How Do Classroom Peers Affect Student Outcomes?: Evidence from a Natural Experiment in Beijing's Middle Schools[1]

Fang Lai[2]

New York University

December, 2007

Abstract: Peer effects are among the most prominent considerations in debates on public policy. However, current empirical research has not reached a consensus on the existence and nature of peer effects because of the difficulty in isolating peer effects from other confounding influences and the simultaneity of peer interactions, also known as reflection problem. This paper exploits the special rule of classroom division in Beijing's Eastern City District and a unique micro-level data set that matches students with their classmates and teachers from 1999 to 2002, and explores not only the mean classroom peer effects but also the effects of the distribution of peer quality on student academic outcomes, as well as the nonlinearity and heterogeneity of peer influence. We control for the school-fixed effects and important factors of classroom divisions according the rule to account for endogenous sorting across schools and classes. Preliminary results find that decreased classroom diversity is beneficial to the student academic performance.

[1] We thank the Education Bureau of Beijing's Eastern City District and the Institute of Education of Tsinghua University for providing the 2002 census and school administrative and survey data. The author is solely responsible for any mistakes in this paper. As additional data collection is still in process, all results in this paper are preliminary and based on the current data available. Please do not cite without permission.

[2] Contact: Fang Lai, Steinhardt School of Culture, Education, and Human Development, New York University, 309 W Kimball Hall, 246 Greene St., New York, NY 10003. Email: fang.lai@nyu.edu. We would highly appreciate any comments on this work in progress.

I. Introduction

The effects of peers on student outcomes is one of the most prominent topics being debated in public and educational policy dealing with issues such as school choice, the moving to opportunity, and academic tracking. With a greater understanding of peer effects, policymakers can utilize its benefits and avoid its pitfalls in policy design. There have been abundant studies on peer influences on a wide range of outcomes including academic achievement and individual behaviors. However, current empirical research has not reached a solid consensus on the existence and nature of peer effects (see e.g., Angrist and Lang, 2004; Arcidiacono and Nicholson, 2005; Bonesrønning, 2007; Gaviria and Raphael, 2001; Lefgren, 2004). Manski (1993, 1995, 2000) reviewed three major challenges in identifying peer effects. The first challenge is to isolate peer effects from the “correlated effects” of unobserved individual characteristics that influence both peer group formation and student behavior. The second challenge is to distinguish the effects of peers’ behavior on individual behavior (endogenous peer effects) from the influences of environment common to all members of the group, and effects of peers’ exogenous background characteristics on individual behavior (contextual effects). The third challenge arises from the simultaneity which occurs in determining both peer and individual performance, also known as reflection problem. The first challenge has been considered the most important one regarding the existence of peer influence, as “correlated effects” do not represent consequence of genuine social interactions. To deal with this challenge, some studies have used instrumental variables to disentangle the correlations between peer characteristics and unobserved individual characteristics (e.g., Evans, Oates, & Schwab, 1992; Lefgren, 2004).[3] Other studies have utilized credibly exogenous variations in peer characteristics (e.g., Dills, 2005; Hoxby, 2000; Lavy & Schlosser, 2007).[4] Additional studies have exploited randomization or natural experiment (e.g., Falk and Ichino, 2006; Sacerdote, 2001; Zimmerman, 2003)[5], while some have benefited from the panel data and teacher-student matched data sets, and control for individual fixed effects as well as other sources of confounding, e.g., teacher influence (e.g., Hanushek et .al, 2003).

This paper, while also briefly acknowledging the other two challenges, will focus primarily on the first challenge. We implement school-fixed effect estimators to account for endogenous sorting across schools, and exploit the within-school class division rule, which creates random variation in classroom peer quality in Beijing's middle schools that is uncorrelated with students' own characteristics, to explore the nature of classroom peer influence on student performance using a cohort of 7,364 students who entered Beijing's middle schools in 1999. McEwan (2003) and Kang (2007), arguing that variations in peer characteristics across classrooms in Chile and Korea, respectively, were random, have found significant peer influences by classmates. However, neither of these studies has formally discussed the actual classroom division rule and incorporated factors affecting classroom division into the peer effect estimation. We will utilize the detailed classroom division information and rich information of student and parents' characteristics in the

[3]It is important to note, however, that research following this line has often been faced with concerns about instrument validity.
[4]Dills (2005) used the introduction of a magnet school into a school district as exogenous source of change in peer quality; Both Hoxby (2000) and Lavy & Schlosser (2007) used exogenous variation in gender ratio across school cohorts to examine contextual peer effects on student performance. Hoxby (2000) additionally identified the effects of the racial composition of a school on student performance in the same manner.
[5]The first paper employed a field experiment to credibly identify peer effects; the latter two papers estimated peer effects using the conditional random assignment of roommates at Williams College and Dartmouth College, respectively.

data set to obtain valid peer effect estimates by controlling for potential confounding influences by sorting across classes, trimming the sample to a subsample that is unlikely to be contaminated by the selection of a peer group, constructing the instrumental variables, and conducting a robustness check. Moreover, we will also explore how sorting across schools and classes might bias the estimates of peer effects by using school choice information of the randomized lotteries of middle school entry in 1999, and by comparing estimates between the full sample and subsamples.

In addition to examining the linear effects of mean peer characteristics on an individual's outcomes, which may alter the distribution of social outcomes instead of improving average social outcomes, an increasing number of studies are exploring the effects of the distribution of peer characteristics on individual outcomes and the issue of nonlinearity in the relationship between peer characteristics and individual outcomes (e.g., Glewwe, 1997; Rangvid, 2003; Schneeweis and Winter-Ebmer, 2007; Ammermueller and Pischke, 2006)**.** Analysis along this line has important policy implications for ways to manipulate peer influences to maximize social outcomes. For example, if the relationship between peer performance and a student's own performance is concave, then mixing students of diverse abilities together might be a more efficient policy to improve the average performance than streaming, as students with lower performance benefit more from being exposed to higher-performing peers than those with higher performance. Ding and Lehrer (2007) found strong evidence for nonlinear and heterogeneous high school peer effects on students' college entrance test scores using unique data from a county in China's Jiangsu Province; these data contained high school entrance test scores, which were the major determinants of school placement and thus school peer characteristics. In addition to positive effects from mean peer quality, they found that reducing the variation in peer performance increased students' performance, and by implementing a partial linear estimation, the curvatures of the relationship between peer quality and students' performance varied across peer quality distribution. These findings offer no clear-cut rules of school assignment (mixing or streaming) that could improve students' performance universally. Hoxby and Weingarth (2007), in their examination of peer influence using random variations in peer characteristics generated by classroom reassignment, reviewed in great detail several models in which peer influences entered in an asymmetric and nonlinear manner. They conducted empirical tests of these models using highly flexible econometric specifications. Likewise, we intend to explore the nonlinearity and heterogeneity of peer influence, test alternative models of the structure of peer influence, and discuss the relevant policy implications.

Background

This study used census and administrative data on a cohort of 7,364 students who entered public middle schools in Beijing's Eastern City District in 1999 and graduated in 2002. The Eastern City District spans 24.7 square kilometers in the east central part of Beijing. As the second largest precinct in the old city section, it has the fourth highest population density among all districts and counties of Beijing, and its GDP per capita was 30,517.24 Yuan in 2002. It also encompasses important Beijing sites such as the Forbidden City and Tian'anmen Square, local and national governmental institutes, and popular commercial areas. The residents come from diverse socioeconomic backgrounds, and many are employed in the district's two major industries: commerce and service. In sum, then, the Eastern City District is a most important district of Beijing, and it also largely represents the metropolitan areas of China's developed regions in overall demographic and socioeconomic composition, education, and economy.

Students in China enter school at the age of 6 and spend 6 years in primary school. Upon graduating from primary school, they take an elementary school graduation examination administered by each individual elementary school on two major subjects, Chinese and Math. However, as of 1998, the test scores from the elementary school graduation examination no longer factor into the students' middle school placements; instead, students participate in randomized lotteries conditional on their applications for school entry. Students then spend 3 years (6 semesters) in middle school. Among all students who entered middle school in 1999, at least 4,948 entered middle school through randomized lotteries, and the rest directly enrolled in middle schools, via recruitment of talented students or other channels such as donations.[6] The implementation of the randomized lotteries greatly increased the within-school diversity of the students' academic and family backgrounds. Even though the assignment of students across middle schools was not completely random, student assignment to classrooms was either random or based on observed characteristics, i.e., the student's primary school performance.

The general rule that schools used for classroom division is called the S-shape division. For example, in a school with five classes; the student with highest elementary school test score is assigned to class No. 1, the student with 2nd highest elementary school test score is assigned to class No. 2, and so on, until the student with the 5th highest score is assigned to class No. 5. Then, the student with the 6th highest test score is again

[6] Among all students, a significant number also transferred from other districts.

assigned to class No. 5, the student with the 7th highest score is assigned to class No. 4, and so on, until the student with the 10th highest score is assigned to class No. 1. The cycle continues, with the student with the 11th highest score being again assigned to class No. 1, and so on, until the student with the 15th highest score is assigned to class No. 5. This rotation back and forth continues until all students are assigned. In practice, some schools implement the S-shape assignment with students from the same primary school to account for systematic differences in grading policies across primary schools; other schools simply randomly assigned students across classrooms. Most regular classrooms in middle schools are filled via this S-shape rule to secure relatively equal average performance across classrooms; as well, educational resources are equally assigned across regular classrooms using the S-shape rule.

Discarding this rule and grouping students by ability is very unlikely, at least in 1999[7] for the majority of students, because of strong objections from parents, teachers, and society. Most parents would not want their children to be sorted and labeled into differential ability groups at such an early age within each school. In addition, relieving the pressures faced by students at the secondary education stage has been one of the main purposes of the educational reform. Finally, academic tracking has been viewed as adding unhealthy physical and psychological burdens on students. Therefore, the public is reluctant to support grouping students across classrooms by ability. From the perspectives of the school administrators and teachers, grouping students across classrooms might make the assessment of teacher performance more difficult. With the S-shape rule, teachers can be assigned to classrooms without much difference in average quality as

[7] Academic tracking within schools were more and more popular several years after the first implementation of the randomized lotteries in 1998 for management and teaching convenience. As of 1999, academic tracking within schools were rare.

they would be with grouping; in fact, barring special cases, teacher assignments across regular classrooms is usually determined by a lottery after classroom division. Of course, as primary school test scores are noisy measures of students' actual academic potential, and only the rank information is used in the class division rule, the S-shape rule is more like a "convenience rule" and does not secure complete equal distribution of students' initial performance across classrooms. The distributional properties of initial test scores within classrooms might also differ as a result. However, because of its relatively simple implementation and goal of achieving equity, parents, school administrators, and teachers have widely embraced this model.

Nevertheless, despite the common practice of S-shape classroom division, a few schools, especially schools with superior performance or large size, have implemented some level of academic tracking. These schools assign students with the highest test scores or special talents or awards to a few special classes before assigning the rest students (and usually the majority) into different classrooms using the S-shape division rule. Other schools have special small-sized classes for the lowest-performing students. In 1999, one school split the student group into half according to their primary school test scores; students in the higher performing half moved into the first four classes, while students in the lower performing half were placed into the last four classes[8].

Once classroom assignment is set, students stay in the same classes throughout all three years of middle school. In China, transferring to another school or classroom during the middle school period is highly restrictive because of concerns for students' equity, management, and educational consistency. In particular, the fixed class is an essential

[8] In our analysis, to account for systematic difference between these two tracks, we used this school as two "schools", one of which consisted of the higher-performing half, and the other of which consisted of the lower-performing half.

means to cultivate the students' spirits of collectivism and teamwork in Chinese middle schools. Classes are the basic unit by which students can participate in all regular school activities, including lectures and lab sessions in all subjects, and school-wide extracurricular activities such as sports tournaments and arts and culture festivals. In this way, students also have sufficient interactions with their classmates over the middle school period, thereby justifying the use of classmates as the peer group that the students actually interact with. As a result of the fixed classes, allocation of teachers and educational resources in most cases is equal across regular classrooms, and thus parents have little incentive to move their children around—nor would schools normally consider such petitions from parents. Transferring to a special class with better educational resources and peers is more enticing; yet it is also more restrictive; in fact, it is next to impossible to justify such a transfer if the student lacks the quality permitting special class entry. In sum, after considering the S-shape class division rule and its exceptions, classroom assignment should not be correlated to student characteristics.

At the end of each semester of middle school, students take tests on all subjects learned during that semester. Five major subjects are taught during middle school: Chinese, Math, English, Physics, and Chemistry; the first three subjects are taught throughout the three years of middle school, while the last two are introduced in the third and fifth semesters, respectively. If students decide to proceed to high school, they are required to take the High School Entrance Exam (HSEE) on the five subjects. Test scores on the HSEE are essential determinants of these high school placements. Students who do not intend to enter a public high school need not take the HSEE.[9] Semester tests and the

[9]In China, the private or semi-private high school education systems, though emerging and growing quickly, are still greatly underdeveloped. Thus, public high schools are by far the first choice of students who desire

HSEE are uniform across the entire district. Grading of semester tests is administered by the individual schools, and the exams are graded by teachers who instruct the examinees; by contrast, the HSEE is graded by a centralized grading committee assigned by the District Education Bureau. This paper intends to explore the classroom peer effects on the student performance measured by these test scores[10].

Empirical Strategies

This empirical analysis is based on data of one cohort of 7,364 students who entered 28 public middle schools in Beijing's Eastern City District. Following an educational production function approach, the model of peer influences we used is

$$(1) \quad y_{isc} = \alpha + m(y_{-isc};\beta) + m(X_{-isc};\gamma) + \eta_s + h(y_{0i;}\theta_e) + X_i\delta + \nu_{cs} + \varepsilon_{isc}$$

where y_{isc} is the final test score of student *i* in school *s* and class *c*; $m(y_{-isc};\beta)$ is a function of the test scores of student *i*'s classmates (excluding *i*); $m(X_{-isc};\gamma)$ is a function of the exogenous characteristics of student *i*'s classmates (excluding *i*); η_s captures the confounding influences from the sorting of students across schools, common shocks at the school level, as well as systematic difference in grading of the semester test scores across schools; $h(y_{0;}\theta_e)$ is a function of student *i*'s primary school test scores and primary school affiliations, which is included to capture any systematic influence on the student's performance from the classroom division rules based on the student's initial

---

a well-rounded high school education rather than a vocational or private school education, and students who seek quality high school education usually take the HSEE.

[10] The current analysis used the semester test scores for reasons we will elaborate later.

performance. X_i is a vector of student *i*'s individual or family characteristics, and it is included in the model to control for additional potentially confounding influences from the individual characteristics, and to improve the efficiency of the estimation. $\nu_{cs} + \varepsilon_{isc}$ is the vector of random error terms clustering within classrooms. The elements in the vector γ measure the exogenous/contextual peer effects, and elements in the vector β measure the endogenous peer effects.

$m(\cdot)$ can be a flexible function of the student's classroom peers' characteristics. Many previous studies have used the "Linear-in-Means" model which only includes mean test scores of student *i*'s classmates ($\overline{y_{-isc}}$) and thus assumes only linear relationship between the mean peer performance and student's own performance, and we use this model as our baseline model. Existing studies on the non-linearity of peer influences have used a parametric specification of $m(y_{-isc};\beta)$, including not only $\overline{y_{-isc}}$, but also the squared term of the mean test scores of student *i*'s classmates ($\overline{y_{-isc}^2}$), and the standard deviation of the peers' test scores ($SD(y_{-ics})$). We use this specification as a starting point to examine the patterns of peer influence. In this case, the coefficient of $\overline{y_{-isc}}$ indicates the linear relationship between peers' average performance and the student's test scores; the coefficient of $\overline{y_{-isc}^2}$ depicts the curvatures (concavity or convexity) of this relationship, in which a positive coefficient indicates convexity and a negative coefficient indicates concavity. In general, a convex peer effect is considered suggesting mixing to be a more efficient class division policy than grouping in improving the average achievement, and the concave one indicating the opposite. The coefficient of $SD(y_{-ics})$ indicates how the dispersion of the classmates' performance affects the student's performance. As for $h(y_{0i;}\theta_e)$, in the baseline model, we include a high order polynomial of the student's primary school test score (y_{0i}), dummies denoting the primary school attended by the student (θ_e), and the interaction between the student's primary school test score and primary school affiliations. Hopefully, this highly flexible specification could capture the influences from class division that affect both classmates' quality and student performance. We also tried alternative specifications of $h(y_{0i;}\theta_e)$ to gauge the robustness of the results to the specifications of $h(y_{0i;}\theta_e)$ [11].

This model faces several problems. First, $h(y_{0i;}\theta_e)$, although highly flexible, might still be too restrictive and insufficient to capture all influences from the primary school test scores that could affect both class division and student performance. Moreover, there might be departures from the original classroom division. For example, a few schools might prefer grouping the talented or high-performing students or students with special talents in one or two single classrooms, and some students might still be transferring schools and classrooms after the classroom division. In another words, even after conditioning on $h(y_{0i;}\theta_e)$, unobserved characteristics might still be correlated to both peer quality through the above channels, and bias the estimates of peer effects. Second, some shocks other than peer effects might be common to students in the same classrooms; we could denote these class-level shocks as Z_{cs}, and neglecting Z_{cs} might also cause omitted variable bias of the classroom peer effect estimates. The most salient example is systematic difference in the teacher assignment across classrooms. In

[11]We also tried to use the rank instead of the values of the primary test scores; thus far, the results were quite similar. Limitation of data availability has to date prevented us from implementing more sophisticated functions based on the division rules.

particular, teacher assignment for special classes might be different from other classes. The estimated peer effects might actually capture the influence of an unobserved teacher of special quality. Third, the reflection problem might remain even after controlling for all of the above confounding influences. In fact, reflection exists even with complete randomization of students across classrooms. When including both $m(y_{-isc};\beta)$ and $m(X_{-isc};\gamma)$ in the model, the student's own performance might indirectly enter the equation, determining the student's performance via its influence on the student's peer performance, and incurring simultaneous equation bias problems (see Manski (1993) for technical details of the reflection problem). As a result, β cannot be fully identified from γ without further assumption.[12] Finally, even while including a quadratic function and the standard deviation of peers' test scores, the $m(.)$ function might still be too restrictive to capture the actual relationship between peer quality and student performance, especially when the actual relationship is highly nonlinear, or heterogeneous across peer quality and the student's own characteristics. A case in point is the "One Bad Apple[13]" or "One Shining Light[14]" model, where peers in the tails of the test score distribution exert disproportionately higher influence on student performance than those in the middle. For another example, girls might only consider other female classmates as their reference group when deciding their academic goal or effort level, or discuss coursework more frequently with girls than boys, and thus female classmates' performance might have a

[12]With some nonlinear specifications of $m(.)$, β and γ could be identified by functional form. However, the estimates might be sensitive to different specifications of the functional forms. Moffitt (2001) indicated that the endogenous peer effects could be identified under a partial-population experiment setting, and this strategy has been applied in studies such as Bobonis and Finan (2007). However, it does not apply to the context of this paper.

[13] In this model, an average student's performance is affected more by classmates at the bottom of the test score distribution than other classmates.

[14] In this model, an average student's performance is affected more by classmates at the top tail of the test score distribution than other classmates.

larger influence on the student than the male classmates' performance. Likewise, the heterogeneity of the classmates' characteristics might also affect different students differently (see. Hoxby and Weingarth (2000) for a review of the various mechanisms by which classroom heterogeneity might either increase or decrease a certain student's performance).

To identify peer effects from the "correlated effects" from unobserved individual characteristics, we used the following strategies. First, our unusually informative data set enabled us to include in the control set X_i many of the characteristics that are normally unobserved in most existing studies. Of course, however rich the control set is, it is difficult to obtain an exhaustive account of all potentially influential characteristics that might confound the estimates. Nonetheless, if controlling for a rich set of factors that are most relevant to both classroom divisions and student performance does not significantly change the estimates, it is unlikely that some remaining unobserved characteristics will substantially confound the peer effect estimates. In this data set, we have each student's primary school attended, primary school graduation test score, self-reported talents and award-winning records, responsibilities in the student body in primary school, gender, whether the student had a relative or acquaintance in the school, and whether the student was assigned through randomized lotteries or not[15]. These are usually among the major considerations in the students' classroom divisions. Moreover, we also included information of students' family backgrounds such as parents' education, party memberships, ages, household income, as well as parents' attitude variables such as how

[15]The schools usually did not divide classes explicitly based on whether the students were enrolled via randomized lotteries or not. However, students enrolled without randomization might be more likely to have special talents, or have more power to choose schools as well as classrooms. Thus, this indicator should be included in the control set.

serious (in retrospect) they were when choosing the first-choice school for their child in 1999—assuming that parents who cared enough about school choice might also have more incentive to choose the classroom when possible. In fact, along with observed abilities (students' primary school attended, primary school test scores, awards, and responsibilities in primary school), the schools rarely sort students into classrooms according to their family backgrounds. Nor do parents have any significant power to influence classroom division.

Second, using the information of special classrooms in the school administrative and survey data and self-reported students' transfer information, we compared estimates using the whole sample with estimates restricting the sample to students in the regular classrooms who never transferred across classes or schools during middle school. Classrooms with more than 5% transferring students were also excluded from the restrictive sample because the measure of peer characteristics might be endogenous for classrooms with a high proportion of transferring students. By doing so, we could look for significant differences between estimates from these two samples. If sorting on unobservables is indeed a serious problem, the estimates of peer effects from sample including special classes or classes with more transferring students, and the estimates of peer effects on students who transferred classes might be more contaminated than estimates from other classrooms and students. Comparing the estimates from the two samples should at least indicate the robustness of the estimates and the direction of the possible bias.

Finally, we can simulate the within-school classroom divisions using the S-shape rule, and construct a 95% confidence interval of the important statistics from multiple simulations. In doing so, we could either check if the actual statistics of classroom characteristics are consistent with those of the simulated classrooms, and exclude classrooms with statistics outside the 95% confidence intervals from the sample; or we could use the peer characteristics of the simulated classrooms as instruments for actual peer characteristics and implement an IV estimation. Because of missing data in the primary test scores and the unavailability of precise information on the actual class division practices of each school[16], we cannot recover the exact classroom divisions via this simulation. However, we hope to at least partially recover the patterns of data; thus, the simulated characteristics of the classroom peers should not be correlated with unobserved variables relevant to sorting and transferring across classrooms while still representing the patterns of classmates' performance[17].

To deal with the issue of common shocks to students within the same classrooms, such as differences in teacher characteristics across classrooms within the same school, we first compared the estimates from the full sample with those from the restrictive sample. The allocation of school resources and teachers are usually even across regular classrooms. Moreover, in China, teachers of different classes are much more cooperative than teachers in the U.S.: teachers instructing the same cohort usually share the same large office space and spend all their non-lecturing working hours there; they have frequent meetings, collaborate in activities, and talk informally most of the time. While it is possible that schools assign the best teachers and provide access to superior

[16]Some schools simply randomized without considering test scores; some schools applied the S-shape division based on global ranking of the students' primary school performance regardless of the students' primary school allocation; others applied the S-shape division among students from the same primary schools.
[17] We are still collecting more detailed information of the actual classroom division for each school, and will conduct the above simulation-based analysis with the augmented data.

educational resources for special classes, it would be hard to imagine that one regular class consistently received worse or better teachers across the three years compared with other regular classes. An individual teacher is also unlikely to be so influential as to account for the whole classroom's peer effects. If the allocation of teachers and resources might have biased the estimates of the classroom's peer effects, it should be a downward rather than upward bias. For example, schools might assign teachers excelling at managerial skills to the most disruptive class of the cohort, because they are usually faced with great parental pressure if one of the regular classrooms consistently and significantly falls behind others. Therefore, this complementary assignment of teachers across classrooms might generate downward bias of peer effect estimates. To manage this, we first matched teachers to classrooms and controlled for classroom-level average teacher characteristics including rank, educational degree, informal degree, and the years of teaching. We also controlled for class size and its squared terms, as it might capture some unobserved characteristics of the classrooms. For the same reason, percentage of transferring students in a class is also controlled for[18]. Second, we use peer performance in a minor subject that as an instrument for peer performance in the average score of the major subjects. In theory, the test score of a minor subject should be correlated to the average test score of the major subjects, but the teacher's assignment of the minor subject should not be correlated to the teacher's assignment of the major subjects because the minor subjects excluded from the HSEE are not considered important by school administrators, and schools in most cases assign teachers of minor subjects across classes in a random manner. Kang (2007) used a similar strategy to disentangle the confounding

[18] A class that accepts significantly more transferring students than other classes might have some unobserved characteristics or experience some special arrangements by the school.

influences from teacher assignment across classrooms from estimates of classroom peer effects on Korean students' math test scores.

The difficulty of identifying endogenous peer effects has plagued existing studies. Many researchers have used lagged measures of peers' performance to alleviate the bias because of simultaneity, assuming no serial correlation among error terms across time. We used the peers' performance measure of the first semester[19] as a proxy of peer performance, assuming that the error term component of the peers' performance measured four semesters (two academic years) before the outcome variable (the 5th semester test score[20]) reflected no significant serial correlation to the error term component of the dependent variable. Moreover, other confounding influences such as adjustment of teacher assignment based on the previous semester performance are unlikely to be significant when using such an early measure of peer performance. However, as Hanushek et al. (2003) point out, the lagged measures might not capture the impact of current peer behavior, and thus more likely are a lower-bound estimate of endogenous peer effects.

Moreover, in this paper, we focused on the variables of the peer performance measure y_{-isc} instead of the exogenous peer characteristics X_{-isc}, although some most important variables of X_{-isc} were controlled in some estimation model specifications. Of course, the endogenous effects measured in this way might capture the contextual effects

[19] We did not use the primary school performance to construct peer performance measures as it as not a good proxy of student genuine academic performance.

[20] As all middle schools busily prepare for the HSEE during the 6th semester of middle school, most middle schools are not serious about administering 6th semester tests or recording the scores. Therefore, the number of 6th-semester test score records available is too small to provide any meaningful empirical evidence, and thus we use the 5th-semester scores instead as the outcome variable. The first semester test score is used to construct peer performance measure because the semester test scores come from the same examination system and thus are more comparable than, for example, the semester test scores and the primary school graduation test scores.

from X_{-isc}; yet it is difficult to completely isolate the influences of y_{-isc} from X_{-isc}: the coefficients of the terms of y_{-isc} might be sensitive to the variables included in X_{-isc}, and it is not feasible to control for all exogenous peer characteristics measures available in the rich data set into X_{-isc} given the "curse of dimensionality", much less the remaining unobserved peer group exogenous characteristics. Moreover, policymakers and school administrators usually make relevant policies based on observations of student performance, which is a most readily observable trait, e.g., mixing of students with diverse abilities or grouping students with same abilities; thus, those measures should be correlated with unobserved exogenous student characteristics. Consequently, it is still meaningful to focus on the coefficients of peer performance measures, even though they might pick up contextual effects, especially given that very few empirical studies in this field have convincingly disentangled endogenous peer effects from contextual peer effects. Of course, in this case, we do not claim that we have identified "endogenous peer effects" per se.

Finally, to accommodate more flexible functional forms of peer interaction, we relieved the functional forms in two dimensions. First, we allowed impacts of peer performance measures (e.g. $\overline{y_{-isc}}$, $\overline{y^2_{-isc}}$, $SD(y_{-isc})$) to vary across different students. Instead of examining peer influence on an average student of the school, we implemented quantile regressions, which examined the influence of peer performance on different percentiles of the student performance distribution, thus allowing the nature of peer influences to vary across the distribution of the unobserved individual factors captured in the educational production function. We also divided the students into different quartiles of initial performance, and included interactions between the quartile indicators and peer performance to gauge the heterogeneity of peer influence on the performance of students belonging to different quartiles of initial test score distribution. Heterogeneity in peer influence across students' demographic and socioeconomic backgrounds can also be explored by including in the model interactions of relevant individual characteristics such as gender, parents' education and income, and peer performance. Similarly, heterogeneity in peer influences across different school or classroom characteristics can also be examined by including relevant interactions between school or classroom characteristics and peer characteristics in the model.

Second, to gauge the differential effects from peers belonging to a different part of the students' performance distribution, we implemented a more flexible functional form of $m(y_{-isc};\beta)$, instead of restricting the peer effect to a quadratic format in the baseline model. Thus, we could depict how the general shape of the relationship between peer performance and students' own performance changed over the distribution of average peer performance by estimating a partial linear regression,

$$(2) \qquad y_{isc5} = \alpha + V(\overline{y_{-isc1}}) + \gamma SD(y_{-ijc1}) + \eta_s + h(y_{0i;}\theta_e) + X_i\delta + \nu_{cs} + \varepsilon_{isc}$$

where y_{isc5} is the 5th semester test score (the outcome variable), and average peer performance constructed by the 1st semester test scores ($\overline{y_{-isc1}}$) enters the model in a nonparametric form $V(\overline{y_{-isc1}})$. The curvatures (convexity vs. concavity) of the estimated relationships indicate whether grouping or mixing is better to improve the performance of an average student across the range of the peer's mean performance.[21] Even with the same average performance of classmates, it is possible that not all classmates exert the same influence on an average student's performance, and thus the specification should also account for the distribution of performance among the classmates. For example, in addition to the mean peer performance (or mean peer performance and the standard deviation of the peer performance), we additionally included percentages of peers falling within each quartile of the performance distribution (dropping the middle section as a comparison group) as predictors,

$$(3) \quad y_{isc5} = \alpha + \gamma_1 P_{-isc1}^{0.25} + \gamma_2 P_{-isc1}^{0.75} + \gamma_3 \overline{y_{-isc1}} + \eta_s + h(y_{0i;}\theta_e) + X_i \delta + \nu_{cs} + \varepsilon_{isc}$$

where $P_{-isc1}^{0.25}$ indicates the percentage of student i's classmates falling below the 25th percentile of the distribution of the first semester test score, and $P_{-isc1}^{0.75}$ indicates the percentage of student i's classmates falling within the top quartile of the first semester test score distribution. Controlling for $\overline{y_{-isc1}}$, the significance of γ_1 and γ_2 indicates whether peers on the tails of the test score distribution have larger effects than peers in the middle. A significantly negative γ_1 suggests the "One Bad Apple" model, while a significantly positive γ_2 suggests the "One Shining Light" model.

More flexible functional forms could be adopted by allowing peer effects to vary depending on both the peers' performance and the student's own performance. For example, Hoxby and Weingarth (2007) tested various hypotheses of the nature of peer

[21] $SD(y_{-ijc1})$ could also enter the model in a nonparametric form; yet the relevant policy implication is not as informative as the current specification.

influences using models consisting of a full set of interactions among indicators of a student's decile and 10 variables representing shares of classmates with initial test scores in each decile. They also permitted the coefficients of the interactions to differ by classroom median performance level (via three-way interactions). We also explored similar extensions of the basic model to examine the patterns of peer influences.

Data Description

The data set used in this study combined two data sources on the students who entered public middle schools in Beijing's Eastern City District in 1999: census data collected by the District Education Bureau in 2002 and school administrative and survey data. A total of 7,004 out of 7,364 students and their parents provided valid responses to the 2002 census of middle school students and parents. The census data contained detailed information on the students' individual characteristics and family background, as well as students' and parents' responses to questions about their experiences and attitudes toward schooling and education. The administrative records contained the students' middle school and primary school affiliations, and middle school applications during the randomized lotteries in 1999; more importantly, they contained the students' test scores on the elementary school graduation exams, all semester tests during middle school, and the HSEE. Because of administrative errors and technical difficulties in combining data from different sources, 5,926 students had available test scores on the elementary school graduation exam in at least one subject; 5,268 students had 5th-semester test scores, and 5,528 students had 1st-semester test scores. In particular, three schools used letter grades instead of the 100-point-scale numeric grade used by the other schools, and thus they were excluded from analysis using the numeric grade as measures of performance. As for the availability of HSEE scores, in addition to the aforementioned reasons, students who chose not to take the HSEE constituted an important reason for the lack of HSEE scores; thus, only 4,586 HSEE scores were available[22]. The within-school variance in student performance seemed to be greatly promoted by the randomized lotteries. One-way analysis of variance showed that 95% of the variance of primary school graduation test scores among all students was left unexplained by the middle school fixed effects; Within-school variation accounted for 90% of variance in the students' 1st semester average test scores, 93% of the variance in the students' 5th semester average test scores, and 90% of the variance in the students' HSEE average scores. The school survey data also contained information on teachers who taught this cohort of students, including teacher's official quality rank, years of teaching, levels of education, and gender.[23]

Twenty-eight public schools and 174 classes were included in the data set. Twenty-nine classes in fifteen schools were identified as special classes either by school administrative and survey data or the data of classroom performance and characteristics.[24] In these special classes were 1,280 (around 17% of the total 7364) students. We only observed the class division of the third year of junior middle school. Based on the census, 740 students reported to transfer schools or classrooms after the 1st semester, 667 of

---

[22] There are many missing HSEE scores in the data, and we are trying to identify from alternative sources which of these missing scores were due to nonattendance to the HSEE. Even though there are also missing values of the 5th semester test scores, the patterns of missingness were reasonably random, especially compared with the patterns of missningness of the HSEE scores. Thus, the current analysis uses the 5th semester test scores as the outcome measure.

[23] The current teacher data available are inadequate to provide a complete match between classes and teachers. We are still trying to obtain additional data that could better link teachers with classes.

[24] For now, in addition to special classes reported by the school administrative and survey data, we considered one class as a special class if the box-plots showed that it significantly differed from other classes in some important characteristics. This approach might be too strict, though. We expect to more accurately identify the special classes via the next round interviews to the school administrative staff.

which at least reported partial information of previous school or class assignment. Among the 677 students, at most 523 students transferred to classrooms within the same schools.[25] This small percentage of classroom transfers is unlikely to have significant influence on classroom peer effect estimates. Nonetheless, we considered classroom transfers in the analysis as described in the empirical strategy.

This data set contained unusually rich information on individual students; a summary of individual characteristics and classroom characteristics is reported in Table 1. The left sub-column reports the means and standard deviations of the relevant characteristics for the whole sample, and the right sub-column reports the same statistics for a restrictive sub-sample excluding all transferring students, special classes, and classrooms with more than 5% transferring students. We could see that the primary school test score was not a very valid measure for students' genuine academic ability as most students got high scores, with the mean test score around 91 points out of 100 points, and a low standard deviation equal to 7. Thus, there might still be remaining difference in average academic ability across classes within the same school with the S-shape class division rule; moreover, this rule did not ensure equal distributional properties of student performance across classes by construction. Thus, with the S-shape class division rule, students' academic ability might still be different across classes in both its mean and its distribution. We then examined whether an individual's own characteristics were in general uncorrelated with classroom peer characteristics after controlling for the factors of classroom division. Table 2 shows the results of regressions of the measures of peer performance, namely, the mean of the classmates' 1st semester average test scores and its

[25]Many of these 523 students might have transferred from another district after the first semester; we will better identify the nature of transfers with additional incoming data.

squared term, and the standard deviation of the classmates' 1st semester average test scores, on important individual characteristics, controlling for variables that might be important to the classroom division: a quintic polynomial of the student's primary school test scores, the dummy of primary school attended, award-winning records, responsibilities in the student body, whether the student's family had any acquaintances in the school, and whether the student was enrolled in the school via randomized lotteries. We found that for the whole sample containing only the regular classes, the mean classmates' 1st semester average test scores and its squared term were not significantly correlated with the set of individual characteristics, as shown by the insignificant F-statistics, and the insignificant or weakly significant (the signs of the significant coefficients did not always suggest sorting) coefficients of individual characteristics. As for the regression with the standard deviation of the classmates' average 1st semester test scores as the dependent variable, the F-test showed that individual characteristics were jointly significant at the 0.05 level; however, only the parents' average years of education was significant at the 0.05 level, and the magnitude of the coefficient, as well as the magnitude of coefficients for the other two weakly significant variables, were small. The F-statistics dropped to insignificance when restricting the sample to students who did not transfer schools or classes during middle school and regular classes with less than 5% transferring students. We noticed that, in both samples, the indicator of disadvantaged family (i.e. family without both parents) seems to be significantly correlated with the standard deviation of the classmates' first semester performance. However, only 14% students in the sample were from disadvantaged family. Thus, after considering the classroom division rules, peer performance measures were in general not significantly correlated to the individual characteristics, indicating no significant patterns of sorting across classrooms.

Preliminary Results[26]

Some preliminary results of the peer effects estimates are summarized in Table 3. As aforementioned, the dependent variable was the student's 5th semester average test scores across subjects, and peers' performance was measure by their 1st semester average test scores across subjects to deal with the simultaneity problem. All regressions controlled for school-fixed effects to capture common shocks at the school level and sorting across schools. Columns 1 to 9 each report results of the corresponding model specification using the full sample. The results of the "Linear-in-Means" model (Columns 1, 3, 5, and 7) show significant positive effects of the classmates' mean test scores on the student's own test scores. However, the magnitudes of the coefficients dropped significantly after controlling for the class division variables (column 3) as aforementioned, possibly due to both the correction for selection bias relevant to the class division and the fact that the student's primary school test score was included as one of the class division variables, which essentially turned the model into a value-added one. Controlling for additional individual characteristics [27] and observed classroom characteristics[28], even though increasing the R-squared of the model, did not cause additional drop in the coefficients of the classmates' mean 1st semester test score in the

[26] Results in this section are preliminary results using the current data and thus are not final. We are still expecting supplemental data and checking with the data source to clarify some questions about the current data set.
[27] The individual variables included parents' average years of education, parents' average income, the age and party membership of both parents, and their profession-based socioeconomic status.
[28] Classroom characteristics included the class size and its squared term, percentage of students enrolling without randomization, percentage of teachers at classroom level with middle school, high school, professional college, university, and graduate degree, teacher gender ratio, and average years of teaching.

"Linear-in-Means" model (columns 5, and 7). When the squared term of the classmates' mean 1st semester test scores and the standard deviation of the classmates' test scores were included, the coefficient of the mean classmates' test score became significantly negative, with coefficient of its squared term being significantly positive, showing a "scoop" shape, and the coefficient of the standard deviation of the classmates' test scores was significantly negative. These patterns indicate that streaming might be more effective in improving an average student's performance, and that the increased dispersion in an average student's classmates' test scores might hurt her academic performance. This is consistent with the intuition: the randomized lotteries greatly increased the within-school diversity of the students; considering that the average class size in the data set is 42, the standard deviation of the performance of the class might have exceeded the range in which the benefits from mixing and mutual learning outweighed the disadvantage from the difficulty management and tutoring of a heterogeneous classroom[29]. This detrimental effect of increased classroom heterogeneity has been predicted by both the "Boutique Model", where students thrive in an environment tailored to their characteristics and demands and with peer group similar to themselves, and the "Focus Model", where more homogeneous classrooms are easier to organize. In fact, after the first couple of years of the randomized lotteries, more and more schools in that district have implemented academic tracking, discarding the traditional S-shape class division rule. This might be the schools' reaction to the negative effects from increased classroom diversity.

In the restrictive sample (columns 10 to15), we excluded all transferring students, special classes, and classes with more than 5% transferring students. The coefficients of the mean classmates' 1st semester test scores in the "Linear-in-Means" models dropped to insignificance with ambiguous signs based on control variables included (see column 10; the results of the other “Linear-in-Means” models showed ambiguous signs and no significance, and are not reported here), indicating that sorting students across classrooms might bias the peer effects estimates in the “Linear-in-Means” model upwards.

Examining models including the mean classmates’ 1st semester test scores, its squared term, and the standard deviation of the classmates’ test scores (columns 11-15), we found that the only robust result for both samples was the negative coefficient of the standard deviation of the classmates’ 1st semester test scores. This result was significant even after controlling for some important classmates’ exogenous characteristics (column 15): classmates’ gender ratio, average parents’ income, average parents’ years of education, their squared terms, and the standard deviation of these classmates’ characteristics. Coefficients of the measures of classmates’ exogenous characteristics also indicated that increased heterogeneity in classmates’ characteristics had negative effects on student’s own performance. We experimented with different cut-off percentage for transferring students in the classrooms to define the restrictive sample and obtained similar results[30], and the significant positive effects of the mean classmates’ average first-semester test scores of the linear-in-means models using the full sample seemed to be driven by a few special classes and classes consisting of a significant numbers of

[30] We might also have used too restrictive criterion in qualifying a class as a non-special one by assuming classes with significant different characteristics from other classes in the same school to be a special one from box-plot observation. We are still trying to obtain more accurate data about the special classes. Moreover, we are still working on the data of transferring students, trying to sort out which of them were error reports, which of them transferred from other schools, which of them transferred classes within the same school, and when they transferred into their current classes, etc., in order to make better use the these data.

transferring students[31]. To examine whether the peer effects estimates were biased by unobserved teacher characteristics, in addition to controlling for observed teacher characteristics from the available matched teacher-classroom data, we also constructed peer performance measures using classmates' 1st semester history test scores as instruments for classmates' average 1st semester test scores in major subjects, assuming the assignment of history teachers was not correlated to the assignment of teachers for the major subjects. The first-stage F-test showed significant correlation between the instrument variables and predictors, and the signs of the IV estimates of the peer effects were consistent with the original results. However, the coefficients were not significant, possibly due to reduced efficiency with the IV estimation.

To depict the actual patterns of the relationship between the classmates' mean performance and the student's own performance that might not be fully captured by the restrictive quadratic form, we relaxed the parametric form of the function $m(.)$ and implemented a partial linear semi-parametric regression as specified in equation (3), where $V(\overline{y_{-isc1}})$ is a nonparametric specification of $\overline{y_{-isc1}}$; the regression linearly controlled for $SD(y_{-ijc1})$, as well as school-fixed effects, class division factors, individual characteristics, and classroom and teacher characteristics. Following the methods in Robinson (1988) and Yatchew (1998), we estimated the model for the full sample and restrictive subsamples excluding special classes, transferring students, and classes with percentage of transferring students above a certain cut-off percentage. Figures 1 to 4 show the smoothed values of the predicted student's own 5th semester average test scores

[31] Whether the "scoop shape" relationship between the mean classmates' test scores and the student's own performance was mainly driven by the special classes and transferring students is ambiguous from current experiments with different cut-off points.

v.s. her classmates' mean 1st semester test score ($\overline{y_{-isc1}}$) for the full sample, restrictive samples with transferring students less than 20%, 10%, and 5% of the class, respectively. With all samples, tests of significance showed that the nonparametric term of the classmates' mean performance had a significant effect on predicting the student's own test score. Moreover, the quadratic form seemed to be inadequate to recover the shape of peer influence across the whole range of classmates' mean performance. Importantly, the pattern of the relationship was also sensitive to how we trimmed the samples according to the percentage of transferring students in a class, even though only around 10% students in the full sample reported to have transferred schools or classes during middle school, and some of them might simply misreported[32]. Nonetheless, comparing these figures, ability sorting across classrooms is likely to cause a spurious positive linear peer effect; although peer effect patterns revealed here were different across samples and different from those found in Ding & Lehrer (2007), we could infer a similar conclusion that curvatures of the relationship between classmates' mean performance and student's own performance might range differ across the range of classmates' mean performance, so that neither "streaming" nor "mixing" policy was a globally optimal strategy.

Finally, to explicitly check the possible bias that might be introduced by sorting across schools in addition to the other common shocks at the school level that were also captured by the school-fixed effects, we restricted the sample to the 4350 students who enrolled in schools via randomized lotteries. We divided these students into subgroups called "selection channels" according to their school applications, so that within each

[32] As it is reasonable to assume students with lower performance were more likely to misreport their school/class transfer status, and thus excluding from the sample these students from the sample and classrooms with a certain percentage of these students might be a potential confounding influence. Thus, we need to further check the school/class transfer data in order to obtain reliable results.

selection channel, the students would be sequentially randomly selected by the same set of schools. Thus, by controlling the selection channel fixed effects instead of the school-fixed effects, we at least partially controlled for the influences by sorting across schools. The estimates of peer effects might not be generalized to students who enrolled without randomization, but are indicative of the direction of bias in peer effects estimates that might be introduced by sorting across schools. Table 4 reports the results for both the full sample of these 4350 students and the restrictive sample excluding special classes, transferring students, and classes with more than 5% transferring students, controlling for the classroom division factors, individual and family characteristics, and classroom characteristics[33]. We found that the peer effects estimates of the “Linear-in-Means” models dropped significantly after controlling for the selection channel fixed effects, especially in the restrictive sample, dropping from 0.16 to 0.04. The effects of the standard deviation of classmates’ performance on the student’s own performance seemed to be larger in magnitude after controlling for the selection channel fixed effects, and were significant using the restrictive sample. We also conducted the same analysis using all students regardless of how they entered the middle school[34]. The same patterns were found using this larger sample (Table 4_1). Moreover, by introducing back to the samples students who were more actively sorting across schools, either via avoiding randomization or transferred schools after the randomization, the peer effects estimates from the “Linear-in-Means” model seemed to be further biased upward compared with

[33] By controlling the full set variables, the number of observations dropped significantly. We also tried smaller sets of covariates with more observations left in the sample, and the results were consistent with ones presented here.

[34] In fact, when we controlled for the selection channel fixed effects using this sample, we restricted the sample to students who entered the randomized lotteries, regardless of whether they stayed in their assigned school or not.

the estimates using the samples containing only students who enrolled in schools via randomized lotteries. Thus, both sorting across schools and across classes might lead to upward bias of peer effects in the "Linear-in-Means" model.

The insignificant effects of the classmates' mean test scores among regular classes might be partially due to the fact that the class division rule explicitly sought to equalize the mean academic ability across regular classrooms, and thus there might not be sufficient variation in mean classroom performance across regular classes. Another possible explanation is that peer effects might exist in a non-linear manner, and might be heterogeneous across different students, different peer groups and different school or classroom contexts. We explored the heterogeneity in peer effects by including interactions between the peer quality measures and the student's own 1st semester test scores, gender, and school's previous performance measures, and no systematic and robust evidences of heterogeneity were observed. We also tested the "One Bad Apple" and "One Shining Light" model specified in equation (3), and did not find robust evidence supporting either. We intend to explore more in this direction with an improved data set after some additional data collection.

Concluding Remarks

This paper exploited the random variations in classmates' characteristics in Beijing's middle schools generated by the S-shape classroom division rule that serves to equalize average initial test scores across classes instead of academic sorting. Controlling for a flexible function of students' primary school test scores and other traits that might affect classroom division, we explored not only mean peer effects but also effects of the distribution of peer characteristics on students' academic outcomes, as well as the heterogeneity of peer influence among different students.

We found that increased variation of the classroom peers' characteristics significantly decreased students' test scores. Moreover, endogenous sorting across schools and classrooms caused upward bias in the estimates of the effects of the classmates' mean performance on the student's own performance in the "Linear-in-Means" models. These findings are instructive when employing classroom peer group dynamics to improve educational outcomes. We are still in the process of additional data collection, and expect to explore more of the mechanisms of peer influences and the heterogeneity and non-linearity of peer influences with improved data.

Reference

Ammermueller, A. and J.-S. Pischke. 2006. "Peer Effects in European Primary Schools: Evidence from PIRLS." NBER Working Paper Series. Cambridge, National Bureau of Economic Research.

Angrist, J. D. and K. Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *The American Economic Review* 94(5): 1613-1634.

Arcidiacono, P. and S. Nicholson. 2005. "Peer effects in medical school." *Journal of Public Economics* 89(1): 327-350.

Bobonis, G and F. Finan. 2007. "Endogenous peer effects in school participation." Manuscript. University of Toronto and University of California-Los Angeles.

Bonesrønning, H. 2007. "Peer group effects in education production: Is it about congestion?" *The Journal of Socio-Economics*.

Dills, A. K. 2005. "Does cream-skimming curdle the milk? A study of peer effects." *Economics of Education Review* 24(1): 19-28.

Ding, W. and S. F. Lehrer. 2006. "Do Peers Affect Student Achievement in China's Secondary Schools?" *Review of Economics and Statistics* 89(2): 300-312.

Evans, W. N., W. E. Oates, et al. 1992. "Measuring Peer Group Effects: A Study of Teenage Behavior." *The Journal of Political Economy* 100(5): 966-991.

Falk, A. and A. Ichino. 2006. "Clean Evidence on Peer Effects." *Journal of Labor Economics* 24(1): 39-57.

Gaviria, A. and S. Raphael. 2001. "School-Based Peer Effects and Juvenile Behavior." *The Review of Economics and Statistics* 83(2): 257-268.

Glewwe, P. 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, E. A., J. F. Kain, J. Markman and S. Rivkin. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics* 18(5): 527-544.

Hoxby, C. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper Series. Cambridge, National Bureau of Economic Research.

Hoxby, C. and G. Weingarth. 2005. "Taking race out of the equation: School reassignment and the structure of peer effects." Mimeo, Harvard University.

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

Lavy, V. and A. Schlosser. 2007. "Mechanisms and Impacts of Gender Peer Effects at School." NBER Working Paper Series. Cambridge, National Bureau of Economic Research.

Lefgren, L. 2004. "Educational peer effects and the Chicago public schools." *Journal of Urban Economics* 56(1): 169-191.

Manski, C. 1993. "Identification of endogenous social effects: The reflection problem." *Review of Economics and Statistics* 60: 531-542.

Manski, C. 1995. *Identification problems in the social sciences*. Cambridge, Mass: Harvard University Press.

Manski, C. 2000. "Economic analysis of social interactions." *Journal of Economic Perspectives* 14: 114-136.

McEwan, P. J. 2003. "Peer effects on student achievement: evidence from Chile." *Economics of Education Review* 22(1): 131-141.

Moffitt, R. 2001. "Policy interventions, low-level equilibria, and social interactions" in Durlauf, Steven N. and H. Peyton Young (eds.). Social Dynamics. Washington, DC: Brookings Institution Press.

Rangvid, B. S. 2003. "Educational Peer Effects Quantile Regression Evidence from Denmark with PISA2000 Data." European Society for Population Economics 2003. New York City.

Robinson, P. 1988. "Root-N-Consistent Semi-Parametric Regression." *Econometrica* 56: 931-54

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

Schneeweis, N. and R. Winter-Ebmer. 2007. "Peer effects in Austrian Schools." *Empirical Economics* 32(2-3): 387-409.

Yatchew, A.1998. "Nonparametric Regression Techniques in Economics." *Journal of Economic Literature* 36: 669-721.

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

Table 1. Descriptive statistics of key variables

	Full sample			Restrictive sample		
	Obs	Mean	SD	Obs	Mean	SD
Individual characteristics						
primary school graduation score	5926	91.36	7.00	3067	91.05	6.83
1st sem. average score (0-100 points)	5528	74.11	14.78	2776	72.38	14.89
5th sem average score (0-100 points)	5268	68.86	19.04	2645	68.03	19.17
gender (0-male; 1-female)	7225	0.50	0.50	3507	0.50	0.50
parents' average income (Yuan)	6666	1591.32	3537.06	3197	1495.11	1295.74
father's years of education	6763	12.97	2.76	3270	12.81	2.66
mother's years of education	6824	12.73	2.41	3294	12.63	2.28
father's party membership (0-no;1-yes)	6159	0.29	0.45	2862	0.29	0.45
mother's party membership (0-no;1-yes)	6140	0.16	0.37	2856	0.16	0.37
disadvantaged family	7001	0.14	0.34	3379	0.14	0.34
carelessness in school choice in 1999 (1-5)	5863	1.23	0.58	2901	1.22	0.56
enrollment via lottery	7364	0.59	0.49	3601	0.65	0.48
primary school award_role model (0-no; 1-yes)	7004	0.18	0.39	3380	0.16	0.36
primary school award_academic (0-no; 1-yes)	7004	0.07	0.26	3380	0.05	0.22
primary school award_science (0-no; 1-yes)	7004	0.09	0.28	3380	0.09	0.28
primary school award_sports and arts (0-no; 1-yes)	7004	0.23	0.42	3380	0.21	0.41
primary school award_leadership (0-no; 1-yes)	7004	0.09	0.29	3380	0.09	0.28
primary school award_other (0-no; 1-yes)	7004	0.02	0.14	3380	0.02	0.13
responsibility in the student body in primary school (1-5)	6959	2.38	1.00	3350	2.33	0.99
having an acquaintance/relative in the school (0-no; 1-yes)	5016	0.07	0.25	2491	0.07	0.25
Classmates' characteristics and classroom traits						
mean primary school test scores	5922	91.37	2.59	3063	91.06	2.10
mean 1st semester test scores	5521	74.11	7.35	2772	72.39	5.56
gender ratio	7221	0.50	0.09	3503	0.50	0.06
mean log(parents' income)	6666	7.01	0.37	3197	6.99	0.34
mean average parents' years of education	6926	12.84	0.95	3345	12.73	0.80
SD(primary school test scores)	5922	6.08	2.56	3063	6.26	2.12
SD(1st semester test scores)	5513	12.55	3.47	2770	13.76	2.63
SD(gender ratio)	7221	0.50	0.02	3503	0.50	0.01
SD(log(parents' income))	6666	1.11	1.05	3197	1.13	1.05
SD(average parents' years of education)	6926	2.06	0.40	3345	2.02	0.38

Disadvantaged family means family without both parents present.

The restrictive sample exclude all special classes, students who transferred schools and classes during middle school, and all classrooms with more than 5% transferring students.

"Carelessness in school choice" comes from parents' responses to how carefully/seriously they were when choosing their first-choice school in 1999. 1 indicates "most serious" and 5 indicates "not serious at all".

(Table 1 continued: class-level characteristics)

	Full sample			Restrictive sample		
	Obs	Mean	SD	Obs	Mean	SD
special class (0-no; 1-yes)	174	0.17	0.37	-	-	-
enrollment via randomization	174	0.63	0.48	84	0.69	0.47
class size	174	42.24	8.33	84	43.05	7.77
gender ratio (0-male; 1-female)	174	0.50	0.50	82	0.50	0.50
Teacher characteristics (classroom level)						
teacher's rank (level 1)%	170	29.67	18.81	82	29.48	19.03
teacher's rank (level 2)%	170	40.96	17.35	82	42.37	16.38
teacher's rank (level 3)%	170	29.13	20.71	82	27.92	17.28
teacher's rank (level 4)%	170	0.24	1.43	82	0.23	1.48
teacher's education						
middle school degree%	170	0.73	2.83	82	1.20	3.67
high school degree%	170	16.40	19.72	82	16.11	19.63
professional college degree%	170	22.67	16.91	82	23.67	15.76
university degree%	170	57.19	22.53	82	57.15	19.11
graduate degree%	170	3.01	10.30	82	1.88	4.78
teachers' gender ratio	170	0.75	0.14	82	0.76	0.11
years of teaching	170	16.08	4.47	82	15.38	3.88

The restrictive sample excludes all special classes, students who transferred schools and classes during middle school, and all classrooms with more than 5% transferring students.

Table 2. Correlations of individual characteristics and peer performance measures conditional on classroom division factors

Dependent variables: peer performance measures	Full sample: Mean (test score) Coef.	Robust Std. Err.	SD (test scores) Coef.	Robust Std. Err.	Squared (mean score) Coef.	Robust Std. Err.	Restrictive sample: Mean (test score) Coef.	Robust Std. Err.	SD (test scores) Coef.	Robust Std. Err.	Squared (mean score) Coef.	Robust Std. Err.
parents' average years of education	0.02	0.03	-0.04**	0.02	2.69	4.38	0.00	0.04	-0.03	0.02	1.09	6.80
parents' average income (Yuan/month)	0.06*	0.03	0.01	0.03	8.53*	4.75	0.06	0.05	0.02	0.03	9.48	7.39
disadvantaged family	-0.28	0.14	0.18*	0.09	-39.53*	20.56	-0.35*	0.20	0.33**	0.14	-49.79	30.17
father's party membership	0.01	0.14	0.17*	0.09	1.62	19.45	0.11	0.12	0.10	0.09	16.60	16.87
mother's party membership	0.04	0.19	0.08	0.12	4.31	27.97	-0.04	0.21	0.20*	0.12	-9.31	31.55
carelessness in school choice in 1999==2	0.13	0.18	-0.16	0.11	19.07	25.62	-0.15	0.22	-0.02	0.15	-21.47	30.78
carelessness in school choice in 1999==3	-0.07	0.25	-0.09	0.14	-7.96	35.74	0.06	0.29	-0.14	0.14	10.47	41.43
carelessness in school choice in 1999==4	0.25	0.65	0.31	0.64	31.62	93.33	0.13	0.45	0.79	0.49	12.81	61.95
carelessness in school choice in 1999==5	1.29*	0.73	-0.55	0.42	176.57*	100.93	0.68	0.47	-0.28	0.70	95.48	68.78
father's age	0.01	0.03	-0.02	0.02	1.29	4.42	0.00	0.04	-0.02	0.02	-0.43	6.68
mother's age	0.01	0.03	-0.01	0.02	1.34	4.31	0.01	0.03	-0.01	0.02	1.57	4.99
F-statistic of joint significance	1.24		1.92		1.22		1.52		1.53		1.48	
p-value of the F test	0.27		0.04		0.28		0.15		0.14		0.16	
Number of observations	2608		2608		2608		1533		1533		1533	
Number of classrooms	111		111		111		61		61		61	
R-squared	0.79		0.57		0.79		0.81		0.64		0.81	

Robust standard errors clustering within classrooms are reported

* significant at 10%; ** significant at 5%; *** significant at 1%

The mean, squared mean, and SD of the student's classmates' first semester average test scores are used as peer performance measures.

The full sample excludes all special classes.

The restrictive sample exclude all special classes, students who transferred schools and classes during middle school, and all classrooms with more than 5% transferring students.

All regressions control for factors of classroom division considerations including a quintic polynomial of primary school test scores, primary school affiliations, students' talents and primary school awards, enrollment status, whether the student has a relative in the school, and the student's responsibilities in the student body during primary school.

Table 3. Estimates of the effects of peer performance measures on student's own performance

Dependent variable: the 5th semester score	Full Sample									Restrictive sample					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
mean test score	1.04***	-4.09***	0.32***	-1.05	0.30***	-1.41	0.30***	-1.59**	-1.59*	0.16	-1.75	2.3	0.64	-0.03	1.06
	[0.07]	[0.85]	[0.09]	[0.83]	[0.09]	[0.85]	[0.08]	[0.80]	[0.82]	[0.21]	[2.86]	[1.67]	[2.03]	[2.80]	[3.75]
sq(mean score)		0.03***		0.01		0.01*		0.01**	0.01**		0.01	-0.02	-0.01	0.00	-0.01
		[0.01]		[0.01]		[0.01]		[0.01]	[0.01]		[0.02]	[0.01]	[0.01]	[0.02]	[0.03]
SD(score)		-0.06		-0.27		-0.47**		-0.46**	-0.29		-0.14	-0.35	-0.97***	-1.04***	-1.07***
		[0.17]		[0.17]		[0.20]		[0.21]	[0.23]		[0.34]	[0.26]	[0.35]	[0.38]	[0.32]
mean(parents' education)									-1.05						-1.3
									[0.90]						[1.82]
SD(parents' education)									-1.57						-5.21**
									[1.13]						[2.27]
mean(gender ratio)									-2.32						-15.68
									[6.46]						[16.62]
SD(gender ratio)									-62.32**						-206.05*
									[31.03]						[122.37]
mean log(parents' income)									3.51						-3.96
									[2.28]						[6.82]
SD(log(parents' income))									1.18*						-0.94
									[0.61]						[1.75]
control															
school-fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
class division variables			Y	Y	Y	Y	Y	Y	Y			Y	Y	Y	Y
individual characteristics					Y	Y	Y	Y	Y				Y	Y	Y
classroom characteristics							Y	Y	Y					Y	Y
Observations	4931	4931	3566	3566	2642	2642	2572	2572	2572	2570	2570	1914	1298	1265	1265
R-squared	0.16	0.17	0.45	0.46	0.54	0.54	0.54	0.54	0.54	0.09	0.09	0.39	0.52	0.52	0.53

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

The mean, squared mean, and SD of the student's classmates' first semester average test scores are used as predictors.

The restrictive sample excludes all special classes, students who transferred schools and classes during middle school, and all classrooms with more than 5% transferring students.

Table 4. The bias in peer effects estimates introduced by sorting across schools

Dependent variable: the 5th semester score	Full sample				Restrictive sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
mean test score	0.47***	0.41***	0.97	0.66	0.16	0.04	1.02	-0.17
	[0.08]	[0.08]	[1.02]	[1.02]	[0.14]	[0.15]	[3.14]	[2.96]
sq(mean score)			0.00	0.00			-0.01	0.00
			[0.01]	[0.01]			[0.02]	[0.02]
SD(score)			-0.11	-0.23			-0.71**	-0.75*
			[0.22]	[0.23]			[0.34]	[0.39]
Control								
selection channel		Y		Y		Y		Y
class division variables	Y	Y	Y	Y	Y	Y	Y	Y
individual characteristics	Y	Y	Y	Y	Y	Y	Y	Y
classroom characteristics	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2236	2236	2236	2236	1152	1152	1152	1152
R-squared	0.5	0.53	0.5	0.53	0.48	0.53	0.48	0.53

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

The mean, squared mean, and SD of the student's classmates' first semester average test scores are used as predictors.

The restrictive sample excludes all special classes, students who transferred schools and classes during middle school, and all classrooms with more than 5% transferring students.

Both the full sample and restrictive sample here only include students who entered their schools via randomized lotteries.

Table 4_1. The bias in peer effects estimates introduced by sorting across schools (for all students)

Dependent variable: the 5th semester score	Full sample				Restrictive sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
mean test score	0.53***	0.43***	0.41	0.4	0.30***	0.19	1.56	-0.88
	[0.07]	[0.08]	[0.84]	[0.92]	[0.10]	[0.14]	[2.64]	[2.82]
sq(mean score)			0.00	0.00			-0.01	0.01
			[0.01]	[0.01]			[0.02]	[0.02]
SD(score)			0.04	-0.27			-0.53	-0.72*
			[0.19]	[0.22]			[0.33]	[0.39]
Control								
selection channel		Y		Y		Y		Y
class division variables	Y	Y	Y	Y	Y	Y	Y	Y
individual characteristics	Y	Y	Y	Y	Y	Y	Y	Y
classroom characteristics	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2800	2419	2800	2419	1378	1227	1378	1227
R-squared	0.52	0.53	0.52	0.53	0.50	0.55	0.50	0.55

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%
The mean, squared mean, and SD of the student's classmates' first semester average test scores are used as predictors.
The restrictive sample excludes all special classes, students who transferred schools and classes during middle school, and all classrooms with more than 5% transferring students.
The samples are based on all students regardless of how they entered the middle schools.

Figure 1: Full Sample

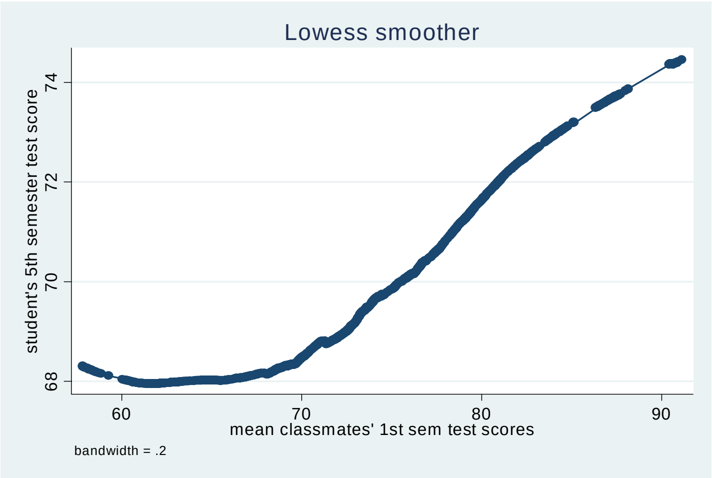

Figure 2: Restrictive sample: % of transferring students< 20%

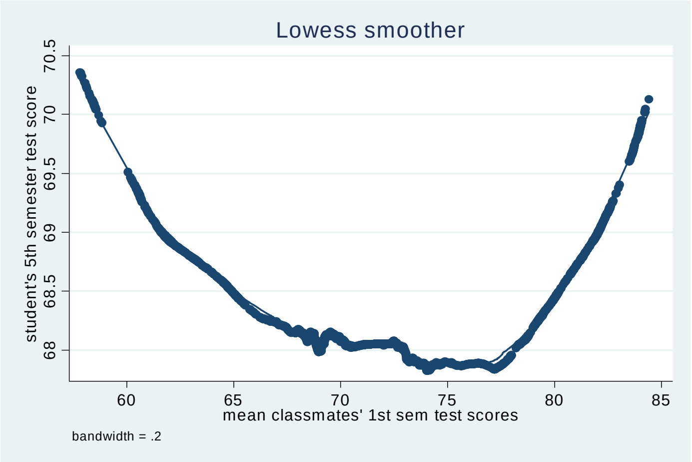

Figure 3: Restrictive sample: % of transferring students<15%

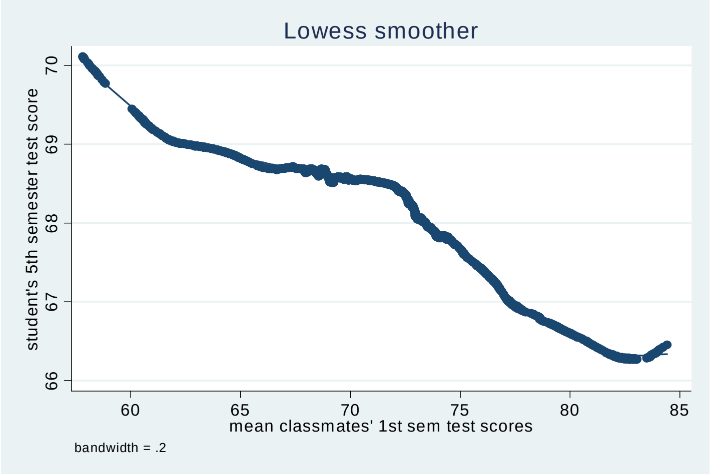

Figure 4: Restrictive sample: % of transferring students<5%