historical students and teacher inputs may be correlated with lagged peer achievement; in that case, the peer effect estimates will be biased. We argue in the last section that this is of less concern with the special features of Chinese middle schools and the panel data; nevertheless, we here provide a sensitivity analysis to test the severity of the problem. In the analysis, we control for each individual's lagged test score, assuming that the lagged test score captures all the past educational inputs. This is a very strong assumption as carefully discussed in Todd and Wolpin (2003), and the test is not to be taken as definitive.[14]

Table 6 reports both average peer effects and peer effects by quartile of initial ability distribution with control for own lagged test score. All fixed effects are includes. The coefficient estimate on own lagged test score is negative and significant for Math and Chinese; with individual fixed effects, this largely reflect mean reversion. Average peer effect estimates become much smaller and insignificant for all three subjects. Peer effects by quartile are similar to the estimates in Table 5 for Math and English, but the magnitude is smaller. Also because of the smaller magnitude, the peer effects by quartile are no longer significant for Chinese. Thus, the ability specific peer effects in Math and English appear to be quite robust. Additionally, small sample size and the large standard errors suggest that we cannot rule out the peer effects definitively simply because of their statistical insignificance.

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 classroom and common teacher effect.

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

[14] In addition, strict exogeneity is violated when lagged dependent variable is included as a regressor on the right hand side of a panel regression model, causing inconsistency.

Controlling for both individual and teacher fixed effects, we find positive and significant average peer effects in Math and 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. Additionally, peer effects estimated from specifications with and without teacher fixed effects differ considerably, suggesting that teachers play an important role in the learning process.

The heterogeneous peer effects for students at different points of ability distribution suggests that reallocating students across classes will bring about efficiency gains and that ability tracking within school is not an efficient way to take advantage of peer effects. Whether this implication can be extended to allocation of students across schools also depends on the teacher quality of different schools. Since the sample middle school is very typical of middle schools in urban China,[15] these results could have broad implications for many students.

[15] Casual observations suggest that Eastern European countries also have similar features of school organization.

References

Ammermueller, Andreas and Jorn-Steffen Pischke. Peer Effects in European Primary Schools: Evidenc 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.

Todd, Petra E. and Kenneth I. Wolpin. On the Specification and Estimation of the Producation Function for Cognitive Achievement. *Economic Journal*, 113 (February): F3-F33, 2003.

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

Table 1: Summary Statistics of Lagged Peer Achievement

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

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

		Math		English		Chinese	
Group	Class	Mean	Variance	Mean	Variance	Mean	Variance
1	1	-0.177	0.00047	-0.098	0.00041	0.022	0.00033
	2	-0.078	0.00028	-0.130	0.00036	0.003	0.00024
2	3	0.291	0.00019	0.051	0.00020	-0.062	0.00024
	4	0.129	0.00032	-0.161	0.00047	-0.207	0.00049
3	5	-0.033	0.00032	0.067	0.00020	-0.019	0.00030
	6	0.377	0.00016	0.065	0.00016	0.013	0.00020
4	7	-0.270	0.00041	-0.099	0.00033	-0.064	0.00037
	8	-0.326	0.00043	-0.047	0.00031	-0.189	0.00027
5	9	-0.015	0.00024	0.050	0.00022	0.253	0.00021
	10	-0.243	0.00034	-0.031	0.00025	0.111	0.00032
6	11	0.151	0.00022	0.080	0.00022	-0.042	0.00021
	12	-0.007	0.00018	0.120	0.00027	0.021	0.00025
7	13	0.184	0.00021	0.033	0.00032	0.093	0.00040
	14	0.043	0.00035	-0.123	0.00042	0.125	0.00020
8	15	-0.273	0.00045	-0.045	0.00045	-0.186	0.00056
	16	0.105	0.00034	0.217	0.00026	0.128	0.00033

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

	1	2	3	4[a]
Dependent Variable: normalized Math test score				
Average of lagged peer performance in Math	0.763	0.198	0.373	0.393
	[0.176]**	[0.118]+	[0.206]+	[0.208]+
St.D. of lagged peer performance in Math	0.25	-0.031	0.095	0.168
	[0.251]	[0.110]	[0.237]	[0.239]
Dependent Variable: normalized English test score				
Average of lagged peer performance in English	0.214	-0.085	0.044	0.068
	[0.219]	[0.118]	[0.207]	[0.207]
St.D. of lagged peer performance in English	-0.067	-0.113	0.038	0.079
	[0.209]	[0.083]	[0.151]	[0.150]
Dependent Variable: normalized Chinese test score				
Average of lagged peer performance in Chinese	0.217	0.148	0.414	
	[0.146]	[0.096]	[0.242]+	
St.D. of lagged peer performance in Chinese	0.21	0.122	0.375	
	[0.186]	[0.116]	[0.249]	
Individual FE	No	Yes	Yes	Yes
Subject Teacher FE	No	No	Yes	Yes
Sample	all	all	all	

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

[a.] Classes 15 and 16 are dropped for the Math regression; Classes 7 and 8 are dropped for English regression

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

Variable	Variance before removing individual FE	Variance after removing individual FE	ratio
Average of lagged peer performance in Math	0.0457	0.0049	0.11
St.D. of lagged peer performance in Math	0.0164	0.0042	0.26
Average of lagged peer performance in English	0.0241	0.0039	0.16
St.D. of lagged peer performance in English	0.0234	0.0073	0.31
Average of lagged peer performance in Chinese	0.0279	0.0143	0.51
St.D. of lagged peer performance in Chinese	0.0133	0.0081	0.61

Table 5: Effects of Peers on Math, English and Chinese by Quartile of Initial Performance on Each Subject

	Math	English	Chinese
Average of lagged peer performance * Quartile 1	0.325	-0.847	0.267
	[0.391]	[0.362]*	[0.317]
Average of lagged peer performance * Quartile 2	0.64	0.529	0.684
	[0.294]*	[0.295]+	[0.280]*
Average of lagged peer performance * Quartile 3	0.537	0.186	0.58
	[0.231]*	[0.252]	[0.277]*
Average of lagged peer performance * Quartile 4	-0.005	0.317	-0.024
	[0.237]	[0.235]	[0.302]
St.D. of lagged peer performance * Quartile 1	0.326	-0.255	0.27
	[0.352]	[0.278]	[0.343]
St.D. of lagged peer performance * Quartile 2	-0.067	0.202	0.577
	[0.327]	[0.202]	[0.315]+
St.D. of lagged peer performance * Quartile 3	0.363	0.136	0.411
	[0.270]	[0.181]	[0.315]
St.D. of lagged peer performance * Quartile 4	-0.285	0.105	0.075
	[0.259]	[0.158]	[0.304]

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.

Table 6: Effects of Peers with Control of Own Lagged Test Score

	Math	Math	English	English	Chinese	Chinese
Average of lagged peer performance	0.241		0.029		0.063	
	[0.212]		[0.209]		[0.236]	
St.D. of lagged peer performance	0.143		0.033		0.206	
	[0.220]		[0.150]		[0.265]	
Average of lagged peer performance * Quartile 1		0.162		-0.866		-0.059
		[0.416]		[0.364]*		[0.311]
Average of lagged peer performance * Quartile 2		0.529		0.519		0.367
		[0.294]+		[0.297]+		[0.279]
Average of lagged peer performance * Quartile 3		0.479		0.174		0.213
		[0.234]*		[0.254]		[0.276]
Average of lagged peer performance * Quartile 4		-0.193		0.3		-0.398
		[0.241]		[0.237]		[0.295]
St.D. of lagged peer performance * Quartile 1		0.334		-0.274		0.03
		[0.323]		[0.279]		[0.359]
St.D. of lagged peer performance * Quartile 2		0.021		0.202		0.442
		[0.300]		[0.202]		[0.331]
St.D. of lagged peer performance * Quartile 3		0.432		0.134		0.236
		[0.257]+		[0.182]		[0.333]
St.D. of lagged peer performance * Quartile 4		-0.236		0.107		-0.022
		[0.245]		[0.158]		[0.316]
Own lagged test score	-0.193	-0.194	-0.029	-0.026	-0.191	-0.192
	[0.021]**	[0.021]**	[0.023]	[0.023]	[0.021]**	[0.021]**

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	initial classsize	in	out	in	out	in	out	in	out
1	51	1	0	1	4	1	0	2	1
2	58	0	0	3	1	0	0	1	3
3	64	0	1	1	2	1	0	0	1
4	55	1	0	3	0	1	1	0	3
5	57	1	0	4	1	4	1	0	3
6	65	2	2	2	2	1	0	0	0
7	55	0	0	0	0	0	2	0	0
8	58	0	0	0	1	0	2	0	1
9	58	2	1	2	0	2	3	2	2
10	56	0	1	0	4	0	1	0	0
11	62	0	0	1	0	0	2	0	1
12	62	1	0	6	2	1	5	0	1
13	62	1	0	3	1	0	2	0	1
14	53	0	0	2	1	1	1	0	2
15	54	4	2	0	2	1	1	0	0
16	52	1	2	1	0	1	0	1	2