# Peer Effects on Student Achievement: Evidence from Middle School in China*

Hongliang Zhang†

Department of Economics

Chinese University of Hong Kong

August 2009

**Abstract**

In this paper, we examine peer effects on student test scores in middle school using a multi-cohort longitudinal data set from China. We base our identification on within-school, between-cohort variation in peer composition, thereby controlling for omitted variables due to unobserved school characteristics and student sorting across schools. The existing peer effects literature pays little attention to the potential positive correlation in measurement errors between the individual- and the peer-level lagged test score variables, which we find important in our data. Such a positive correlation in measurement errors arises because the individual- and the peer-level lagged test score variables are subject to transitory common shocks due to the continuing presence of a student's former peers in her current peer group. We derive formally that the presence of transitory common shocks on lagged test scores will lead to a negative bias in the estimate of peer effects. We propose an empirical strategy to address this problem by using the lagged test score measures of new peers to instrument for the corresponding lagged test score measures of all peers. Our within-school IV estimate of the linear-in-means model shows little evidence that having peers of higher average lagged test score significantly improves a student's test score. Estimates of heterogeneous peer effects models, however, show some evidence in favor of ability tracking. We find that a rightward shift in the distribution of lagged peer test scores benefits high-achieving students relative to low-achieving students, while a mean-preserving contraction in the distribution of lagged peer test scores benefits all students, but to a greater extent for those in the middle of a school's lagged test score distribution.

*I am grateful to Joshua Angrist, Esther Duflo, and Karen R. Polenske for their advice and support. I thank David Autor, Abhijit Banerjee, Amy Finkelstein, Rongzhu Ke, Tao Jin, Weifeng Li, and participants at MIT summer applied micro lunch for valuable comments and suggestions. I gratefully acknowledge financial support from the MIT Shultz Fund and the MIT Center for International Studies Summer Study Grant. All remaining errors are my own.

†E-mail: hongliang@cuhk.edu.hk/ Address: Department of Economics, Chinese University of Hong Kong, Shatin, Hong Kong SAR.

# 1 Introduction

The effects of peer groups on students' academic performance play a prominent role in various education policy debates. Many current education interventions – for example, school choice, ability tracking, and affirmative action – have the potential to influence student outcomes through their impacts on peer composition. Understanding the structure and the magnitude of peer effects is therefore a critical ingredient in evaluating these policies. However, despite the importance of peer influences for education policies, empirical research has not yet reached a consensus on the existence and the nature of peer effects. While some studies show large positive effects of peer quality on academic achievement (e.g., McEwan, 2003; Kang, 2007), others find small or insignificant effects (e.g., Angrist and Lang, 2004; Lefgren, 2004).

The lack of consensus on peer influences reflects various challenges confronted by empirical research on peer effects (Manski, 1993; Moffitt, 2001; Brock and Durlauf, 2001). The first challenge is to isolate peer effects from "correlated effects" due to the correlation between peer composition and the omitted individual or institutional characteristics that can affect student outcomes. The second challenge, known as the "reflection problem," arises from the reciprocal nature of peer interactions: a student influences her peers and is also influenced by her peers, which causes a classical simultaneity problem of econometrics. These two challenges have engaged much of the attention of the peer effects literature and have been treated intensively. The past decade has seen the development of a variety of empirical strategies to identify exogenous sources of variation in peer characteristics to deal with the endogeneity problem. These recent studies have exploited within-school (grade) variation (e.g., Hoxby, 2000; Hanusheck et al., 2003; McEwan, 2003; Vigdor and Nechyba, 2004, 2006; Lavy and Schlosser, 2007; Ammermueller and Pischke, 2009; Gould, Lavy, and Paserman, forthcoming), within-student variation (e.g., Betts and Zau, 2004; Lavy, Silva, and Weinhardt, 2009), subgroup reassignment (e.g., Angrist and Lang, 2004; Hoxby and Weingarth, 2005), instrumental variables (IV) (e.g., Kang, 2007; Zabel, 2008), and random assignment (e.g., Sacerdote, 2001; Zimmerman, 2003; Duflo, Dupas, and Kremer, 2007, 2008). Identifying the structural parameters under the simultaneity problem, however, has proved difficult or impossible without imposing severe restrictions on the econometric model.[1] The empirical peer

[1] Necessary conditions for identification of the structural parameters can be found in Brock and Durlauf (2001). Two principle methods for identification are either to introduce some type of nonlinearity into the model (e.g., suppose

effects literature has often resorted to estimating the exogenous relationship between individual outcomes and predetermined measures of peer composition to circumvent the reflection problem (Nechyba, 2006). While some studies have focused on the relationship between exogenous peer characteristics (such as race, gender, and family background) and individual outcomes (e.g., Hoxby, 2000; Angrist and Lavy, 2004; Ammermueller and Pischke, 2009; Lavy and Schlosser, 2007; Gould, Lavy, and Paserman, forthcoming), other studies have benefited from panel data to include lagged outcome measures rather than contemporary values (e.g., Hanusheck et al., 2003; Ding and Lehrer, 2003; Vigdor and Nechyba, 2004, 2006; Lavy, Silva, and Weinhardt, 2009).[2]

In this paper, we examine peer effects on students' math scores using a matched panel data set from China. The data set consists of 7,435 students from three successive cohorts of all the 15 middle schools in a school district and tracks their academic histories from finishing primary school (grades 1-6) to completing middle school (grades 7-9). By taking advantage of the panel data set, we address the reflection problem by focusing our interest on the exogenous relationship between predetermined peer characteristics – gender and lagged achievement in particular – and individual outcomes. We base our identification on within-school, between-cohort variation in peer composition, thereby controlling for omitted variables due to unobserved school characteristics and student sorting across schools. In terms of the identification strategy, the papers closest to ours are Hoxby (2000) and Gould, Lavy, and Paserman (forthcoming), both of which use comparisons in adjacent cohorts' peer composition within schools. Our identification strategy is also similar in spirit to studies that assume random classroom assignment within schools and use comparisons across classrooms for the same cohort (grade) in the same school (e.g., McEwan, 2003; Vigdor and Nechayba, 2004, 2006; Kang, 2007; Ammermueller and Pischke, 2009).

A surprising result for our data is that the within-school estimate of the coefficient of the average lagged peer test score is negative and significant. We argue that the unexpected negative sign of the estimated peer coefficient is explained by correlation in measurement errors between the individual- and the peer-level regressors. In our within-school estimation, we simultaneously control for lagged individual test score and the average lagged peer test score. These two variables,

individual behavior varies with other moments of group behavior) or to assume there exists one individual variable whose group-level average has no direct influence on individual outcomes or vice versa.

[2] Hanushek et al. (2003) provide a thorough discussion of how this does and does not fully address the reflection problem.

however, are subject to transitory common shocks due to the continuing presence of former peers in a student's current peer group. In this paper, we refer to transitory common shocks as group-specific contextual, or environmental, influences that have only transitory effects on students' observed outcomes, i.e., these influences affect the observed test scores of all students in a group, but not their abilities. For example, if a teacher happens to cover in the classroom some materials that for random reasons are tested in the exam, the test scores of all students in the class will be inflated for this particular exam. The presence of such transitory common shocks will lead to a positive correlation in measurement errors between the individual- and the peer-level lagged test score variables. We derive formally that such a positive correlation in measurement errors will lead to a negative bias in the estimated peer coefficient. In our context, this negative bias due to transitory common shocks on lagged test scores dominates the within-school estimator, making the point estimate negative and significant. The source of this bias is the presence of a student's former peers in her current peer group. The longitudinal structure of our data allows us to track the primary school origins of a student's peers and to distinguish between new peers and old peers. In our sample, on average three-quarters of a student's middle school peers are new peers. Our way to address this transitory-common-shock problem is to use the lagged test score measures of new peers to instrument for the corresponding lagged test score measures of all peers. Under the assumption that transitory shocks in lagged test scores are uncorrelated for students from different primary schools, the measurement error in the lagged test scores of new peers is expected to be unrelated to the measurement error in lagged individual test score. Hence, the transitory-common-shock problem can be circumvented by using the lagged test score measures of new peers as instruments. The existence of transitory common shocks has been well documented in the school accountability literature, in which it leads a regression-to-the-mean problem in school or teacher assessment (e.g., Kane and Staiger, 2002; Betts and Dannenberg, 2002). The potential effect of transitory common shocks on lagged individual and peer test scores, however, has largely been ignored in the peer effects literature. This paper clarifies the econometric problem of transitory common shocks on lagged test scores and makes an important methodological contribution to the existing peer effects literature by proposing an IV strategy to correct this problem.

In our linear-in-means peer effects models, we examine the effect of peer gender mix and average lagged peer test score on a student's 9th-grade math score in a school fixed-effect framework. As discussed earlier, we instrument the average lagged peer test score with the average lagged test score of new peers to circumvent the transitory-common-shock problem. Unlike some previous studies that find positive spillover effects of girls on math scores (Hoxby, 2000; Whitemore, 2003; Lavy and Schlosser, 2009), we find no evidence that peer gender composition has an impact on students' 9th-grade math scores. Our within-school IV estimate of the linear-in-means model also shows little evidence that having peers of higher average lagged test score significantly improves a student's test score in math, although the IV coefficient is quite imprecisely estimated. This finding contrasts with some existing studies on educational peer effects in China, which have found significant and positive effects of average peer achievement (Ding and Lehrer, 2007; Lai, 2007; Carman and Zhang, 2008). We believe, however, that the within-school, between-cohort variation in peer composition we rely on for identification is more credibly exogenous than those in these previous studies.[3] Some recent well-identified studies also find no evidence of a significant positive effect of average peer achievement in the linear-in-means specifications and suggest other alternative peer effects models (e.g., Hoxby and Weingarth, 2005; Duflo, Dupas and Kremer, 2008).

We then explore some simple nonlinear peer effects models, allowing peer influences to operate through the dispersion of the distribution of lagged peer test scores or through the interaction between the distribution of lagged peer test scores and a student's initial achievement. Our results on the effect of peer heterogeneity, measured by the inter-quartile range (IQR) of lagged peer test scores, show that students benefit from having more homogeneous peers: the point estimate indicates that a $0.2\sigma$ reduction in the IQR of 6th-grade peer test scores, a magnitude of change over two-thirds of the middle schools in our sample had experienced among three adjacent cohorts, can increase a student's test score by $0.1\sigma$. Estimates of heterogeneous peer effects models also show some interesting findings. First, a rightward shift in the distribution of lagged peer test scores benefits high-achieving students relative to low-achieving students, making the overall effect of the average lagged peer test score insignificant. Second, a mean-preserving contraction in the

[3] Ding and Lehrer (2007) rely on variation in peer quality across schools for identification. Even though they include observed school and teacher characteristics as control variables, the observed peer quality might still be endogenous to the unobserved school and teacher characteristics, resulting in omitted-variable biases in their estimates. Lai (2007) and Carman and Zhang (2008) both exploit variation in peer quality across classrooms within the same school-grade (cohort). They both argue that students are randomly assigned into classrooms within schools. However, even if the assignment rule is indeed random, imperfect compliance with initial classroom assignment would still lead to upward biases in their estimates. Zhang (2009) shows that a substantial proportion of students opt out of their assigned middle school in China.

distribution of lagged peer test scores benefits all students, but to a greater extent for those in the middle of a school's lagged test score distribution. Both of these findings are in favor of ability tracking for math learning.

The remainder of this paper proceeds as follows: Section 2 provides background and describes the data; Section 3 discusses the econometric problem and presents the empirical strategy; Section 4 presents the empirical results on peer effects on student test scores; and Section 5 provides some concluding remarks.

# 2 Background and Data

## 2.1 Institutional Background and Data Construction

The cornerstone of this research is the analysis of peer effects on student achievement in middle school. The data come from administrative records of a school district in the capital city of a central China province. Based on the district's administrative records, we construct a matched panel data set that tracks three successive cohorts of middle school students in the district who had completed middle school between 2005 and 2007. The middle school system of the district includes 13 public neighborhood schools and two semi-private magnet schools. Upon graduation from primary school, each student is assigned to one of the 13 neighborhood middle schools based on their residency. The zoning scheme for middle school, however, is not fixed over time. The school district creates a new zoning scheme every year and announces it in June after 6th-grade students complete their primary school. As proximity has been taken into consideration in creating the zoning scheme, students know in advance the set of nearby neighborhood middle schools they might be assigned to, but not the exact school until the announcement by the school district. Since primary school enrollment is also based on residency, a student usually has some of her former schoolmates from primary school assigned to the same neighborhood middle school with her. Students also have the option to apply to one of the two semi-private magnet schools in the district and will be selected based on an admission lottery (Zhang, 2009).[4] On average, about 30 percent of the students in the district opt out from their assigned neighborhood middle school to a semi-private magnet school.

[4] Neighborhood middle schools are tuition-free under the compulsory education law, but magnet schools charge additional tuition. Zhang (2009) provides details about the magnet school admission process.

Our panel data set is constructed by matching administrative student records from two sources. Student information at the end of middle school comes from the city's Middle School Graduation Exam (MSGE) database, which includes each student's middle school of graduation and test scores in four subjects examined in the citywide uniform MSGE: math, science, Chinese, and English. Student information before the start of middle school comes from the district's records of students' primary school of graduation and their math scores in a district-wide uniform exam taken at the end of 6th grade.[5]

Some limitations remain in the structure of the matched panel data set. First, the two databases do not share perfect individual identification information to guarantee unique tracking of the academic histories of all students. Specifically, we can only use the combination of name and gender to match student records in the two databases. Consequently, some students cannot be uniquely tracked due to multiple matches to common names. In addition, some students in the MSGE database have no matched primary school records, either because they attended a primary school outside the district or because their names were misspelled in one or both databases.[6] Second, we can only identify peer composition at the cohort (grade) level but not the classroom level. The classroom-level measures of peer composition may be more desirable if peer externalities take place mainly through classroom interactions. However, classroom-level measures are likely to be endogenous as school administrators and parents can have some discretion in placing students in different classes within a grade. Because of the potential sorting of students across classrooms within a grade, we would still use the cohort-level measures even if classroom-level measures were available.

Our sample consists of 7,435 students from three successive cohorts in the district whose academic histories are uniquely tracked. Students in our sample account for about 86 percent of the universe of 8,620 students who had completed middle school in the district during 2005-2007. Our data are ideal for analyzing peer effects in education for two reasons. First, we measure individual and peer abilities by lagged test scores, which are much more precise proxies than other individual

[5]Students also take an exam in Chinese at the end of 6th grade. The Chinese exam includes a writing section. Students' Chinese test scores are largely non-comparable across schools as grading standards differ substantially across schools.

[6]Misspelling is more likely to occur in the primary school information records as we have obtained these records in handwritten paper documents and coded them into an electronic database. The MSGE records are obtained in electronic format.

and peer characteristics such as mother's education (McEwan, 2003) and number of books at home (Ammermueller and Pischke, 2009). Second, there is a large amount of reshuffling of peers during the transition from primary school to middle school and the longitudinal structure of the data allows us to distinguish between new peers and old peers. As we will discuss in further detail in Section 3.3, separating new peers from old peers is very important for the identification of peer effects when lagged test scores are used. For concerns about the sensitivity of our results to the inclusion of two magnet schools, we replicate all our analyses to a subsample of 5,191 students from neighborhood middle schools only. Results of this subsample remain qualitatively the same as the full sample. We therefore report only the results of the full sample in this paper.

### 2.2 Descriptive Statistics

Table 1 presents descriptive statistics for the matched sample. Panel A shows statistics for three individual-level variables: gender, 6th-grade math score, and 9th-grade math score. Panel B reports four exogenous measures of peer composition in middle school: proportion of female peers, proportion of new peers, average 6th-grade peer math score, and inter-quartile range (IQR) of 6th-grade peer math scores. For the latter two measures of lagged test scores, Panel B also reports separate statistics for new peers only. Column 1 shows the means for these variables. The sample is balanced in gender, consisting of roughly 50 percent girls. On average, three-quarters of a student's peers in middle school are new peers from other primary schools. For ease of interpretation, we normalize student test scores by cohorts to have zero means and standard deviations of one. As some students are not included in our matched sample for reasons discussed above, any observed deviation of our sample mean test score from zero reflects selection into the matched sample. For instance, the average 6th-grade test score ($0.035\sigma$) in our sample is slightly higher than the district average (which is normalized to be zero). Our explanation of this difference is that a disproportionate share of students from high mobility families (e.g., rural migrants), who on average have lower academic achievement, opt out of the district's middle school system and are therefore not tracked in our matched sample. The IQR of lagged peer test scores in our sample is $1.04\sigma$. For reference, the IQR of a standard normal distribution is $1.35\sigma$, which is what we would expect to see had students been randomly assigned to middle schools. The observation of a smaller IQR of lagged peer test scores than the case of random assignment indicates student sorting in peer group formation.

Column 2 reports the standard deviations of these individual- and peer-level variables. As any between-school variation is removed in the school fixed-effect framework, our source for identification is variation across cohorts within the same school. Hence, column 3 reports a measure of the within-school dispersion of the individual- and the peer-level variables: the standard deviation of the residual of each variable after removing school and cohort fixed effects. Figures 1a to 1c plot the within-school variation in peer composition measured by gender mix, average 6th-grade test scores, and IQR of 6th-grade math scores, respectively. These figures show that there is a fair amount of cohort-to-cohort variation in peer composition within schools. For example, over the three consecutive cohorts observed in the sample, 11 out of a total of 15 schools have experienced a more than seven-percentage-point change in the proportion of female peers, 10 schools have experienced a more than $0.3\sigma$ change in the average 6th-grade math scores, and 11 schools have experienced a $0.2\sigma$ change in the IQR of 6th-grade math scores.

# 3 Empirical Strategy

## 3.1 The Model Framework

We start from a simple linear-in-means education production function of the following form:

$$Y_{ics} = \beta A_{ics} + \lambda \overline{A}_{(-i)cs} + \phi_s + \kappa_{cs} + \upsilon_{ics} \tag{1}$$

where $Y_{ics}$ is a student outcome, such as a test score, for student $i$ of cohort $c$ in school $s$; $A_{ics}$ is the exogenous predetermined ability of student $i$; $\overline{A}_{(-i)cs}$ is the average ability of student $i'$s peers; $\phi_s$ represents the school-specific common-shock effects, arising from school-level unobserved common contextual, or environmental, influences that affect the outcomes of all students in that school; $\kappa_{cs}$ represents variation in the common-shock effects across cohorts within schools and has a zero-mean within each school; and $\upsilon_{ics}$ is an individual-level stochastic error term that has a zero-mean within each cohort in each school. Note that the model is set up by assuming no cohort-to-cohort evolution in student ability $A_{ics}$ or outcome $Y_{ics}$. In practice, such cohort-to-cohort evolution can be easily controlled by including a cohort fixed effect.

The identification of peer effects $\lambda$ in equation (1) faces two major challenges. First, $A_{ics}$ and $\overline{A}_{(-i)cs}$ are latent variables and cannot be observed directly. Second, the two common-shock effects $\phi_s$ and $\kappa_{cs}$ reflect correlated effects and will give rise to a bias in the estimated peer coefficient $\widehat{\lambda}$ if they are correlated with $\overline{A}_{(-i)cs}$. Let us pretend for a moment that we have perfect measures of $A_{ics}$ and $\overline{A}_{(-i)cs}$ and focus on the second challenge of isolating peer effects from correlated effects. Random assignment of students to groups, where a group refers to a cohort in a school in our context, can solve this problem because randomization breaks the potential link between peer composition ($\overline{A}_{(-i)cs}$) and the common shock effects ($\phi_s$ and $\kappa_{cs}$). However, true random assignment rarely exists outside experimental settings (Sacerdote, 2001; Zimmerman, 2003; Duflo, Dupas, and Kremer, 2007, 2008). In practice, parents choose a school based on its quality and the composition of its peers, and schools also have some discretion in choosing students for admission. Hence, peer quality $\overline{A}_{(-i)cs}$ will be systematically correlated with common shock effects $\phi_s$ at the school level, causing the OLS estimator of $\lambda$ to be biased. A possible way to account for such school-level correlated effects is to use within-school estimation that exploits variation in peer composition across adjacent cohorts within the same school.

As shown in Appendix A, the within-school specification of the education function can be written as

$$y_{ics} = \beta a_{ics} + \lambda \overline{a}_{(-i)cs} + \kappa_{cs} + \upsilon_{ics} \tag{2}$$

where $y_{ics}$, $a_{ics}$, and $\overline{a}_{(-i)cs}$ are derivations from their school means. The basic idea behind the within-school estimation is to examine whether, for students from adjacent cohorts in the same school, those who have more favorable peers (in terms of average peer ability) in their cohort score higher conditional on their own abilities. The identification assumption of the within-school estimation is that the within-school, between-cohort variation in peer composition $\overline{a}_{(-i)cs}$ is uncorrelated with the within-school, between-cohort variation in common-shock effects $\kappa_{cs}$. Under this identification assumption, the cohort-to-cohort variation in common-shock effects $\kappa_{cs}$ can be subsumed into a general error term $\epsilon_{ics}$ such that $\epsilon_{ics} = \kappa_{cs} + \upsilon_{ics}$. Consequently, the within-school model estimates the following equation:

$$y_{ics} = \beta a_{ics} + \lambda \overline{a}_{(-i)cs} + \epsilon_{ics} \tag{3}$$

where $cov(a_i, \epsilon_{ics}) = cov(\overline{a}_{(-i)cs}, \epsilon_{ics}) = 0$.

Although equation (3) is not confounded by correlated effects given the above assumption, it still cannot be estimated directly because the de-meaned ability measures $a_i$ and $\overline{a}_{(-i)cs}$ are not directly observed. Lagged test scores are often used as proxies for latent abilities. Let $x_{ics}$ denote the deviation of the observed lagged individual test score from its school mean, and $w_{ics}$ denote the deviation of the observed average lagged peer test score from its school mean. Appendix A shows that $x_{ics}$ and $w_{ics}$ are related to $a_i$ and $\overline{a}_{(-i)g}$, respectively, as follows:

$$x_{ics} = a_{ics} + v_{ics} \tag{4a}$$

$$w_{ics} = \overline{a}_{(-i)cs} + u_{ics} \tag{4b}$$

where $v_{ics}$ is a stochastic error term that has a zero mean within each school and is uncorrelated with $\epsilon_{ics}$, and $u_{ics} = \overline{v}_{(-i)cs}$. Substituting equations (4a) and (4b) into equation (3) yields

$$y_{ics} = \beta x_{ics} + \lambda w_{ics} + \psi_{ics} \tag{5}$$

where $\psi_{ics} = \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics}$. Note that $x_{ics}$ ($w_{ics}$) and $\psi_{ics}$ are correlated because they both contain $v_{ics}$ ($u_{ics}$).

Ammermueller and Pischke (2009) consider a similar within-school estimation of peer effects in the presence of measurement errors. While we rely on variation across cohorts within the same school for identification, they use comparisons across classes within the same grade in the same school. As they do not have students' lagged test scores, they use parents' reports of number of books at home as a measure of peer composition. They argue that classes are formed roughly randomly in European primary schools and the measurement errors in books at homes are uncorrelated within classes. In our context, these assumptions would imply that the within-school, between-cohort variation in peer quality $\overline{a}_{(-i)cs}$ is idiosyncratic and not related to $a_{ics}$ or $\epsilon_{ics}$, and that the error terms $v_{ics}$ and $u_{ics}$ are uncorrelated. Under these assumptions, the within-school estimators $\widehat{\beta}_W$ and $\widehat{\lambda}_W$ converge to:

$$p\lim \widehat{\beta}_W = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_v^2}\beta \tag{6a}$$

$$p\lim \widehat{\lambda}_W = \frac{\sigma_{\overline{a}}^2}{\sigma_{\overline{a}}^2 + \sigma_u^2}\lambda \tag{6b}$$

where $\sigma_a^2$, $\sigma_v^2$, $\sigma_{\overline{a}}^2$, and $\sigma_u^2$ denote the variances of $a_{ics}$, $v_{ics}$, $\overline{a}_{(-i)cs}$, and $u_{ics}$. Although $x_{ics}$ and $w_{ics}$ are both correlated with the error term $\psi_{ig}$ in equation (5), they are uncorrelated with each other under the above assumptions. Hence, $\widehat{\beta}_W$ and $\widehat{\lambda}_W$ are both subject to attenuation biases in the classical errors-in-variables (EIV) problem. Specifically, the plims of $\widehat{\beta}_W$ and $\widehat{\lambda}_W$ are regression coefficients in the following model:

$$y_{ics} = \widetilde{\beta} x_{ics} + \widetilde{\lambda} w_{ics} + \mu_{ics} \tag{7}$$

where $\widetilde{\beta} = \frac{\sigma_a^2}{\sigma_a^2+\sigma_v^2}\beta_0$, $\widetilde{\lambda} = \frac{\sigma_{\overline{a}}^2}{\sigma_{\overline{a}}^2+\sigma_u^2}\lambda$, and $\mu_{ics} = (\beta - \widetilde{\beta})x_{ics} + (\lambda - \widetilde{\lambda})w_{ics} + (\epsilon_{ics} - \beta v_{ics} - \lambda u_{ics})$. Because of the unobservable nature of individual and peer abilities, the structural coefficients $\beta$ and $\lambda$ in equation (1) cannot be estimated directly. The regression coefficients $\widetilde{\beta}$ and $\widetilde{\lambda}$ in equation (7), however, are estimable and can be interpreted as important policy parameters of interest: the marginal effect of an individual's lagged test score and the marginal effect of the average lagged peer test score.

Ammermueller and Pischke (2009) argue that, if there exists another set of independent measures of the same individual and peer variables $x'_{ics}$ and $w'_{ics}$ (e.g., students' reports of number of books at home in addition to parents' reports), using $x'_{ics}$ and $w'_{ics}$ as instruments for $x_{ics}$ and $w_{ics}$ can correct the measurement error problem and provide consistent estimates of $\beta$ and $\lambda$. However, we do not always have two measurements of the same variables, and, even if we do, the errors in the two measurements may well be correlated. Hence, their IV approach to correct the measurement error problem may not always be feasible. Despite the classical attenuation biases, the within-school estimators $\widehat{\beta}_W$ and $\widehat{\lambda}_W$ are still informative for at least two reasons. First, they provide consistent estimators of policy parameters $\widetilde{\beta}$ and $\widetilde{\lambda}$ as defined in the previous paragraph. Second, the attenuation biases decrease with the precision of the proxy ability measures. We expect students' lagged test scores used in our study are more precise measures of abilities than other indirect measures used in some previous studies (such as mother's schooling and number of books at home).

### 3.2 The Problem of Transitory Common Shocks on Lagged Test Scores

Table 2 reports the OLS and the within-school estimations that regress students' 9th-grade math scores on their 6th-grade math scores and the average 6th-grade math scores of their peers. Each column corresponds to a separate regression and the standard errors reported in parentheses are adjusted for clustering within each cohort in each school. Column 1 reports the least square estimation that does not control for school fixed effects. The OLS estimator $\widehat{\lambda}_{OLS}$ (0.591 with a standard error of 0.104) shows a very strong positive relationship between one's 9th-grade test score and the average 6th-grade test score of one's peers in a cross-sectional setting. The large positive OLS estimator of peer coefficient, however, almost certainly confounds peer effects with "correlated effects" because of student sorting across schools based on the unobserved school characteristics. The fact that the estimated peer coefficient $\widehat{\lambda}_{OLS}$ (0.591) is even larger than the estimated coefficient of own lagged test scores $\widehat{\beta}_{OLS}$ (0.442) also implies that $\widehat{\lambda}_{OLS}$ is likely to be biased upward due to correlated effects and that the magnitude of the bias may be quite large. As we have discussed earlier, introducing school fixed effects to the model can mitigate the bias due to school-level correlated effects. Column 2 reports the results of the within-school estimation that includes both school and cohort fixed effects. The F-test of the joint significance of the school fixed effects has a p-value below 0.001, showing evidence of the existence of school-level correlated effects. Not surprisingly, $\widehat{\lambda}_W$ is reduced considerably in the within-school estimation. What is perhaps surprising is that $\widehat{\lambda}_W$ is now negative and significant (with a point estimate of $-0.250$ and a standard error of 0.112). Although the empirical literature has not reached a consensus on the existence and the magnitude of peer effects, the true peer coefficient $\lambda$ is unlikely to be negative. Hence, we take the negative and significant point estimate of $\lambda$ as evidence that our within-school estimator $\widehat{\lambda}_W$ is subject to a negative bias that cannot be simply explained by the attenuation bias.

Next, we revisit the assumptions used to derive the within-school estimator, equation (6b), to examine the potential sources of such a negative bias. First, we assume that within-school, between-cohort variation in peer quality $\overline{a}_{(-i)cs}$ is unrelated to the de-meaned individual ability $a_{ics}$. This assumption implies that student sorting only occurs across schools but not across cohorts within the same school. This is plausible in our context as parents are unlikely to be well informed and sophisticated enough to condition their school choice decision on the cohort-to-cohort variation in peer quality within a school. Moreover, with classical measurement errors, within-school student sorting will introduce an upward bias in $\widehat{\lambda}_W$, opposite to what we have seen in the data. Specifically, if there exists within-school student sorting by ability, i.e., $\pi = cov(a_{ics}, \overline{a}_{(-i)cs}) > 0$, the within-school estimator $\widehat{\lambda}_W$ will converge to $\lambda - \frac{\sigma_u^2}{(\sigma_a^2+\sigma_u^2)-\frac{\pi^2}{(\sigma_a^2+\sigma_v^2)}}\lambda + \frac{\sigma_v^2}{(\sigma_a^2+\sigma_v^2)}\frac{\pi}{(\sigma_a^2+\sigma_\mu^2)-\frac{\pi^2}{(\sigma_a^2+\sigma_v^2)}}\beta$. The third component of this expression can be interpreted as a correlation bias, which arises from correlation between $a_{ics}$ and $\overline{a}_{(-i)cs}$, and has the same sign as own ability effect $\beta$. Our second assumption is that, within the same school, the cohort-to-cohort variation in peer quality $\overline{a}_{(-i)cs}$ is uncorrelated with the cohort-to-cohort variation in common-shock effects $\kappa_{cs}$. A downward bias in $\widehat{\lambda}_W$ would arise if $\overline{a}_{(-i)cs}$ is instead negatively correlated with $\kappa_{cs}$. This would be the case if, when a cohort quality is relatively poor in a school, a principal who cares about within-school equity assigns high-quality teachers to that cohort to partly compensate for the poor student quality. However, the extent of such endogenous teacher assignment, if it exists at all, is likely to be quite limited as teachers usually rotate their grade assignment on a three-year basis (grades 7 to 9). Moreover, we would not expect a principal to manipulate teacher assignment to the extreme extent to more than fully compensate the difference in peer quality such that the net effect $(\lambda\overline{a}_{(-i)cs} + \kappa_{cs})$ is negatively correlated with peer quality $\overline{a}_{(-i)cs}$.

The third assumption to derive equation (6b) is that measurement errors in the individual- and the peer-level lagged test scores are uncorrelated, i.e., $\rho = cov(v_{ics}, u_{ics}) = 0$, a condition that would hold if the error terms of lagged individual test scores are i.i.d. within each cohort in each school. However, students usually take some former peers with them when moving to the next schooling phase. In our sample, about a quarter of a student's peers in middle school are her former peers from the same primary school. To the extent that the lagged test scores of students from the same primary school are subject to transitory common shocks, the presence of a student's former peers in her current peer group leads to a positive correlation between $v_{ics}$ and $u_{ics}$, i.e., $\rho > 0$. In this paper, we refer to transitory common shocks as group-specific contextual, or environmental, influences that have only transitory effects on students' observed outcomes, i.e., these influences affect the observed test scores of all students in a group, but not their permanent abilities.[7] As we

[7] The literature usually uses the terminology "common shocks" to refer to common contextual factors, such as school resources and teacher quality, that affect students' test scores through their effects on students' abilities. However, the transitory common shocks we consider here have no (permanent) influences on students' abilities, but

have mentioned earlier, random overlapping between testing contents and teachers' instructions is one source of such transitory common shocks. Another source of such transitory common shocks is the across-school difference in grading standards, causing students' grades to be inflated in some schools but deflated in others. Note that random assignment of teachers to grading at the school or class level cannot alleviate the second type of transitory common shocks, although random assignment of grading at the individual level will work. Moreover, there is no direct way to correct the first type of transitory common shocks.

When the imperfect individual and peer ability measures $x_{ics}$ and $w_{ics}$ are subject to transitory common shocks, the correlation between the error terms $v_{ics}$ and $u_{ics}$ will carry over to $x_{ics}$ and $w_{ics}$. Hence, the standard attenuation formulation no longer applies. Appendix A shows that, in the presence of transitory common shocks in lagged test scores, the within-school estimator of peer coefficient $\widehat{\lambda}_W$ converges as follows:

$$p\lim(\widehat{\lambda}_W - \lambda) = -\frac{\sigma_u^2 - \frac{\rho^2}{\sigma_a^2+\sigma_v^2}}{\sigma_{\bar{a}}^2 + \sigma_u^2 - \frac{\rho^2}{\sigma_a^2+\sigma_v^2}}\lambda - \frac{\sigma_a^2}{\sigma_a^2+\sigma_v^2}\frac{\rho}{\sigma_{\bar{a}}^2 + \sigma_u^2 - \frac{\rho^2}{\sigma_a^2+\sigma_v^2}}\beta \qquad (8)$$

Assuming that $\beta$ and $\lambda$ are both positive, the within-school estimator $\widehat{\lambda}_W$ underestimates $\lambda$ as both bias components in equation (8) are negative (see proof in Appendix A). The first bias component, which we refer to as the "attenuation bias," is similar to the classical attenuation bias formula except that it has an additional adjustment component $(-\frac{\rho^2}{\sigma_a^2+\sigma_v^2})$ in both the numerator and the denominator to correct for the correlation between the $v_{ics}$ and $u_{ics}$. Note that when $\rho > 0$, this attenuation bias is smaller in magnitude than the classical attenuation bias $(-\frac{\sigma_u^2}{\sigma_{\bar{a}}^2+\sigma_u^2}\lambda)$. The second component, which we refer to as the "transitory-common-shock bias," arises because the peer-level regressor $w_{ics}$ $(= a_{ics} + u_{ics})$ is negatively correlated with $\psi_{ics}$ $(= \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics})$ also through the correlation between $u_{ics}$ and $v_{ics}$ in the presence of transitory common shocks. Specifically, $cov\,(w_{ics}, \psi_{ics}) = -\lambda\sigma_u^2$ when measurement errors $v_{ics}$ and $u_{ics}$ are independent, while $cov\,(w_{ics}, \psi_{ics}) = -\beta\rho - \lambda\sigma_u^2$ when $v_{ics}$ and $u_{ics}$ are correlated. The second bias component in equation (8) can dominate the true coefficient $\lambda$ and reverse the sign of $\widehat{\lambda}_W$ when $\beta$ is sufficiently large compared to $\lambda$. Hence, our explanation of the negative and significant within-school estimator $\widehat{\lambda}_W$ is that the lagged individual and peer test scores used in our estimation are subject to transitory

only transitory effects on test scores through their effects on the measurement errors.

common shocks because of the presence of former peers in one's current peer group.

### 3.3 The IV Approach

Our foremost concern about the within-school estimator $\widehat{\lambda}_W$ is that it is subject to a transitory-common-shock bias. The empirical evidence indicates that the transitory-common-shock bias dominates the within-school estimator $\widehat{\lambda}_W$ and reverses its sign. The source of this bias is the correlation between $v_{ics}$ and $u_{ics}$ arising from the presence in a student's current peer group of her former peers, whose lagged test scores are subject to transitory common shocks just like her own lagged test score. An idea for overcoming the transitory-common-shock bias is to use the average lagged test score of a student's new peers ($w_{ics,new} = \overline{x}_{(-i)cs,new}$) as an instrument for the average lagged test score of all peers ($w_{ics}$), so as to consistently estimate an intermediate model similar to equation (7). The error term $u_{ics,new}$ in the peer-level regressor $w_{ics,new}$ is expected to be uncorrelated with the error term $v_{ics}$ in the individual-level regressor $x_{ics}$, under the assumption that transitory shocks in lagged test scores are uncorrelated for students from different primary schools. However, there remains the question of what is the formulation of the peer coefficient $\lambda^*$ we actually estimate in this IV approach and to what extent $\lambda^*$ is informative regarding the true structural coefficient $\lambda$. This is shown in Proposition 1.

**Proposition 1** *Let $y_{ics}$ denote an outcome of interest for student $i$ of cohort $c$ in school $s$; let $x_{ics}$ denote the lagged test score of student $i$, which is an imperfect measure of student $i'$s latent ability $a_{ics}$ such that $x_{ics} = a_{ics} + v_{ics}$, where $v_{ics}$ is a stochastic individual error term; let $w_{ics}$ denote the average lagged test score of student $i'$s peers such that $w_{ics} = \overline{x}_{(-i)cs} = \overline{a}_{(-i)cs} + u_{ics}$, where $u_{ics} = \overline{v}_{(-i)cs}$; and let $w_{ics,new}$ denote the average lagged test score of student $i'$s new peers such that $w_{ics,new} = \overline{x}_{(-i)cs,new} = \overline{a}_{(-i)cs,new} + u_{ics,new}$. Suppose the latent education production function takes the following linear-in-means form:*

$$y_{ics} = \beta a_{ics} + \lambda \overline{a}_{(-i)cs} + \epsilon_{ics}$$

*Assume all the covariances between $a_{ics}$, $v_{ics}$, $\overline{a}_{(-i)cs}$, $u_{ics}$, $\overline{a}_{(-i)cs,new}$, $u_{ics,new}$, and $\epsilon_{ics}$ are zero except for $cov(v_{ics}, u_{ics})$, which is denoted as $\rho$ and is assumed to be positive. Then, using $w_{ics,new}$*

*as an instrument for $w_{ics}$ can provide consistent IV estimators of the following intermediate model:*

$$y_{ics} = \beta^* x_{ics} + \lambda^* w_{ics} + \varphi_{ics} \tag{9}$$

*where* $\beta^* = \frac{\sigma_a^2}{\sigma_a^2+\sigma_v^2}\beta - \frac{\rho}{\sigma_a^2+\sigma_v^2}\frac{\sigma^2_{\overline{a}_{new}}}{\sigma^2_{\overline{a}_{new}}+\sigma^2_{u_{new}}}\lambda$, $\lambda^* = \frac{\sigma^2_{\overline{a}_{new}}}{\sigma^2_{\overline{a}_{new}}+\sigma^2_{u_{new}}}\lambda$, *and* $\varphi_{ics} = (\beta-\beta^*)x_{ics} + (\lambda-\lambda^*)w_{ics} + (\epsilon_{ics} - \beta v_{ics} - \lambda u_{ics})$. *In the preceding formulas,* $\sigma_a^2$, $\sigma_v^2$, $\sigma^2_{\overline{a}_{new}}$, *and* $\sigma^2_{u_{new}}$ *denote, respectively, the variances of* $a_{ics}$, $v_{ics}$, $\overline{a}_{(-i)cs,new}$, *and* $u_{ics,new}$.

The proof of Proposition 1 is provided in Appendix B. As lagged test scores are imperfect measures of abilities, attenuation biases remain in both $\widehat{\beta}_{W,IV}$ and $\widehat{\lambda}_{W,IV}$. However, using $w_{ics,new}$ as an instrument for $w_{ics}$ removes the transitory-common-shock bias in $\widehat{\lambda}_{W,IV}$ because the error term $u_{ics,new}$ in the instrument $w_{ics,new}$ is uncorrelated with the error term in lagged individual test score $v_{ics}$. The transitory-common-shock bias component remains in $\widehat{\beta}_{W,IV}$ as we do not have an instrument for individual lagged test scores. Despite the attenuation bias, $\widehat{\lambda}_{W,IV}$ is still informative because it is a consistent estimator of an interesting intermediate parameter ($\frac{\sigma^2_{\overline{a}_{new}}}{\sigma^2_{\overline{a}_{new}}+\sigma^2_{u_{new}}}\lambda$), which converges to the policy parameter of interest $\widetilde{\lambda}(= \frac{\sigma^2_{\overline{a}}}{\sigma^2_{\overline{a}}+\sigma^2_{u}}\lambda)$ if the information-to-noise ratio in the lagged test scores of new peers ($\frac{\sigma^2_{\overline{a}_{new}}}{\sigma^2_{u_{new}}}$) is the same as that in the lagged test scores of all peers ($\frac{\sigma^2_{\overline{a}}}{\sigma^2_{u}}$).

# 4 Empirical Results

## 4.1 First-stage Results

Results in Table 2 show a possibly serious negative transitory-common-shock bias in the within-school estimator of peer coefficient $\widehat{\lambda}_W$. Our approach to correct this transitory-common-shock problem is to use the lagged test score measures of new peers to instrument for the corresponding measures of all peers. Specifically, we are interested in the causal effects of two lagged peer test score measures: the mean and the IQR. In order to overcome the transitory-common-shock problem, we instrument the average lagged peer test score using the average lagged test score of new peers, and instrument the IQR of lagged peer test scores with the IQR of lagged test scores of new peers. Table 3 shows the first-stage relationships. All specifications include as covariates the student's own lagged test score, the proportion of female peers, a female dummy, middle school dummies, and cohort dummies. Column 1 shows that the two average lagged test score measures are highly positively correlated: a one standard deviation increase in the average lagged test score of new peers is associated with 0.56 standard deviation increase in the average lagged peer test score. The first-stage coefficient estimate is less than the average proportion of new peers in our sample (75 percent), suggesting a negative association between the average lagged test score of old peers and that of new peers.[8] This observation indicates that the school district may have some equity concern in mind when determining the middle school zoning scheme every year. In such case, studies examining solely the reduced-form effect of new peers (e.g., Gibbons and Telhaj, 2006; Lavy, Silva, and Weinhardt, 2009) may underestimate the magnitude of peer effects. Column 3 estimates the first-stage relationship between two IQR measures: a one standard deviation increase in the IQR of lagged test scores of new peers is associated with a 0.30 standard deviation increase in the IQR of lagged test scores of all peers. Columns 2 and 4 include both instruments and estimate the first-stage models, respectively, for the average lagged peer test score and the IQR of lagged peer test scores. The first-stage coefficient of the relevant instrument (i.e., the average lagged test score of new peers in column 2 and the IQR of lagged test scores of new peers in column 4) remains virtually unchanged after the inclusion of the other instrument.

### 4.2 Basic Results on Peer Effects

Table 4 summarizes our basic results on peer effects from homogeneous models. Columns 1-3 present results from reduced-form regressions and columns 4-6 report the corresponding IV results. Let us first consider the coefficients of two individual-level regressors: the female dummy and the lagged individual test score. The coefficients of both these individual-level regressors are consistently estimated across all the reduced-form and IV specifications. The coefficient of the female dummy is statistically insignificant in all specifications, indicating no significant gender gap in 9th-grade math scores. The coefficient of lagged individual test score is highly significant and is estimated to be virtually the same (around 0.43) across all the six specifications. As shown in Proposition 1, $\widehat{\beta}_{W,IV}$ puts a lower bound of the structural coefficient of own ability effect $\beta$ when the peer ability effect $\lambda$ is nonnegative.

[8] Another way to examine such a negative association is to regress the lagged average test score of old peers on that of new peers and the same set of covariantes. The estimated coefficient of the lagged average test score of new peers in that regression is -0.166 (with a standard error of 0.022).

We now turn to the results on peer effects. We are interested in three measures of peer composition: peer gender mix, average lagged peer test score (a proxy measure for the average peer ability), and IQR of the lagged peer test scores (a proxy measure for the spread of the distribution of peer ability). Unlike some previous studies that find positive spillover effects of girls on math scores (Hoxby, 2000; Whitemore, 2003; Lavy and Schlosser, 2009), we find no evidence that peer gender composition has an impact on students' 9th-grade math scores. The coefficient of the proportion of female peers variable is insignificant in all the reduced-form and the IV specifications. Columns 1 reports the reduced-form effect of the average lagged test score of new peers. Once we replace average lagged achievement of a student's peers with the same measure of her new peers, the negative peer coefficient for the within-school estimator ($-0.250$ with a standard error of 0.112) disappears. Column 4 presents the corresponding IV estimator of the average lagged peer test score. Unfortunately, both the reduced-form (0.121 with a standard error of 0.087) and the IV coefficients (0.218 with a standard error of 0.175) are very imprecisely estimated. Although we cannot reject the null hypothesis of no linear-in-means peer effects, we also cannot reject very large peer effects. The imprecise reduced-form and IV estimators are likely to be because of the relatively small number of clusters (cohorts $\times$ schools) in our sample. Despite the imprecise results, the pattern of change from the within estimator to the reduced-form and IV estimators still shows strong evidence for the existence of a severe negative bias of using the average lagged peer test score measure when this measure is subject to transitory common shocks just like the student's own lagged test score.

We next examine the effect of peer group heterogeneity, measured by the IQR of lagged peer test scores, on student achievement. Columns 2 and 5, report, respectively, the reduced-form and the IV estimates of the effect of the IQR of lagged peer test scores. Both estimators are negative and significant at the five percent level, suggesting that students benefit from having homogeneous peers. The point estimate of the IV coefficient ($-0.566$ with a standard error of 0.327) indicates that a $0.2\sigma$ reduction in the IQR of 6th-grade peer test scores, a magnitude of change 11 out 15 schools in our sample had experienced, can increase a student's test score by $0.1\sigma$. Column 3 presents the results of reduced-form estimation that includes both the mean and the IQR of lagged test scores of new peers. Column 6 shows the corresponding IV results that control for both the mean and the IQR of lagged peer test scores. The IV estimate of the coefficient of the IQR of lagged peer test scores ($-0.571$ and a standard error of 0.321) is insensitive to the inclusion of average peer test score. The point estimate of the IV coefficient of the average lagged peer test score ($-0.007$ with a standard error of 0.199), however, has been reduced substantially once we control for the IQR of lagged peer test scores. The reduction in the point estimate is due to a negative correlation between the residual average peer test score and the residual IQR of lagged test scores (after controlling for school fixed effects).

### 4.3 Allowing Heterogeneity for Peer Effects

Peer influences, however, may be heterogeneous and operate through the interaction between the distribution of peer ability and a student's own ability. For instance, some existing computational models of peer sorting in schools assumes that peer effects exhibit single crossing, i.e., an increase in average peer ability affects high-achieving students more than low-achieving students (Nechyba, 2006). In addition, several recent empirical studies find that students seem to benefit from having peers with similar characteristics as themselves, evidence in support of tracking (e.g., Hoxby and Weingarth, 2005; Duflo, Dupas, and Kremer, 2008).

We explore these alternative models of heterogeneous peer effects in this subsection by interacting measures of lagged peer test scores – the mean and the IQR in particular – with student's own lagged test scores. To implement the estimation of these heterogeneous peer effects models in an IV framework, we instrument each interaction term between measures of lagged peer test scores and a student's lagged test score with the corresponding interaction term between measures of lagged test score of new peers and the student's lagged test score. Tables 5 and 6 present the first-stage and reduced-form results of these heterogeneous models. We focus our discussion in the text on the IV results reported in Table 7. Column 1 examines whether peer effects exhibit the single-crossing property. The IV coefficient of the interaction term between the average lagged peer test score and a student's own lagged test score (0.113 with a standard error of 0.052) is positive and significant at the five percent level, evidence in support of the single-crossing property. Column 2 examines whether a change in the dispersion of lagged peer test scores in a school affects students at the middle of the school's lagged test score distribution differently than those at the two tails. To do so, we interact the IQR of the distribution of lagged peer test scores with the absolute deviation of a student's own lagged test score from the school-cohort median, and instrument this interaction term with the corresponding interaction term using the IQR of her new peers. The IV coefficient of this interaction is positive and significant at the one percent level (0.093 with a standard error of 0.024), indicating that students in the middle of the lagged test score distribution benefit most from a contraction in the spread of the distribution of lagged peer test scores. Column 3 provides estimates of the full specification that includes both the mean and the IQR of lagged peer test scores as well as their interactions with students' own lagged test scores. Results of column 3 can summarize our findings. First, a rightward shift in the distribution of lagged peer scores benefits high-achieving students relative to low-achieving students, making the overall effect of average lagged test score insignificant. Second, a mean-preserving contraction in the distribution of lagged peer scores benefits all students, but to a greater extent for those in the middle of a school's lagged test score distribution. Both of these findings are in favor of ability tracking.

## 5 Conclusion

We provide empirical evidence on the existence and the structure of peer effects in middle school using a unique longitudinal data set from China. The peer effects literatue seems to be dominated by discussions on the relection problem and the selection issues, whereas little attention is being paid to the potential correlation in measurement errors between the individual- and the peer-level regressors, which we find important in our data. Such a correlation in measurement errors would arise if we simultaneously control for the lagged individual and peer test scores as the two measures are subject to transitory common shocks due to the continuing presence of former peers in a student's current peer group. An important contribution of this paper is to clarify the impact of lagged transitory common shocks on estimates of peer effects. We derive formally that a positive correlation in measurement errors between the individual- and the peer-level regressors will lead to a negative bias in the estimate of peer coefficient, and provide empirical evidence that the transitory-common-shock problem is more than theoretical. We propose an empirical strategy to circumvent the transitory-common-shock problem by using the lagged test score measures of new peers, whose measurement error is uncorrelated with the measurement error in lagged individual test score, to instrument for the corresponding lagged test score measures of all peers.

Our main identification strategy uses within-school variation in peer composition across adjacent cohorts to control for student sorting across schools and the unobserved school characteristics that affect student outcomes. Our within-school IV estimate of the linear-in-means model shows little evidence that having peers of higher average lagged test score significantly improves a student's test score in math. The coefficients of the average lagged peer test scores, however, are not very precisely estimated. While we cannot reject the null hypothesis of no peer effects, we also cannot reject relatively large peer effects that have been found in the previous literature. Estimates of heterogeneous peer effects models show some evidence in favor of ability tracking for math learning. We find that a rightward shift in the distribution of lagged peer test scores benefits high-achieving students relative to low-achieving students, while a mean-preserving contraction in the distribution of lagged peer test scores benefits all students, but to a greater extent for those in the middle of a school's lagged test score distribution.

# 6 Appendices

## 6.1 Appendix A The Within-School Estimation

We are interested in the within-school estimation of equation (1) in the text

$$Y_{ics} = \beta A_{ics} + \lambda \overline{A}_{(-i)cs} + \phi_s + \kappa_{cs} + \upsilon_{ics} \tag{A1}$$

Taking average of equation (A1) for all students in school $s$ yields

$$\overline{Y}_s = \beta \overline{A}_s + \lambda \overline{A}_s^* + \phi_s \tag{A2}$$

where $\overline{Y}_s = \frac{1}{n_s}\sum\limits_i Y_{ics}$, $\overline{A}_s = \frac{1}{n_s}\sum\limits_i A_{ics}$, $\overline{A}_s^* = \frac{1}{n_s}\sum\limits_i \frac{n_{cs}-1}{n_{cs}} A_{ics}$, and $n_s$ and $n_{cs}$ represent, respectively, the total number of students in school $s$ and the total number of students in cohort $c$ in school $s$. Here $\overline{A}_s^*$ differs from $\overline{A}_s$ because we use leave-out average peer ability in equation (A1). Note that $\kappa_{cs}$ and $\upsilon_{ics}$ are not included in equation (A2) as they both have zero means at the school level. We can transform $Y_{isc}$, $A_{isc}$, and $\overline{A}_{(-i)cs}$ into derivations from their school means such that $y_{ics} = Y_{isc} - \overline{Y}_s$, $a_{ics} = A_{ics} - \overline{A}_s$, and $\overline{a}_{(-i)cs} = \overline{A}_{(-i)cs} - \overline{A}_s^*$. Substracting equation (A2) from equation (A1), the within-school specification of the education production function is

$$y_{ics} = \beta a_{ics} + \lambda \overline{a}_{(-i)cs} + \kappa_{cs} + \upsilon_{ics} \tag{A3}$$

Consider the following model generating the lagged test score ($X_{ics}$):

$$X_{ics} = A_{ics} + V_{ics}$$

where $V_{ics}$ is a stochastic individual error term that is uncorrelated with $A_{ics}$ and $\epsilon_{ics}$. Let $W_{ics}$ denote the average lagged test score of student $i's$ peers such that:

$$W_{ics} = \overline{X}_{(-i)cs} = \overline{A}_{(-i)cs} + U_{ics}$$

where $U_{ics} = \overline{V}_{(-i)cs}$. The within-school transformation of $X_{ics}$ and $W_{ics}$ can be written as follows

$$x_{ics} = X_{ics} - \overline{X}_s = (A_{ics} - \overline{A}_s) + (V_{ics} - \overline{V}_s) = a_{ics} + v_{ics} \tag{A4a}$$

$$w_{ics} = W_{ics} - \overline{W}_s = (\overline{A}_{(-i)cs} - \overline{A}^*_s) + (U_{ics} - \overline{U}_s) = a_{(-i)cs} + u_{ics} \tag{A4b}$$

Note that the above within-school transformation allows the possibility that $\overline{V}_s$ and $\overline{U}_s$ are nonzero. For example, if a middle school always draw students from a primary school that manipulate the test scores of its students by lowering the grading standards, $\overline{V}_s$ and $\overline{U}_s$ would both be positive. Such across-school variation in measurement errors, however, is accounted for in the within-school estimation. Substituting equations (A4a) and (A4b) into equation (A3) yields

$$y_{ics} = \beta x_{ics} + \lambda w_{ics} + \psi_{ics} \tag{A5}$$

where $\psi_{ics} = \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics}$. We assume all covariances between $a_{ics}$, $v_{ics}$, $\overline{a}_{(-i)cs}$, $u_{ics}$, and $\epsilon_{ics}$ are zero except for $cov(v_{ics}, u_{ics})$, which is denoted as $\rho$ and is assumed to be nonnegative.

Let $n$ denotes the total number of students in the sample, the plims of the variance and covariance terms are

$$\begin{aligned}
p\lim \sum \frac{(x-\overline{x})^2}{n} &= \sigma_a^2 + \sigma_v^2 \\
p\lim \sum \frac{(w-\overline{w})^2}{n} &= \sigma_{\overline{a}}^2 + \sigma_u^2 \\
p\lim \sum \frac{(w-\overline{w})(x-\overline{x})}{n} &= \rho \\
p\lim \sum \frac{(x-\overline{x})(y-\overline{y})}{n} &= \beta\sigma_a^2 \\
p\lim \sum \frac{(w-\overline{w})(y-\overline{y})}{n} &= \lambda\sigma_{\overline{a}}^2
\end{aligned}$$

The within-school estimator $\widehat{\beta}_W$ is

$$\widehat{\beta}_W = \frac{\sum(w-\overline{w})^2 \sum(x-\overline{x})(y-\overline{y}) - \sum(w-\overline{w})(x-\overline{x}) \sum(w-\overline{w})(y-\overline{y})}{\sum(x-\overline{x})^2 \sum(w-\overline{w})^2 - (\sum(w-\overline{w})(x-\overline{x}))^2} \tag{A6}$$

Taking the plim of (A6) and substituting the above plims of the variance and covariance terms yield,

$$
\begin{aligned}
p\lim \widehat{\beta}_W &= \frac{(\sigma_{\overline{a}}^2+\sigma_u^2)\beta\sigma_a^2-\rho\lambda\sigma_{\overline{a}}^2}{(\sigma_a^2+\sigma_v^2)(\sigma_{\overline{a}}^2+\sigma_u^2)-\rho^2} \\
&= \beta-\frac{\sigma_v^2-\frac{\rho^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)}}{(\sigma_a^2+\sigma_v^2)-\frac{\rho^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)}}\beta-\frac{\sigma_{\overline{a}}^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)}\frac{\rho}{(\sigma_a^2+\sigma_v^2)-\frac{\rho^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)}}\lambda
\end{aligned} \tag{A7}
$$

By the same argument,

$$
\begin{aligned}
p\lim \widehat{\lambda}_W &= p\lim \frac{\sum(x-\overline{x})^2\sum(w-\overline{w})(y-\overline{y})-\sum(w-\overline{w})(x-\overline{x})\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2\sum(w-\overline{w})^2-(\sum(w-\overline{w})(x-\overline{x}))^2} \\
&= \frac{(\sigma_a^2+\sigma_v^2)\lambda\sigma_{\overline{a}}^2-\rho\beta\sigma_a^2}{(\sigma_a^2+\sigma_v^2)(\sigma_{\overline{a}}^2+\sigma_u^2)-\rho^2} \\
&= \lambda-\frac{\sigma_u^2-\frac{\rho^2}{(\sigma_a^2+\sigma_v^2)}}{(\sigma_{\overline{a}}^2+\sigma_u^2)-\frac{\rho^2}{(\sigma_a^2+\sigma_v^2)}}\lambda-\frac{\sigma_{\overline{a}}^2}{(\sigma_{\overline{a}}^2+\sigma_\mu^2)}\frac{\rho}{(\sigma_a^2+\sigma_v^2)-\frac{\rho^2}{(\sigma_{\overline{a}}^2+\sigma_\mu^2)}}\beta
\end{aligned} \tag{A8}
$$

As $\rho^2=[cov(v,u)]^2<var(v)var(u)=\sigma_v^2\sigma_u^2$, the last two terms in (A7) and (A8) are negative when $\rho>0$. For the special case in which $\rho=0$,

$$
\begin{aligned}
p\lim \widehat{\beta}_W &= \beta-\frac{\sigma_v^2}{(\sigma_a^2+\sigma_v^2)}\beta \\
p\lim \widehat{\lambda}_W &= \lambda-\frac{\sigma_u^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)}\lambda
\end{aligned}
$$

### 6.2 Appendix B Proof of Proposition 1

We are interested in using $w_{ics,new}$ as an instrument for $w_{ics}$ to estimate the following intermediate model of interest:

$$
y_{ics}=\beta^*x_{ics}+\lambda^*w_{ics}+\phi_{ics} \tag{A9}
$$

*where* $\phi_{ics} = (\beta - \beta^*)x_{ics} + (\lambda - \lambda^*)w_{ics} + (\epsilon_{ics} - \beta v_{ics} - \lambda u_{ics})$. For $w_{ics,new}$ to be a valid instrument for $w_{ics}$, the formulations of $\beta^*$ and $\lambda^*$ need to satisfy the following two conditions:

$$cov(x_{ics}, \phi_{ics}) = 0 \quad (1A)$$

$$cov(w_{ics,new}, \phi_{ics}) = 0 \quad (1B)$$

Condition (1A) implies

$$\begin{aligned}
& cov(x_{ics}, \phi_{ics}) \\
= \; & (\beta - \beta^*)var(x_{ics}) + (\lambda - \lambda^*)cov(x_{ics}, w_{ics}) - \beta cov(x_{ics}, v_{ics}) - \lambda cov(x_{ics}, u_{ics}) \\
= \; & (\beta_0 - \beta^*)(\sigma_a^2 + \sigma_v^2) + (\lambda - \lambda^*)\rho - \beta\sigma_v^2 - \lambda_0\rho \\
= \; & \beta\sigma_a^2 - \beta^*(\sigma_a^2 + \sigma_v^2) - \lambda^*\rho \\
= \; & 0 \qquad (1A')
\end{aligned}$$

Condition (1B) implies

$$\begin{aligned}
& cov(w_{ics,new}, \phi_{ics}) \\
= \; & (\beta - \beta^*)cov(w_{ics,new}, x_{ics}) + (\lambda - \lambda^*)cov(w_{ics,new}, w_{ics}) - \beta cov(w_{ics,new}, v_{ics}) - \lambda cov(w_{ig,new}, u_{ics}) \\
= \; & (\beta - \beta^*)0 + (\lambda - \lambda^*)(1-p)(\sigma_{\bar{a}_{new}}^2 + \sigma_{u_{new}}^2) - \beta 0 - \lambda(1-p)\sigma_{u_{new}}^2 \\
= \; & (1-p)[\lambda\sigma_{\bar{a}_{new}}^2 - \lambda^*(\sigma_{\bar{a}_{new}}^2 + \sigma_{u_{new}}^2)] \\
= \; & 0 \qquad (1B')
\end{aligned}$$

The formulations of $\beta^*$ and $\lambda^*$ that satisfies both (1A') and (1B') are:

$$\beta^* = \frac{\sigma_a^2}{\sigma_a^2+\sigma_v^2}\beta - \frac{\rho}{\sigma_a^2+\sigma_v^2}\frac{\sigma_{\bar{a}_{new}}^2}{\sigma_{\bar{a}_{new}}^2+\sigma_{u_{new}}^2}\lambda$$

$$\lambda^* = \frac{\sigma_{\bar{a}_{new}}^2}{\sigma_{\bar{a}_{new}}^2+\sigma_{u_{new}}^2}$$

Therfore, using $w_{ics,new}$ as an instrument for $w_{ics}$ provides consistent IV estimates of $\beta^*$ and $\lambda^*$ as defined above.

## References


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

[2] Joshua D. Angrist and Kevin Lang. Does school integration generate peer effects? Evidence from Boston's Metco program. *American Economic Review*, 94(5):1613–1634, 2004.

[3] Julian R. Betts and Anne Danenberg. School accountability in california: An early evaluation. *Brookings Papers on Education Policy*, 2002:123–197, 2002.

[4] Julian R. Betts and Andrew Zau. Peer groups and academic achievement: Panel evidence from administrative data. Unpublished manuscript, University of California, San Diego, 2004.

[5] William A. Brock and Steven N. Durlauf. Interactions-based models. In James J. Heckman and Edward Leamer, editors, *Handbook of Econometrics*, volume 5, chapter 54, pages 3297–3367. Elsevier Science, Amsterdam, 2001.

[6] Katherine Carman and Lei Zhang. Classroom peer effects and academic achievement: Evidence from a Chinese middle school. Unpublished manuscript, Clemson University, 2008.

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

[8] Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer effects, pupil-teacher ratios, and teacher incentives: Evidence from a randomized evaluation in Kenya. unpublished manuscript, 2007.

[9] Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer effects and the impact of tracking: Evidence from a randomized evaluation in Kenya. unpublished manuscript, 2008.

[10] Stephen Gibbons and Shqiponje Telhaj. Peer effects and pupil attainment: Evidence from secondary school transition. CEE Discussion Papers 0063, Centre for the Economics of Education, LSE, May 2006. LSE Center for the Economics of Education Discussion Papers No. 0063.

[11] Eric D. Gould, Victor Lavy, and M. Daniele Paserman. Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *The Economic Journal*, Forthcoming.

[12] Eric A. Hanushek, John F. Kain, Jacob M. Markman, and Steven G. Rivkin. Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18:527–544, 2002.

[13] Caroline M. Hoxby. Peer effects in the classroom: Learning from gender and race variation. NBER Working Papers No. 7867, 2000.

[14] Caroline M. Hoxby and Gretchen Weingarth. Taking race out of the equation: School reassignment and the structure of peer effects. Unpublished manuscript, Harvard University, 2005.

[15] Thomas J. Kane and Douglas O. Staiger. The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, 16(4):91–114, 2002.

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

[17] Fang Lai. How do classroom peers affect student outcomes? Evidence from a natural experiment in Beijing's middle schools. Unpublished manuscript, New York University, 2007.

[18] Victor Lavy and Analia Schlosser. Mechanisms and impacts of gender peer effects at school. NBER Working Papers No. 13292, 2007.

[19] Victor Lavy, Olmo Silva, and Felix Weinhardt. The good, the bad and the average: Evidence on the scale and nature of ability peer effects in schools. Unpublished manuscript, Hebrew University of Jerusalem, 2009.

[20] Lars Lefgren. Educational peer effects and the Chicago public schools. *Journal of Urban Economics*, 56(2):169–191, 2004.

[21] Charles F. Manski. Identification of endogenous social effects: The reflection problem. *Journal of Economic Studies*, 14(3):115–136, 1993.

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

[23] Robert A. Moffitt. Policy interventions, low-level equilibria, and social interactions. In Steven N. Durlauf and H. Peyton Young, editors, *Social Dynamics*, chapter 3, pages 45–82. Brookings Institution Press, Washington, D.C., 2001.

[24] Thomas J. Nechyba. Income and peer quality sorting in public and private schools. In Eric A. Hanushek and Finis Welch, editors, *Handbook of the Economics of Education*, volume 2, chapter 22, pages 1327–1368. Elsevier Science, Amsterdam, 2006.

[25] Bruce Sacerdote. Peer effects with random assignment: Results from dartmouth roomates. *Quarterly Journal of Economics*, 116(2):681–704, 2001.

[26] Jacob L. Vigdor and Thomas Nechyba. Peer effects in elementary school: Learning from "apparent" random assignment. Unpublished manuscript, Duke University, 2004.

[27] Jacob L. Vigdor and Thomas Nechyba. Peer effects in North Carolina public schools. In Ludger Woessmann and Paul E. Peterson, editors, *Schools and the Equal Opportunity Problem*. MIT Press, Cambridge, MA, 2006.

[28] Diane Whitemore. Resource and peer impact on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review Papersand Proceedings*, 95(2):199–203, 2005.

[29] Jeffrey E. Zabel. The impact of peer effects on student outcomes in New York City public schools. *Education Finance Policy*, 3:197–249, 2008.

[30] Hongliang Zhang. Magnet schools and student achievement: Evidence from a randomized natural experiment in China. Unpublished manuscript, Massachusetts Institute of Technology, 2009.

[31] David Zimmerman. Peer effects in academic outcomes: Evidence from a natural experiment. *Journal of Urban Economics*, 85(1):9–23, 2003.

Table 1 Descriptive Statistics

| | Mean | s.d. | within-school s.d. |
|---|---|---|---|
| Panel A: Individual characteristics: | | | |
| Female | 0.497 | 0.500 | 0.500 |
| 6th-grade math score | 0.035 | 0.927 | 0.865 |
| 9th-grade math score | -0.005 | 1.002 | 0.914 |
| Panel B: Peer-group characteristics | | | |
| All peers: | | | |
| Proportion of female peers | 0.497 | 0.036 | 0.029 |
| Proportion of new peers | 0.749 | 0.281 | 0.151 |
| Average 6th-grade peer math score | 0.035 | 0.363 | 0.140 |
| Inter-quartile range of 6th-grade peer math scores | 1.043 | 0.233 | 0.118 |
| New peers: | | | |
| Average 6th-grade math score | 0.010 | 0.383 | 0.183 |
| Inter-quartile range of math scores | 1.065 | 0.292 | 0.210 |
| Number of observations | | 7,435 | |

Table 2 OLS and Within-School Estimation

| | Dependent variable: 9th-grade math score | |
|---|---|---|
| | OLS | Within |
| | (1) | (2) |
| Own lagged test score | 0.442*** | 0.436*** |
| | (0.019) | (0.019) |
| Average lagged peer test score | 0.591*** | -0.250*** |
| | (0.104) | (0.112) |
| Middle school fixed effects | no | yes |
| Number of observations | 7,435 | |

Notes: All specifications control for a female dummy, the proportion of female peers, and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level.

Table 3 The First-Stage Effects

| | Dependent Variables | | | |
|---|---|---|---|---|
| | Average lagged peer test score | | IQR of lagged peer test scores | |
| | (1) | (2) | (3) | (4) |
| Average lagged test score | 0.555*** | 0.546*** | | -0.058 |
| of new peers | (0.071) | (0.068) | | (0.078) |
| IQR of lagged test scores | | -0.015 | 0.297*** | 0.271*** |
| of new peers | | (0.042) | (0.060) | (0.058) |

Notes: All specifications control for own lagged test score, a female dummy, the proportion of female peers, middle school fixed effects and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level.

Table 4 Basic Results of Within-school IV Estimates of Peer Effects

| | Reduced-form | | | IV | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Female | -0.020<br>(0.023) | -0.020<br>(0.022) | -0.020<br>(0.023) | -0.019<br>(0.022) | -0.025<br>(0.024) | -0.025<br>(0.023) |
| Proportion of female peers | -0.182<br>(0.614) | 0.330<br>(0.592) | 0.313<br>(0.587) | 0.281<br>(0.696) | -0.747<br>(0.839) | -0.759<br>(0.869) |
| Own lagged test score | 0.433***<br>(0.019) | 0.432***<br>(0.019) | 0.432***<br>(0.019) | 0.427***<br>(0.020) | 0.426***<br>(0.020) | 0.426***<br>(0.020) |
| Average lagged peer test score | | | | 0.218<br>(0.175) | | |
| Average lagged test score of new peers | 0.121<br>(0.087) | | 0.030<br>(0.087) | | | -0..007<br>(0199) |
| IQR of lagged peer test scores | | | | | -0..566*<br>(0.327) | -0..571*<br>(0.321) |
| IQR of lagged test scores of new peers | | -0.168**<br>(0.080) | -0.155*<br>(0.077) | | | |

Notes: All specifications control for middle school fixed effects and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported parentheses. A triple asterisk (***) denotes significant at the 1 percent level. A single asterisk (*) denotes significant at the 10 percent level.

Table 5 The First-stage Results of Heterogeneous Peer Effects Models

| | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| Average lagged test score of new peers | 0.550*** | | 0.541*** |
| | (0.072) | | (0.067) |
| Average lagged test score of new peers * lagged individual test score | 0.821*** | | 0.821*** |
| | (0.083) | | (0.083) |
| IQR of lagged test scores of new peers | | 0.291*** | 0.265*** |
| | | (0.060) | (0.056) |
| IQR of lagged test scores of new peers * Deviation of own lagged test score from the school-cohort median | | 0.955*** | 0.953*** |
| | | (0.032) | (0.031) |

Notes: Each cell of the table reports the coefficient of regressing the corresponding instrumented variable on the instrument. All specifications control for the set of other instruments used in the model, own lagged test score, a female dummy, the proportion of female peers, middle school fixed effects, and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level.

Table 6 The Reduced-form Results of Heterogeneous Peer Effects Models

| | Dependent variable: 9th-grade math scores | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Average lagged test score of new peers | 0.140 | | 0.052 |
| | (0.090) | | (0.090) |
| Average lagged test score of new peers * Own lagged test score | 0.087** | | 0.091* |
| | (0.041) | | (0.042) |
| IQR of lagged test scores of new peers | | -0.221* | -0.206*** |
| | | (0.080) | (0.056) |
| IQR of lagged test scores of new peers * Deviation of own lagged test score from the school-cohort median | | 0.084*** | 0.086*** |
| | | (0.023) | (0.022) |

Notes: All specifications control for own lagged test score, a female dummy, the proportion of female peers, middle school fixed effects, and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level. A double asterisk (**) denotes significant at the 5 percent level. A single asterisk (*) denotes significant at the 10 percent level.

Table 7 Results of Within-school IV Estimates of Heterogeneous Peer Effects

| | Dependent variable: 9th-grade math scores | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Average lagged peer test score | 0.246 | | 0.027 |
| | (0.180) | | (0.208) |
| Average lagged peer test score * Own lagged test score | 0.113** | | 0.097* |
| | (0.052) | | (0.057) |
| IQR of lagged peer test scores | | -0.613* | -0.585* |
| | | (0.327) | (0.330) |
| IQR of lagged peer test scores * Deviation of own lagged test score from the school-cohort median | | 0.093*** | 0.094*** |
| | | (0.024) | (0.024) |

Notes: All specifications control for own lagged test score, a female dummy, the proportion of female peers, middle school fixed effects, and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level. A double asterisk (**) denotes significant at the 5 percent level. A single asterisk (*) denotes significant at the 10 percent level.

Figure 1a Within-school Variation in Peer Gender Mix

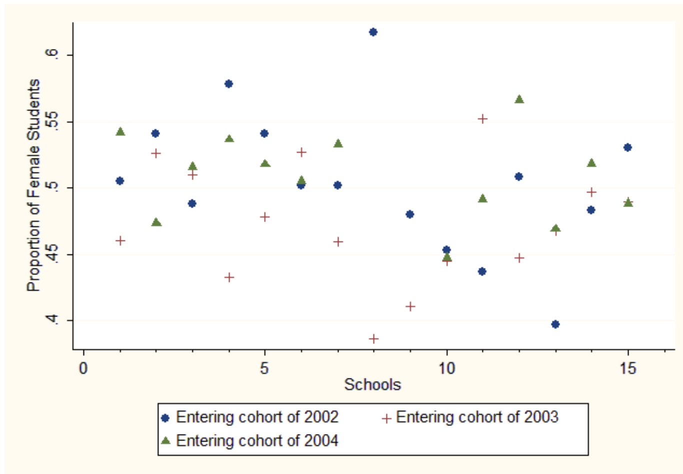


Figure 1b Within-school Variation in the Average 6th-Grade Peer Test Score

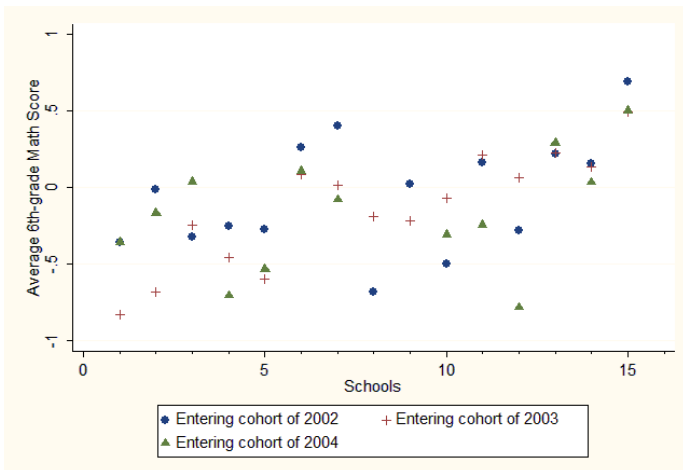


Figure 1c Within-school Variation in the Inter-Quartile Range of 6th-Grade Peer Test Scores