HCEO WORKING PAPER SERIES

# Classroom Peer Effects and Teachers: Evidence from Quasi-random Assignment in a Chinese Middle School

Leshui He
Stephen L. Ross

Working Paper

**2017-014**

02/2017

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

# Classroom Peer Effects and Teachers: Evidence from Quasi-random Assignment in a Chinese Middle School


Leshui He*[1] and Stephen L. Ross†[2]

[1]Department of Economics, Bates College

[2]Department of Economics, University of Connecticut


February 16, 2017


Abstract:

This paper examines peer effects in a Chinese middle school where 1. classes are randomly assigned to teachers and 2. student quality across classes varies because student assignment is based on a noisy measure of student quality. Peer effects are concentrated primarily on math scores, as opposed to Chinese or English scores. Improvements in peers at the bottom of the distribution of ability leads to improvements in student performance both for the student's own class and for sibling classes that share the same teachers, but are not connected in any other way. For middle and top tercile peers, improvements in peers appear to reduce student test scores. The positive effects of peers at the bottom of the distribution are primarily associated with the ability of boys in the class and the sibling class, while the negative effect of peers seems to be driven by same gender peers. Finally, the positive own class and sibling class peer effects arise primarily when the head teacher of the class or the sibling class, respectively, teaches math.



JEL Codes: I21

Key Words: Peer Effects, Teachers, Middle School, Chinese Education

## 1. Introduction

Carrell, Fullerton and West (2009) find very large peer effects on academic outcomes in higher education when examining peer groups of 30 students at the Air Force Academy who are required to spend large amounts of their time together. The findings stand in contrast to small and often insignificant effects found in studies that examine random assignment to roommates or dormitory floors in higher education (Sacerdote, 2001; Stinebrickner and Stinebrickner, 2006; Foster, 2006). The large peer effects in Carrell Fullerton and West (2009) might arise because other studies utilize much noisier proxies for individuals' actual peers or because the extensive exposure in Air Force Academy work groups causes much larger peer effects than those observed in less structured educational environments. Other than a few exceptions such as West Point companies (Lyle 2009),[1] this extensive exposure context is not representative of the typical educational experiences in the U.S. at any level other than primary education or elementary school.

On the other hand, secondary education in Chinese private schools regularly creates a similar concentrated exposure to peers environment.[2] In Chinese cities, virtually all public and private middle and high schools organize incoming cohorts of students into classes that will take virtually all of their courses together. In middle school, Chinese students typically have no electives, and while students are sorted into tracks after 10$^{th}$ grade in high school the students are re-organized after track assignment into classes that again take most if not all of their courses

[1] While companies at West Point are 120 students, these companies only include about 35 students per class, who arrive together as new cadets.

[2] In traditional, western primary and secondary school settings, Gould, Lavy, and Passerman (2009), Lavy and Schlosser (2011), Lavy, Passerman and Schlosser (2012), Friesen and Krauth (2011) and Bifulco, Fletcher and Ross (2011) find substantial peer effects exploiting across cohort variation in the composition of broad peer groups (Hoxby 2000). In an international context, some studies have found large effects of high school quality (Ding and Lehman 2007; Jackson 2010; Pop-Eleches and Urquiola 2013; Hoekstra, Mouganie and Wang 2016), but Jackson (2013) and Hoekstra, Mouganie and Wang (2016) find that peer quality differences can explain only a small fraction of such effects.

together. In the middle school that we study, each incoming cohort of seventh graders is divided into 24 classes of approximately 60 students each, and these 60 students take all of their classes together for their three years at the middle school. These students spend the vast majority of their time at school together, and have relatively minimal interactions with students in other classes. This design is used in tens of thousands of secondary schools across Chinese cities, and likely affects millions of students.

A second feature of the subject middle school is that teachers are organized into teams, and each team is assigned two classes from an incoming cohort and remain with those two classes for all three years. Increasingly studies of peer effects in education have examined evidence on the mechanisms underlying those peer effects including evidence on disruptive behavior (Lavy and Schlosser 2011, Lavy, Paserman and Schlosser 2014, Aizer 2009, Hoekstra and Carrell 2010) or imitation (Bifulco, Fletcher and Ross 2011, Bifulco, Fletcher, Oh and Ross 2014). Lavy and Schlosser (2011) and Lavy, Paserman and Schlosser (2014) are unique in that they observe information on teacher pedagogy and conclude that peers influence teachers' pedagogical practices. However, it is impossible to know from their study whether the peer effects are transmitted through changes in pedogogy or whether teachers are adjusting in this case to negative peer effects in order to minimize the peer impacts on other students. Using our data, we can examine whether peers from one class affect student outcomes in a second class where the only connection between these classes is having a common team of teachers, whose pedagogical practices, might be affected by the peers in both of their classes.

We observe data on the 2009-2013 entering year cohorts in a large, elite private middle school covering grades 7-9 in a major Chinese city. Initial assignment to classes is based on each student's cohort rank on a short placement test that is given to the entire incoming cohort, and then class composition is manually adjusted to assure virtually identical mean class scores on the placement tests, similar to company assignment at West Point (Lyle 2009). However, the class assignment process ignores a much more accurate evaluation test that is used for admission to the school. This high stakes, more comprehensive test, potentially provides a very accurate measure of average student ability over which the classrooms vary, and this resulting variation is expected to be quasi-random conditional on the placement test because the admissions test is not used in classroom assignment. To validate this identification strategy, we demonstrate that 1. the within cohort variation in classroom average placement test is less than one tenth of the equivalent variation on average admissions tests, and 2. we show that most teacher and student attributes are independent of classroom average admissions test scores conditional on the student's rank on the placement test.

Based on Lavy, Paserman and Schlosser's (2014) evidence of heterogeneous peer effects, we measure the average peer composition on the admissions test for the bottom, middle and top tercile of the score distribution in each class. Similar to their findings, we find positive effects associated with the average test scores of the first tercile of peers in each class. Specifically, we find that a change from the $25^{th}$ to the $75^{th}$ percentile on average admissions test score of first tercile peers is associated with an increase of 4 percent of a standard deviation on end of year seventh grade test scores. Further, when we include the peer composition of the sibling class, we find significant effects of both first tercile own peers and first tercile sibling class peers on all test scores examined. The estimated effects are very similar in magnitude across tests and between the own and sibling peer variables ranging between 3 and 7 percent of a standard deviation for a change in exposure from the $25^{th}$ to $75^{th}$ percentile class. The peer effects are concentrated primarily on math scores and to some extent on English and physics scores.

We also examine gender differences. We find larger estimated effects for boys, but we still find positive, but less statistically precise, peer effects for girls. Further, when the model is estimated separately by gender, some evidence of negative peer effects associated with the second and third ability terciles appears for both girls and boys and for both own and sibling class peers. For boys, the negative effect appears concentrated on the second tercile average test scores, while for girls the negative effect is associated with the third tercile or the highest ability students in the class and sibling class. This evidence of negative peer effects on performance are consistent with findings of Murphy and Weinhadt (2014) concerning the negative effects of having a low rank among classmates. Additional analyses suggest that the positive effect of peer quality of the first tercile of students is primarily associated with the average test scores of boys in both the class and the sibling class, consistent with earlier findings by Lavy and Schlosser (2011).

Next, we test whether these peer effects differ based on the identity of the head teacher since the head teacher has the most contact with the class. Focusing on the Math tests, we examine whether having the Math teacher being assigned as the head teacher or as the head teacher of the sibling class mediates the estimated peer effects. For both own and sibling class peers, the first tercile effects are concentrated in classes where the head teacher teaches math, consistent with the head teacher playing a significant role in transmitting the effects of peers within and between classrooms.

## 2. Empirical Context

### *The middle school*

The data used in this study include the complete administrative academic records from one particularly large elite middle school (henceforth the middle school) located in a major city in

China.[3] The city has a population of about 3 to 6 million people, a typical size of major cities at the provincial level in China. This school is different from regular middle schools for two reasons—its semi-private nature and its unusually large scale. This middle school is one of the top elite middle schools in the entire province.

Founded in the 1920s, this school is an affiliated middle and high school (grades 7-12) with a large research university. Such semi-public schools are prevalent in all major cities in China with most affiliated with universities, research institutions and large state-owned enterprises. These affiliated school systems can range from pre-K to high school. In an earlier time, they primarily served as dedicated care and education providers exclusively for the employees of their parent organizations. Overtime, many such schools, especially those affiliated with universities, developed significant reputations as very high quality schools. Unlike public schools that are under the direct management of local education officials, these semi-public schools are managed by their parent organization, and are consequently unrestricted by public school district assignment policies. Thus, these schools have the ability to impose admission tests and selectively enroll outstanding students from the general public, a feature resembling that of elite exam schools in many U.S. cities such as New York and Boston (Abdulkadiroglu, Angrist and Pathak 2014).

***Graduation test and admission***

This middle school enrolls 1400 to 1500 students each year to its main campus. Our sample covers all students during grades 7 to 9 for the incoming cohorts of 2009 through 2013. The middle school system in China teaches grades 7 to 9, while high school teaches grades 10 to 12. At the end of the $9^{th}$ grade, all graduating students take a municipal level high-school entrance exam (HSEE), or *zhongkao*, which is a very high-stakes exam serving as the sole selection criterion for

---

[3] The specific location of the school cannot be disclosed for confidentiality reasons.

high school admission.[4] All students from the same city take identical exams at exactly the same time. The exam typically lasts 2 to 3 days, testing 2 to 3 subjects each day. Similar to the more well-known national level college entrance exam, or *gaokao*, in China, all students who submit preferences for the same objective school are first ranked according to the exam score, then lower ranking students are rejected for admission once higher ranking students have filled all admission slots. Such pressure on the selection system led to exceptional demand for high-performing middle schools, as well as strong incentives for middle school administrators to coordinate curriculum across classrooms to maximize HSEE outcomes.

Each year in the summer, this middle school accepts open applications from the general public. Out of the 8000 to 9000 applicants each year, only 1400 to 1500 students are admitted due to space limitations. To cope with such enormous enrollment pressure, the school administers a strict admissions test for all grade-6 applicants in a procedure similar to that of HSEE—Chinese and math are the only two subjects tested. The students are then admitted based on their rank on this test. Although all students affiliated with the parent organization receive automatic admission, most still take the admissions test. Some affiliated parents use the test as a tool to evaluate their children's performances; some more extreme parents withhold this "free-pass" policy from their children to incentivize effort in preparation for middle school.

[4] Some students earn bonus scores on top of their total HSEE score due to outstanding performances in middle school, social activities or awards, but such bonus scores are relatively insignificant compared to weight given to the total HSEE score.

***Quasi-random assignment***

After admission, students in each cohort are assigned to one of 24 classes with sizes of around 60 students. To avoid conflicts among teachers in the assignment process, the school adopts a double-blind quasi-random assignment process.

On the student side, the school administers yet another smaller, less accurate placement or pre-test among all admitted students, which in recent years tests three subjects—Chinese, math and English.[5] The students are ranked within gender based on the pre-test, and then sorted into classes based on that rank. Sorting takes place in the following way. Classes are labeled as A-X. While assignment rules vary slightly by year, the primary rule is based on the following rotation pattern for each gender: the first ranked student by gender is assigned to A, the second ranked to class B, until the 24th and 25th are both assigned to class X. Similarly, the 23rd and 26th students are assigned to class W. This rule continues in a zig-zag pattern of assignment with for example the 47th in class B, the 48th and 49th in class A, and then the 50th in class B. However, because this strategy does not yield perfectly balanced classes on the pre-test, the school provost re-arranges students after this assignment to assure very similar class averages on the pre-test for the entire cohort of students. While parents can influence the school and sometimes successfully cause their kids to be re-assigned after this process is completed, we have the original assignments prior to any intervention of parents. These original assignments are used for the calculation of measures of average peer ability for those classrooms, producing intent-to-treat estimates.

On the teacher side, the provost assigns 24 head teachers, a role similar to a dean of students, and pairs these teachers in advance. The head teachers teach one of the subjects, while also

[5] Unlike the admissions test, the pre-tests are written in a much shorter time frame, in a shorter format and graded faster with no partial credit.

disciplining students in their class and communicating directly with the parents. The majority of head teachers teach Chinese or Mathematics. Each head teacher participates in a raffle drawing a letter from A to W. The letters drawn by each pre-assigned pair of teachers creating the pairs of sibling classes. These classes are both taught by the same team of teachers across most subjects. Each head teacher is then assigned a classroom from 1 to 24.

***Teachers, peers and sibling-classes***

Table 1 summarizes the subjects students take betweem 7th and 9th grade. From the 7th grade forward, students study 8 subjects—Chinese, math, English, social study, biology, geography, history, IT/computer—plus physical education (PE). From the 8th grade, they start taking physics, and from the 9th grade, chemistry is introduced to the course plan. These subjects take very different weights in the HSEE, where Chinese, math and English each account for 20% of the total score, physics for 13.3%, chemistry and PE each for 8.3%. The remainder is divided evenly between social studies, biology, geography, history, IT/computer, and music/arts each only accounting for 1.6% and are graded on a pass/fail basis.[6] These weights in the HSEE exam determined the emphasis on different subjects during middle school, thus driving the allocation of contact hours within the classroom, as well as the different teaching loads for the teachers. Chinese, math and English courses are assigned the most contact hours. Teachers of these subjects only teach 2 classes in the same cohort. Physics and chemistry teachers teach 3 classes, and teachers for other subjects teach more classes, but with less hours for each class.

Each teacher in this school typically works with the same group of students throughout their entire middle school experience. All classes and subjects are mandatory, and students stay in

[6] These score weights are based on 2016 policy. The exact weights varied slightly in our sample period from 2012 to 2016, but remain largely similar.

the same classroom except for lab or PE classes. These unique features of Chinese middle schools lead to two facts that serve as the cornerstones of our identification strategy: large and well-identified peer groups throughout the students' time in middle school, and well-defined *sibling-class* groups who have virtually no direct academic interactions, but share common teachers. We do not observe the pre-paired sets of teachers, and so we use having a common math teacher to define sibling classes. Chinese, math and English teachers only teach two courses for a given cohort, and so they are virtually always assigned to sibling classes and their assignments overlap for the vast majority of classes. These three subjects also represent a substantial majority of the weight in the HSEE total score.[7]

Even if students do not directly respond to each other's behavior, peer effects can still arise through teacher's behavior. If a teacher has a smart student in her class, she may raise her expectations for others and become less patient; she may teach faster to accommodate the outstanding student; or she may spend more/less time with this talented student and leave less/more time for other students. In any of these theories, exposure to a better peers may have an effect, although not directly through peer to peer interactions. We call such teacher-transmitted peer effects *indirect peer effects*.

In practice, it is very difficult to tease out indirect peer effects from direct peer effects arising from student interactions because peers who share the same teachers also interact with each other. However, our empirical context provides a unique opportunity to separately identify indirect peer effects. In most Chinese secondary schools, including this one, students from different classes have virtually no direct academic interactions. Thus, students in one class should have no impact

[7] Using sibling classes defined by Chinese and English yield very similar score measures and consequent regression results.

on those in another class. However, if two classes share a common teacher who behaves differently in response to students in one class, her behavior may spillover to the other class. Thus estimating effects of sibling-class student quality on student performance could provide evidence for indirect peer effects.

***Data and Empirical Strategy***

Our sample includes all students in this middle school from the incoming cohorts of 2009 to 2013: 6510 students in 5 cohorts. The data contain the students' placement or pre-test score, admissions test score, original class-assignment letter, classroom number, all end of year test scores by subject throughout the entire middle school career including the score on the HSEE. We also observe teacher names for each subject, graduating elementary school, family background, including parents' workplaces and job titles, and basic demographic information, including gender, birth day, ethnicity, etc.

Table 2 illustrates the variation that we use to identify the model. The table shows the variance of the standardized total pretest score (placement test), the math and Chinese subscores, and the admissions test on which our peer analysis is based. The first column presents the within cohort variance across students. The second column shows the variation across assigned classes. Since classrooms have been balanced based on ability, the within class variance is quite smaller in all cases. Most significantly, since the class balance is based on the pretest average, the variance of the pre-test plummets in comparison to the admissions test with the pre-test variance smaller by a factor of 10. The greater variation across classes in average admissions test scores is the source of variation that we will use to identify the effects of peers.

Given the large class sizes and previous findings suggesting that positive peer effects are associated with the bottom of a class' ability distribution, we calculate the 1st, 2nd and 3rd tercile means of the admissions scores for each originally assigned class. The means and standard deviations of these variables are shown in Table 3 panel a. The across class variation in the tercile means is substantially larger than the variation in the admissions test class averages. However, some of this variation may arise due to the zig-zag assignment process and the follow-up equalization of class mean pre-test scores. In order to focus on idiosyncratic variation and remove variation associated with the systematic process for creating classes, we will include class letter fixed effects in all models.

We primarily examine five test score outcomes. First, we examine combined Chinese and math scores on the HSEE, a high-stakes test used for high school admissions. This test provides a very precise measure of short-term academic achievement that reflects the full effort of all involved parties. Moreover, because the exam is highly coordinated, intensively monitored and systematically graded, the scores are directly comparable across students in each cohort. We also examine end of year combined Chinese and math exam scores, and the average of these end of year exams over all three middle school years, a measure close to the student's graduating GPA. We primarily focus on combined Chinese and math scores because the admissions test that we use to measure peer ability covers Chinese and math scores, and we do not observe the scores on the individual sections, math and Chinese, of the admissions test. The means and standard deviations of these variables along with observed student and teacher attributes are also shown in Table 3 panels b and c.

Given that the noisier pre-test is the only factor in class assignment, the assignment process randomly exposes students in the same cohorts to peers of different quality—students assigned to seemingly identical classes in terms of pre-tests that have in fact been exposed to classes with peers of different quality measured by the more accurate admissions test. This feature in the assignment process creates a situation where exposure to peers (at least the intent to treat exposure) is independent of student and family attributes.

We conduct two balancing tests to validate this identification strategy. First, in Table 4, we regress each teacher attribute on dummy variables for both class assignment letters and classroom numbers where each table entry presents the significance level associated with an F-test for whether the entire set of estimates on either class letter or classroom number dummy variables are zero. The F-tests on whether the class letters can explain observable teacher attributes are all statistically insignificant, supporting the school administration's description of the process where teachers are assigned to class letters via a raffle. For classroom number, one of the 18 F-tests rejects the null at better than a 5% significance level, and two of the tests reject the null at better than 10%. This suggests that assignment of teachers to classroom may not be entirely random. Given that rejections of the null are associated with experience, age and having tenure, more senior teachers may have the opportunity to selection better classrooms. All models control for classroom number fixed effects to address any systematic sorting of teachers or class letters into specific classrooms.

The resulting model specification is

$$y_{ijkc} = \beta \bar{a}_{jkc} + \gamma \bar{s}_{jkc} + r_{ijkc} + f(a_{ijkc}) + \theta X_{ijkc} + \alpha_j + \delta_k + \mu_c + \epsilon_{ij} \qquad (1)$$

where $y_{ijkc}$ is the in-school test score of student $i$ in class letter $j$ and classroom $k$ for a given admissions cohort of students $c$, $\bar{a}_{jkc}$ is the vector of tercile mean admissions test scores of the student's class letter/number for their cohort, $\bar{s}_{jkc}$ is the vector of tercile mean admissions test scores for the sibling class of the student's class, $r_{ijkc}$ is the students within cohort rank on the pretest, $f$ is a fifth order polynomial of the admissions test score $a_{ijkc}$, $X_{ijkc}$ represents the inclusion of student and teacher attributes, the next three terms represent class letter, class number and cohort fixed effects, respectively, and the final term is an idiosyncratic error.[8] Student characteristics include the students' pretest math score, gender, age at admission, majority Han-ethnicity indicator, an indicator of whether the student is eligible for admission without admissions test, indicators of whether one of the parents work in academic, in government, or hold a higher leader/administrator/manager title, such as CEO, head of local government bureau, etc. Teacher characteristics include title (coded as 1-4 based on teacher's title rank), gender and total years of teaching career for head teacher, Chinese and math teachers. Standard errors are clustered at the class level.

Table 5 reports the results of a set of balancing tests where we replace the dependent variable with the class mean and tercile mean test scores on the admissions test $\bar{a}_{jkc}$ and drop own and sibling test scores from the list of regressors, $\bar{a}_{jkc}$ and $\bar{s}_{jkc}$. Column 1 presents the estimates on $X_{ijkc}$ where the dependent variable is the class average admissions test scores, and columns 2-4 present the estimates on $X_{ijkc}$ for the class tercile averages of the admissions tests scores. The F-tests on student and teacher attributes separately are presented at the bottom of the Table. For first tercile test score means, the coefficients on admin test waived and parent job academic are statistically significant, but the F-test for significance on all student attributes is only statistically significant at the 0.147 level. For the second and third tercile test score means, however, the coefficient on teacher female is large and highly significant, and the F-test on all attributes is significant at the 0.071 and 0.014 levels, respectively. No other attributes are significant at the 5

[8] Bifulco, Fletcher and Ross (2011) explain the importance of including own test score in balancing tests of these types, otherwise the balancing tests will suffer from the bias identified by Guryan, Kroft and Notowidigdo (2009). Given the different forms used for controlling for own test score in different papers, we use a higher order polynomial function to provide a relatively general conrtrol.

percent level, and all estimates other than teacher female are small in magnitude. While the result on teacher female is surprising, the sample passes balance on all other student and teacher attributes.

## 3. Empirical Model and Findings

### *Peer and sibling effects*

Table 6 reports the intent-to-treat estimates for the combined Chinese and math scores for the HSEE exam, for the mean of the end of year exam for each grade, and for the end of year exams separately for each grade. We focus on outcomes on Chinese and math combined scores because out peer ability measures are based on combined Chinese and math scores on the admissions test. The point estimates on the $1^{st}$ or bottom tercile are all positive, and the estimates are statistically significant for the average final exams in $7^{th}$ grade (average of finals in fall and spring semesters) and for the average across all grades. In terms of magnitidue, a move from a class at the $25^{th}$ percentile of the cohort to a class at the $75^{th}$ percentile on $1^{st}$ tercile mean standardized admissions test score implies a 0.22 improvement in mean score, and so based on the estimate of 0.190 for $7^{th}$ grade end of year scores this proposed change across classes is associated with a 4.2% increase in student test scores. The bottom of the Tables reports the p-value of the joint F-test for the coefficients on the three tercile variables in order to address concerns that our model is conducting tests on three different controls. For the $7^{th}$ grade tests, the null of no peer effects for the three tercile variables is rejected with a confidence level of 0.069.

Next, we add additional controls for the tercile mean test scores of the sibling class. These results are shown in Table 7. We now observe positive, statistically significant $1^{st}$ tercile peer effects on all five test scores with magnitudes ranging from 3.7 to 6.5 percent of a standard deviation for a $25^{th}$ to $75^{th}$ percentile change in peer quality among the weakest classmates. While the point estimates of the peer effects for sibling classes always exceed the estimates for own class, the differences in the estimates are well within the estimation error. It is notable that we now observe sizable and consistently negative effects of the peer quality of students in the top tercile of peers, but these effects are only statistically significant for two of the 10 tests conducted (5 tests by own vs. sibling class). Many of the estimates on the second tercile peer means are sizable and negative, as well.

***Heterogeneity in peer and sibling effects***

Table 8 shows the estimates of peer and sibling effects on Chinese and math separately in panels a and b, respectively. The format of the tables is identical to the Tables presented above. Panels (a) and (b) show starkly different results with no peer or sibling effects in Chinese test scores, but strong and significant effects in math scores with own 1$^{st}$ tercile peers and sibling class 1$^{st}$ tercile peers having effects between 4.9 and 8.2 percent of a standard deviation higher math test scores for a 25$^{th}$ to 75$^{th}$ percentile change. The F-tests are highly statistically significant for peer and sibling peer effects for most of the outcome variables. Again, we find some evidence of negative effects on the math test of 3$^{rd}$ tercile peer quality for both own and sibling class.

Table 9 investigates other major individual subjects in middle school and the HSEE—English, physics and chemistry. We find sizable effects of both own and sibling class 1$^{st}$ tercile peers on the 8$^{th}$ and 9$^{th}$ grade physics exams, panel b. The statistical evidence for peer effects on English and Chemistry scores are significantly weaker.

Table 10 re-estimates the models in Table 7 separately for boys (Panel a) and girls (Panel b). The strongest peer and sibling effects appear to be concentrated among boys with 1$^{st}$ tercile effect sizes from a 25$^{th}$ to 75$^{th}$ percentile move ranging between 3.9 and 11.6 of a standard deviation. The largest effects are clearly associated with sibling class peers. Further, we find much stronger evidence of negative peer effects associated with middle tercile peers especially on the end of year

9th grade exam and on the HSEE, again with the strongest effects arising for the sibling class peers. The 2nd tercile peer mean test score of the sibling class (25th to 75th percentile move of 0.18) decreases HSEE scores by 11.4 percent of a standard deviation, while the same change for own class 2nd tercile peers implies a 6.7 percent decrease. For girls in panel b, results are weaker both in magnitude and statistically, but we observe a very similar pattern with better 1st tercile peers being associated with higher test scores and in the case of girls better 3rd tercile peers being associated with lower test scores.

While the resulting estimates become substantially noisier, we also estimate models that include controls for tercile means based separately the boy and girl peers in each class. Focusing on our strongest estimates, Table 11 presents results with separate peer effects of boys' and girls' average test scores on boys' and girls' math test scores. The estimates suggest that most of the effects of 1st tercile average test scores arise from boy peers regardless of the students' gender, while any negative effects of higher tercile average test scores, 2nd tercile for boys and 3rd tercile girls, appear to arise from same gender peers. However, given the magnitude of the estimated standard errors, these results are primarily suggestive.

***Role of head teachers***

In order to further investigate the role of teachers, we focus on the teachers who are assigned as head teachers. Due to the primary status of Chinese and math in the HSEE, most head teachers are drawn from those teaching these two subjects (74 out of 120 classes, or 62%).[9] A head teacher basically acts as a Dean of students for their specific class. The head teachers have much stronger authority over students in their own classes, and the students' academic performances are

---

[9] Head teachers teaching either Chinese, math or English accounts for 106 out of the 120 classes in our sample, or 88%.

also tied more closely to the teacher's job evaluation and performances. It is thus reasonable to hypothesize sibling class effects being concentrated in the head teacher's subject. For example, students with a math teacher as their head teacher may be subject to different sibling effects in math compared to students whose math teacher is the head teacher of the sibling class, or comparing to students whose math teacher is not a head teacher of either class. Given that our peer effects are concentrated in performance on math tests, we examine tercile peer effects on math test performance and allow those effects to vary by whether your head teacher is your math teacher (34 out of 120 classes, or 28%), the head teacher of your sibling class is your math teacher, or your math teacher is not the head of either your class or your sibling class. The omitted category contains classes where the math teacher is the head teacher.

These intent-to-treat results are shown in Table 12. The first six rows present the estimates on the tercile peer variables for own and sibling class, capturing the peer effects for the omitted category. The next six rows present the interaction of the math teacher not being a head teacher for either class with the tercile peer variables, and the last six rows present the interaction with having the math teacher be the head teacher of the sibling class. The first feature of these estimates is that the positive own class $1^{st}$ tercile peer effects (row 1) arise for omitted category, the class' head teacher teaches math, and the $1^{st}$ tercile estimates for own class peers for both interactions (rows 7 and 13) are consistently negative and comparable in magnitude to the row 1 estimates. Own class peer effects on math scores primarily arise when the class head teacher also teaches math. Similarly, the sibling class $1^{st}$ tercile estimates are small and insignificant when the head teacher teaches math, and the sizable effects of sibling class $1^{st}$ tercile peers are only observed on the interaction with whether the sibling class head teacher teaches math. Therefore, the majority of the math score peer effects associated with $1^{st}$ tercile peers (both own and sibling class effects) arise when the head teacher for those peers also teaches math. Admittedly, the statistical pattern is weak for some of these results with the F-tests on the relevant interaction terms, rows 7-9 and rows 13-15, only being significant at the 5% level for the HSEE exam.

**4. Summary and Conclusion**

In this paper, we find positive effects associated with the average test scores for the first tercile of students in each class of a large, private Chinese middle school where students remain together in their assigned classes for all three years. A change from the 25th to the 75th percentile on average admissions test scores of first tercile peers is associated with an increase of 4 percent of a standard deviation on end of year seventh grade test scores. Further, when we include the peer composition of a sibling class that shares the same teachers, we find significant positive effects of both first tercile own peers and first tercile sibling class peers with estimated effects ranging between 3 and 7 percent of a standard deviation for change in exposure from the 25th to 75th percentile. These peer effects are concentrated primarily on math scores and to some extent English and Physics scores. However, none of the estimated effects compare in magnitude to the peer effects identified in Carrell, Fullerton and West (2009).

When we divide the sample by gender, we find some evidence of negative peer effects associated with the average scores of the second and third terciles for both girls and boys and for both own and sibling class peers. For boys, the negative effect appears concentrated on the second tercile average test scores, while for girls the negative effect is associated with the third tercile or the highest ability students in the class and sibling class. The positive effects of lower tercile peers persist for both boys and girls, but the effect size is larger for boys. Further, these positive peer effects of the bottom tercile of students are primarily associated with the average test scores of boys for both the class and the sibling class, regardless of whether the student affected is male or female.  On the other hand, the negative effects associated with the higher terciles appear to arise from the peer quality of same gender students.

Finally, we test whether these peer effects differ based on the identity of the head teacher focusing on math scores because that is where we saw the largest peer effects.  The math score effects of own class first tercile peers primarily when the head teacher also teaches math.  Similarly, the first tercile sibling class peer effects are concentrated in classes where the sibling class is taught math by the head teacher.  In both cases, the teacher appears to play an especially significant role in transmitting the effects of peers when the teacher is the head teacher for those peers.

**References**


Abdulkadiroglu, A., Angrist, J., Pathak, P., 2014. The Elite Illusion: Achievement effects at Boston and New York Exam Schools. Econometrica 82(1): 137-196.

Aizer, Anna. 2009. Peer Effects and Human Capital Accumulation: The Externalities of ADD. NBER Working Paper #14354.

Bifulco, Robert, Fletcher, Jason and Stephen L. Ross. 2011. The effect of classmate characteristics on individual outcomes: Evidence from the Add Health. American Economic Journal: Economic Policy 3 (1): 25-53.

Bifulco, Robert, Fletcher, Jason, Sun Jung Oh and Stephen L. Ross. 2011. Do High School Peers Have Persistent Effects on College Attainment and Other Life Outcomes? Labour Economics 29, 83-90.

Carrell, Scott E., Richard L. Fullerton and James E. West. 2009. Does Your Cohort Matter? Measuring Peer Effects in College Achievement. Journal of Labor Economics, 27 (3), 439-464.

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

Foster, G. (2006). It's not your peers, and it's not your friends: Some progress toward understanding the educational peer eect mechanism. Journal of Public Economics 90 (8-9), 14551475.

Friesen, Jane and Brian Krauth. 2011. Ethnic Enclaves in the Classroom. Labour Economics 18 (5): 656-66.

Gould, Eric, Victor Lavy, and Daniele Paserman. 2009. Long term classroom peer effects: Evidence from random variation in enrollment of disadvantaged immigrants. Economic Journal, 119 (540), 1243-69.

Guryan, Jonathan, Kory Kroft, and Matt Notowidigdo. “Peer effects in the workplace: Evidence from random groupings in professional golf tournaments.” American Economic Journal: Applied Economics 1, no.4 (2009): 34-68.

Hoekstra, Mark L. and Scott Carrell. 2010. Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone. American Economic Journal: Applied Economics 2 (1), 211-28.

Hoxby, Caroline. Peer effects in the classroom: Learning from gender and race variation. No. w7867. National Bureau of Economic Research, 2000.

Jackson, C.K, 2010. Do Students Benefit From Attending Better Schools? Evidence From Rule-based Student Assignments in Trinidad and Tobago. Economic Journal 120(549): 1399-1429.

Jackson, C.K, 2013. Can higher-achieving peers explain the benefits to attending selective schools? Evidence from Trinidad and Tobago. Journal of Public Economics 108: 63-77.

Lavy, Victor, M. Daniele Paserman and Analia Schlosser. 2012. Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. The Economic Journal 122, 208–237.

Lavy, Victor, and Analia Schlosser. "Mechanisms and impacts of gender peer effects at school." American Economic Journal: Applied Economics 3, no. 2 (2011): 1-33.

Lyle, David S. 2009. The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point. American Economic Journal: Applied Economics 2009, 1:4, 69– 84.

Murphy, Richard and Felix Weinhadt. 2014. Top of The Class: The Importance of Ordinal Rank. CESIFO Working Paper #1485.

Pop-Eleches, C., Urquiola M., 2013. Going to a Better School: Effects and Behavioral Responses. American Economic Review 103(4): 1289-1324.

Sacerdote, B. 2001. Peer eects with random assignment: results for dartmouth roommates. The Quarterly Journal of Economics 116 (2), 681-704.

Stinebrickner, R. and T. R. Stinebrickner (2006). What can be learned about peer effects using college roommates? evidence from new survey data and students from disadvantaged backgrounds. Journal of Public Economics 90 (8), 1435-1454.

Zhang, H. In Press. Identification of Local Average Treatment Effects under Imperfect Matching: with an Application to Chinese Elite Schools. Journal of Public Economics.

Table 1: Student Subjects by Grade and Teachers Course Loads by Subjects

<table>
<tr><th colspan="3">grade</th><th rowspan="2">weight in HSEE</th><th rowspan="2">teaching load</th></tr>
<tr><th>7</th><th>8</th><th>9</th></tr>
<tr><td colspan="3">Chinese</td><td>120/600 (20%)</td><td>2</td></tr>
<tr><td colspan="3">math</td><td>120/600 (20%)</td><td>2</td></tr>
<tr><td colspan="3">English</td><td>120/600 (20%)</td><td>2</td></tr>
<tr><td colspan="3">social studies</td><td>10/600 (1.3%)</td><td>6-7</td></tr>
<tr><td colspan="3">history</td><td>10/600 (1.3%)</td><td>6-7</td></tr>
<tr><td colspan="3">geography</td><td>10/600 (1.3%)</td><td>4</td></tr>
<tr><td colspan="3">biology</td><td>10/600 (1.3%)</td><td>4-5</td></tr>
<tr><td colspan="3">IT/Computer</td><td>10/600 (1.3%)</td><td>6-14</td></tr>
<tr><td colspan="3">music/arts</td><td>10/600 (1.3%)</td><td>6-14</td></tr>
<tr><td colspan="3">gym</td><td>50/600 (8.3%)</td><td>8</td></tr>
<tr><td></td><td colspan="2">physics</td><td>80/600 (13.3%)</td><td>3</td></tr>
<tr><td colspan="2"></td><td>chemistry</td><td>50/600 (8.3%)</td><td>3</td></tr>
</table>

Note: The table presents middle school subjects by grade taught. The second column shows the weight given to each subject on the HSEE, citywide high school entrance exam, and the third column shows the teacher course load by subject at the middle school.

Table 2: Variation in Test Scores, Demeaned by Cohort

| score | variance of student score | variance of class mean score |
|---|---|---|
| TTL Admin, std. score | 0.96 | 0.0094 |
| TTL PreT, std. score | 0.91 | 0.0006 |
| CHN PreT, std. score | 0.86 | 0.0059 |
| MTH PreT, std. score | 0.90 | 0.0024 |

Note: The within cohort variance of the admissions test, the pre-test, and the subscores of the pre-test are shown in this Table. The second column presents the variance of the student level scores, and the third column presents the variance of the class means.

Table 3: Summary Statistics

(a) Key Independent Variables

| | mean | sd | min | max |
|---|---|---|---|---|
| Admin Total, class peer 1Tercile mean std. score | -1.16 | 0.16 | -1.67 | -0.56 |
| Admin Total, class peer 2Tercile mean std. score | 0.13 | 0.13 | -0.23 | 0.45 |
| Admin Total, class peer 3Tercile mean std. score | 1.01 | 0.10 | 0.77 | 1.29 |
| Admin-test pseudo-class peer 1Tercile mean std. score | -1.17 | 0.16 | -1.68 | -0.61 |
| Admin-test pseudo-class peer 2Tercile mean std. score | 0.13 | 0.13 | -0.23 | 0.45 |
| Admin-test pseudo-class peer 3Tercile mean std. score | 1.01 | 0.10 | 0.74 | 1.33 |
| Admin-test Sibling-class 1Tercile mean std. score | -1.16 | 0.17 | -1.61 | -0.63 |
| Admin-test Sibling-class 2Tercile mean std. score | 0.13 | 0.13 | -0.21 | 0.42 |
| Admin-test Sibling-class 3Tercile mean std. score | 1.01 | 0.10 | 0.81 | 1.26 |
| Admin-test pseudo-sibling-class 1Tercile mean std. score | -1.17 | 0.16 | -1.65 | -0.68 |
| Admin-test pseudo-sibling-class 2Tercile mean std. score | 0.13 | 0.13 | -0.21 | 0.42 |
| Admin-test pseudo-sibling-class 3Tercile mean std. score | 1.01 | 0.10 | 0.77 | 1.29 |
| Observations | 6510 | | | |

Note: The table presents the mean, variance, minimum and maximum of key variables with panel a containing the class and assigned class mean for the bottom, middle and top terciles of students for both the class and the sibling. Due to uniformity in class size, the descriptive statistics are virtually identical between own and sibling classes. Panel b presents descriptive statistics for all end of year tests administered during middle school, and Panel c presents statistics on student and assigned teacher attributes.

(b) Dependent Variables

| | mean | sd | min | max |
|---|---|---|---|---|
| HSEE CHN&MTH | 0.03 | 0.96 | -7.18 | 1.70 |
| g7-g9 FN Mean CHN&MTH | 0.08 | 0.84 | -4.92 | 1.51 |
| g7 FN Mean CHN&MTH | 0.09 | 0.84 | -5.46 | 1.68 |
| g8 FN Mean CHN&MTH | 0.08 | 0.89 | -5.06 | 1.73 |
| g9 FN Mean CHN&MTH | 0.04 | 0.92 | -5.62 | 1.65 |
| HSEE Chinese | 0.02 | 0.98 | -5.46 | 2.26 |
| g7-g9 FN Mean Chinese | 0.06 | 0.79 | -4.17 | 1.86 |
| g7 FN Mean Chinese | 0.08 | 0.82 | -4.55 | 2.45 |
| g8 FN Mean Chinese | 0.06 | 0.88 | -4.59 | 2.30 |
| g9 FN Mean Chinese | 0.03 | 0.89 | -5.20 | 2.21 |
| HSEE Math | 0.03 | 0.96 | -8.56 | 1.36 |
| g7-g9 FN Mean Math | 0.07 | 0.82 | -5.24 | 1.29 |
| g7 FN Mean Math | 0.09 | 0.84 | -5.60 | 1.44 |
| g8 FN Mean Math | 0.07 | 0.89 | -5.14 | 1.39 |
| g9 FN Mean Math | 0.04 | 0.90 | -6.49 | 1.26 |
| HSEE English | 0.03 | 0.96 | -7.77 | 1.42 |
| g7-g9 FN Mean English | 0.06 | 0.85 | -4.85 | 1.36 |
| g7 FN Mean English | 0.08 | 0.84 | -5.48 | 1.42 |
| g8 FN Mean English | 0.06 | 0.92 | -4.25 | 1.61 |
| g9 FN Mean English | 0.04 | 0.92 | -5.86 | 1.36 |
| HSEE Physics | 0.02 | 0.97 | -7.75 | 1.43 |
| g7-g9 FN Mean Physics | 0.04 | 0.58 | -3.30 | 0.97 |
| g8 FN Mean Physics | 0.07 | 0.90 | -4.23 | 1.57 |
| g9 FN Mean Physics | 0.04 | 0.92 | -5.94 | 1.47 |
| HSEE Chemistry | 0.02 | 0.98 | -8.25 | 1.46 |
| g9 FN Mean Chemistry | 0.03 | 0.91 | -5.69 | 1.38 |
| Observations | 6510 | | | |

(c) Stduent and Teacher Attributes

| | mean | sd | min | max |
|---|---|---|---|---|
| Admin-test Std. Total | 0.05 | 0.98 | -3.6 | 2.9 |
| PreT Cohort Ranking | 712.44 | 433.17 | 1 | 1567 |
| PreT Math | 0.09 | 0.95 | -8.8 | 2.6 |
| female | 0.49 | 0.50 | 0 | 1 |
| Age at Admission | 12.52 | 0.56 | 9.8 | 42 |
| EthnicHan | 0.79 | 0.41 | 0 | 1 |
| Admin_waived | 0.10 | 0.30 | 0 | 1 |
| One/Both Parent Job Academic | 0.34 | 0.47 | 0 | 1 |
| One/Both Parent Administrator | 0.72 | 0.45 | 0 | 1 |
| One/Both Parent Job Government | 0.38 | 0.49 | 0 | 1 |
| Top 2 Elementary School Grad | 0.27 | 0.44 | 0 | 1 |
| g7s1.head.Tchr.titleCont | 2.13 | 0.61 | 1 | 3 |
| g7s1.head.Tchr.female | 0.72 | 0.45 | 0 | 1 |
| g7s1.head.Tchr.startYrCount | 7.21 | 4.83 | 0 | 26 |
| g7s1.CHN.Tchr.titleCont | 2.10 | 0.74 | 1 | 3 |
| g7s1.CHN.Tchr.female | 0.80 | 0.40 | 0 | 1 |
| g7s1.CHN.Tchr.startYrCount | 7.48 | 6.21 | 0 | 28 |
| g7s1.MTH.Tchr.titleCont | 2.37 | 0.60 | 1 | 3 |
| g7s1.MTH.Tchr.female | 0.54 | 0.50 | 0 | 1 |
| g7s1.MTH.Tchr.startYrCount | 7.98 | 4.40 | 0 | 21 |
| Observations | 6510 | | | |

Table 4: Regressing Teacher Attributes on Assignment Letters and Class Numbers
(p-values on F-stats)

| Teacher Attribute | Assignment Letter | Class Number |
|---|---|---|
| Chinese Teacher Age | 0.580 | 0.063* |
| Chinese Teacher Female | 0.727 | 0.976 |
| Chinese Teacher Tenure (years) | 0.642 | 0.501 |
| Chinese Teacher has Tenure | 0.611 | 0.110 |
| Chinese Teacher Title Level | 0.521 | 0.054* |
| Chinese Teacher Total Experience (years) | 0.618 | 0.034** |
| Head Teacher Age | 0.686 | 0.297 |
| Head Teacher Female | 0.600 | 0.922 |
| Head Teacher Tenure (years) | 0.899 | 0.649 |
| Head Teacher has Tenure | 0.600 | 0.702 |
| Head Teacher Title Level | 0.895 | 0.041** |
| Head Teacher Total Experience (years) | 0.756 | 0.450 |
| Math Teacher Age | 0.794 | 0.963 |
| Math Teacher Female | 0.768 | 0.917 |
| Math Teacher Tenure (years) | 0.963 | 0.761 |
| Math Teacher has Tenure | 0.555 | 0.700 |
| Math Teacher Title Level | 0.699 | 0.982 |
| Math Teacher Total Experience (years) | 0.773 | 0.914 |

Note: Each row presents the results of a regression of teacher attributes on class letter and classroom number FE's. The first column contains the significance level for an F-test associated with the null hypothesis that the estimates on the class letter FE's are jointly zero, and the second column contains the significance level for an equivalent test for the classroom number FE's.

Table 5: Balancing Test, Admin Scores

| | Intent-to-treat Classes | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Admin-test Pseudo-Class Std. Mean | 1 Tercile pseudo-class peer mean | 2 Tercile pseudo-class peer mean | 3 Tercile pseudo-class peer mean |
| PreT Math | -0.000232 | -0.000240 | -0.000774 | -0.000176 |
| | (0.00200) | (0.00277) | (0.00216) | (0.00217) |
| female | 0.0000341 | -0.000734 | 0.0000971 | -0.000800 |
| | (0.000887) | (0.00153) | (0.00136) | (0.00109) |
| Age at Admission | 0.000904 | -0.000262 | 0.000972 | 0.00199 |
| | (0.00159) | (0.00244) | (0.00175) | (0.00182) |
| EthnicHan | -0.00400* | -0.00240 | -0.00510** | -0.00464** |
| | (0.00211) | (0.00328) | (0.00245) | (0.00224) |
| Eligible for Admission without Exam | 0.00593* | 0.0113** | 0.00781* | 0.000482 |
| | (0.00322) | (0.00553) | (0.00439) | (0.00300) |
| One/Both Parent Job Academic | 0.00362* | 0.00790*** | 0.00177 | 0.000225 |
| | (0.00184) | (0.00284) | (0.00226) | (0.00194) |
| One/Both Parent Administrator | -0.00148 | -0.00386 | -0.000692 | 0.000664 |
| | (0.00201) | (0.00310) | (0.00248) | (0.00229) |
| One/Both Parent Job Government | -0.000676 | -0.00240 | -0.000190 | 0.00149 |
| | (0.00214) | (0.00380) | (0.00205) | (0.00211) |
| Top 2 Elementary School Grad | 0.00329 | 0.00521 | 0.00168 | 0.00289 |
| | (0.00216) | (0.00346) | (0.00277) | (0.00201) |
| g7s1 head Teacher Title | -0.0291 | -0.0449 | -0.0179 | -0.0155 |
| | (0.0233) | (0.0391) | (0.0293) | (0.0206) |
| g7s1 head Teacher female | -0.0409** | 0.00337 | -0.0776*** | -0.0468** |
| | (0.0206) | (0.0303) | (0.0238) | (0.0220) |
| g7s1 head Teacher Tenure (Years) | 0.00479* | 0.00692 | 0.00413 | 0.00200 |
| | (0.00260) | (0.00428) | (0.00276) | (0.00230) |
| g7s1 CHN Teacher Title | 0.0107 | 0.000510 | 0.00823 | 0.00697 |
| | (0.0164) | (0.0251) | (0.0179) | (0.0176) |
| g7s1 CHN Teacher female | 0.0134 | -0.0240 | 0.0335 | 0.00244 |
| | (0.0253) | (0.0465) | (0.0276) | (0.0221) |
| g7s1 CHN Teacher Tenure (Years) | -0.00160 | -0.000261 | -0.00235 | -0.00190 |
| | (0.00198) | (0.00324) | (0.00221) | (0.00177) |
| g7s1 MTH Teacher Title | 0.0313 | -0.00632 | 0.0267 | 0.0519** |
| | (0.0207) | (0.0329) | (0.0262) | (0.0238) |
| g7s1 MTH Teacher female | -0.00383 | -0.0222 | -0.00433 | 0.0279 |
| | (0.0228) | (0.0336) | (0.0254) | (0.0234) |
| g7s1 MTH Teacher Tenure (Years) | -0.00165 | 0.00253 | -0.00411 | -0.00214 |
| | (0.00259) | (0.00436) | (0.00335) | (0.00310) |
| Observations | 6243 | 6243 | 6243 | 6243 |
| Adjusted $R^2$ | 0.340 | 0.320 | 0.470 | 0.331 |
| p-value for F-stat on student attributes | 0.258 | 0.147 | 0.401 | 0.440 |
| p-value for F-stat on teacher attributes | 0.331 | 0.672 | 0.071 | 0.014 |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes |
| student admin test score control | yes | yes | yes | yes |
| student PreT rank control | yes | yes | yes | yes |
| clustered s.e. | class | class | class | class |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Each column represents a different dependent variable covering in order the class mean admissions test, the mean of admissions test for the 1st tercile of students in the class, for the 2nd tercile and the 3rd tercile. The rows in the main section present the estimates on the student and teacher attributes. The bottom section presents significance levels for F-tests associated with the null hypothesis that the coefficients on all student or on all teacher attributes are jointly zero. All regressions contain controls for pre-test rank and admissions test score plus class letter, classroom number and cohort FE's. Standard errors are clustered at the class level.

## Table 6: Peer Effects by Tercile (intent-to-treat)

| | Intent-to-treat Peers | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | HSEE CHN&MTH | g7-g9 FN Mean CHN&MTH | g7 FN Mean CHN&MTH | g8 FN Mean CHN&MTH | g9 FN Mean CHN&MTH |
| 1 Tercile pseudo-class peer mean | 0.111 | 0.149* | 0.190** | 0.0956 | 0.141 |
| | (0.103) | (0.0883) | (0.0898) | (0.112) | (0.0972) |
| 2 Tercile pseudo-class peer mean | -0.138 | -0.0587 | -0.0219 | -0.129 | -0.0342 |
| | (0.212) | (0.168) | (0.155) | (0.199) | (0.188) |
| 3 Tercile pseudo-class peer mean | 0.0734 | -0.0754 | -0.142 | -0.0376 | -0.00372 |
| | (0.158) | (0.135) | (0.134) | (0.151) | (0.155) |
| Observations | 6229 | 6146 | 6222 | 6209 | 6192 |
| Adjusted $R^2$ | 0.449 | 0.597 | 0.599 | 0.542 | 0.501 |
| Clustered s.e. | class | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes | yes |
| Student Attributes | yes | yes | yes | yes | yes |
| Teacher Attributes | yes | yes | yes | yes | yes |
| H0: Joint Peer Effect F-test, p-value | 0.745 | 0.269 | 0.069 | 0.748 | 0.448 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Each column represents a different dependent variable covering in order the standardize scores on the combined Chinese and Math portions of the HSEE test, the average of 7th to 9th grade end of year exams, 7th grade, 8th grade and 9th grade end of year exams . The rows in the main section present the estimates on the tercile mean admissions test scores for the student's assigned class. The bottom section presents significance level for F-tests associated with the null hypothesis that the coefficients on the three tercile means are jointly zero. All regressions contain controls for pre-test rank and admissions test score plus class letter, classroom number and cohort FE's. Standard errors are clustered at the class level.

Table 7: Peer and Sibling Effects by Tercile (intent-to-treat)

| | Intent-to-treat Peers | | | | |
|---|---|---|---|---|---|
| | (1) HSEE CHN&MTH | (2) g7-g9 FN Mean CHN&MTH | (3) g7 FN Mean CHN&MTH | (4) g8 FN Mean CHN&MTH | (5) g9 FN Mean CHN&MTH |
| 1 Tercile pseudo-class peer mean | 0.173** | 0.202*** | 0.216*** | 0.166* | 0.199** |
| | (0.0867) | (0.0762) | (0.0817) | (0.0962) | (0.0852) |
| 2 Tercile pseudo-class peer mean | -0.173 | -0.0859 | 0.0307 | -0.192 | -0.0816 |
| | (0.181) | (0.148) | (0.150) | (0.172) | (0.165) |
| 3 Tercile pseudo-class peer mean | -0.123 | -0.245* | -0.297** | -0.223 | -0.183 |
| | (0.160) | (0.135) | (0.134) | (0.153) | (0.156) |
| 1 Tercile pseudo-sibling-class mean | 0.296*** | 0.249*** | 0.221** | 0.253** | 0.279*** |
| | (0.104) | (0.0872) | (0.0906) | (0.0981) | (0.100) |
| 2 Tercile pseudo-sibling-class mean | -0.235 | -0.161 | -0.0572 | -0.155 | -0.235 |
| | (0.154) | (0.134) | (0.134) | (0.166) | (0.145) |
| 3 Tercile pseudo-sibling-class mean | -0.208 | -0.176 | -0.184 | -0.207 | -0.167 |
| | (0.161) | (0.156) | (0.162) | (0.177) | (0.171) |
| Observations | 6115 | 6032 | 6108 | 6095 | 6078 |
| Adjusted $R^2$ | 0.452 | 0.600 | 0.599 | 0.545 | 0.504 |
| Clustered s.e. | class | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes | yes |
| Student Attributes | yes | yes | yes | yes | yes |
| Teacher Attributes | yes | yes | yes | yes | yes |
| Joint Peer Effect F-test p-value | 0.086 | 0.004 | 0.001 | 0.039 | 0.043 |
| Joint Sibling Effect F-test p-value | 0.007 | 0.006 | 0.031 | 0.009 | 0.009 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents estimates after adding controls for the peer quality in the sibling class. Each column represents a different dependent variable covering in order the standardize scores on the combined Chinese and Math portions of the HSEE test, the average of 7th to 9th grade end of year exams, 7th grade, 8th grade and 9th grade end of year exams. The rows in the main section present the estimates on the tercile mean admissions test scores for the student's assigned class and for the sibling class of the student's assigned class. The bottom section presents significance levels for F-tests associated with the null hypothesis that the coefficients on the three own class tercile means are jointly zero and that the coefficient on the three sibling class tercile means are jointly zero. All regressions contain controls for pre-test rank and admissions test score plus class letter, classroom number and cohort FE's. Standard errors are clustered at the class level.

Table 8: Peer and Sibling Effects by Tercile on Chinese and Math (intent-to-treat for all semesters)

(a) Effects on Chinese Score

| | Intent-to-treat Peers | | | | |
|---|---|---|---|---|---|
| | (1) HSEE Chinese | (2) g7-g9 FN Mean Chinese | (3) g7 FN Mean Chinese | (4) g8 FN Mean Chinese | (5) g9 FN Mean Chinese |
| 1 Tercile pseudo-class peer mean | -0.00623 | 0.0497 | 0.149 | -0.0359 | 0.00280 |
| | (0.101) | (0.0923) | (0.115) | (0.115) | (0.0966) |
| 2 Tercile pseudo-class peer mean | -0.365* | -0.228 | -0.0683 | -0.303 | -0.273 |
| | (0.202) | (0.195) | (0.229) | (0.236) | (0.185) |
| 3 Tercile pseudo-class peer mean | 0.0751 | -0.0507 | -0.187 | -0.0718 | 0.124 |
| | (0.188) | (0.173) | (0.199) | (0.214) | (0.179) |
| 1 Tercile pseudo-sibling-class mean | 0.124 | 0.0709 | 0.133 | -0.00405 | 0.104 |
| | (0.117) | (0.100) | (0.118) | (0.121) | (0.106) |
| 2 Tercile pseudo-sibling-class mean | -0.240 | -0.0897 | 0.0969 | -0.127 | -0.198 |
| | (0.172) | (0.172) | (0.206) | (0.199) | (0.167) |
| 3 Tercile pseudo-sibling-class mean | -0.187 | -0.195 | -0.310 | -0.231 | -0.0659 |
| | (0.195) | (0.169) | (0.215) | (0.204) | (0.180) |
| Observations | 6115 | 6037 | 6108 | 6097 | 6082 |
| Adjusted $R^2$ | 0.407 | 0.559 | 0.505 | 0.479 | 0.456 |
| Clustered s.e. | class | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes | yes |
| Student Attributes | yes | yes | yes | yes | yes |
| Teacher Attributes | yes | yes | yes | yes | yes |
| Joint Peer Effect F-test p-value | 0.144 | 0.395 | 0.368 | 0.102 | 0.377 |
| Joint Sibling Effect F-test p-value | 0.128 | 0.302 | 0.291 | 0.163 | 0.393 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents estimates using Chinese (panel a) and Math (panel b) scores separately. Each column represents a different dependent variable covering in order the standardize scores on the HSEE test, the average of 7th to 9th grade end of year exams, 7th grade, 8th grade and 9th grade end of year exams. The rows in the main section of each panel present the estimates on the tercile mean admissions test scores for the student's assigned class and for the sibling class of the student's assigned class. The bottom section presents significance levels for F-tests associated with the null hypothesis that the coefficients on the three own class tercile means are jointly zero and that the coefficient on the three sibling class tercile means are jointly zero. All regressions contain controls for pre-test rank and admissions test score plus class letter, classroom number and cohort FE's. Standard errors are clustered at the class level.

(b) Effects on Math Score

| | Intent-to-treat Peers | | | | |
|---|---|---|---|---|---|
| | (1) HSEE Math | (2) g7-g9 FN Mean Math | (3) g7 FN Mean Math | (4) g8 FN Mean Math | (5) g9 FN Mean Math |
| 1 Tercile pseudo-class peer mean | 0.280*** | 0.254** | 0.221* | 0.240* | 0.289*** |
| | (0.100) | (0.105) | (0.117) | (0.122) | (0.104) |
| 2 Tercile pseudo-class peer mean | 0.00730 | 0.0153 | 0.0980 | -0.111 | 0.0686 |
| | (0.189) | (0.175) | (0.183) | (0.189) | (0.188) |
| 3 Tercile pseudo-class peer mean | -0.259 | -0.326* | -0.335* | -0.261 | -0.344* |
| | (0.186) | (0.170) | (0.175) | (0.187) | (0.185) |
| 1 Tercile pseudo-sibling-class mean | 0.374*** | 0.311*** | 0.243** | 0.350*** | 0.341*** |
| | (0.111) | (0.108) | (0.113) | (0.119) | (0.113) |
| 2 Tercile pseudo-sibling-class mean | -0.189 | -0.167 | -0.137 | -0.161 | -0.203 |
| | (0.181) | (0.183) | (0.191) | (0.218) | (0.182) |
| 3 Tercile pseudo-sibling-class mean | -0.187 | -0.140 | -0.0712 | -0.141 | -0.208 |
| | (0.185) | (0.201) | (0.205) | (0.226) | (0.205) |
| Observations | 6115 | 6036 | 6108 | 6096 | 6081 |
| Adjusted $R^2$ | 0.374 | 0.552 | 0.549 | 0.487 | 0.446 |
| Clustered s.e. | class | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes | yes |
| Student Attributes | yes | yes | yes | yes | yes |
| Teacher Attributes | yes | yes | yes | yes | yes |
| Joint Peer Effect F-test p-value | 0.015 | 0.025 | 0.051 | 0.105 | 0.007 |
| Joint Sibling Effect F-test p-value | 0.002 | 0.013 | 0.157 | 0.010 | 0.005 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Peer and Sibling Effects by Tercile on English, Physics and Chemistry (intent-to-treat for all semesters)

(a) Effects on English Score

| | Intent-to-treat Peers | | | | |
|---|---|---|---|---|---|
| | (1) HSEE English | (2) g7-g9 FN Mean English | (3) g7 FN Mean English | (4) g8 FN Mean English | (5) g9 FN Mean English |
| 1 Tercile pseudo-class peer mean | 0.180* | 0.202** | 0.157 | 0.222** | 0.184* |
| | (0.0938) | (0.0861) | (0.0954) | (0.0956) | (0.0953) |
| 2 Tercile pseudo-class peer mean | -0.230 | -0.141 | -0.0208 | -0.134 | -0.214 |
| | (0.141) | (0.145) | (0.168) | (0.163) | (0.146) |
| 3 Tercile pseudo-class peer mean | -0.0859 | -0.0903 | -0.0176 | -0.108 | -0.128 |
| | (0.146) | (0.147) | (0.176) | (0.170) | (0.154) |
| 1 Tercile pseudo-sibling-class mean | 0.137 | 0.117 | 0.115 | 0.187* | 0.0649 |
| | (0.0947) | (0.0930) | (0.102) | (0.106) | (0.102) |
| 2 Tercile pseudo-sibling-class mean | -0.143 | 0.156 | 0.286* | 0.146 | 0.0378 |
| | (0.167) | (0.156) | (0.162) | (0.176) | (0.168) |
| 3 Tercile pseudo-sibling-class mean | -0.0245 | -0.195 | -0.198 | -0.258 | -0.139 |
| | (0.191) | (0.172) | (0.186) | (0.196) | (0.193) |
| Observations | 6115 | 6037 | 6108 | 6097 | 6082 |
| Adjusted $R^2$ | 0.377 | 0.461 | 0.425 | 0.444 | 0.404 |
| Clustered s.e. | class | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes | yes |
| Student Attributes | yes | yes | yes | yes | yes |
| Teacher Attributes | yes | yes | yes | yes | yes |
| Joint Peer Effect F-test p-value | 0.112 | 0.109 | 0.354 | 0.128 | 0.109 |
| Joint Sibling Effect F-test p-value | 0.460 | 0.095 | 0.051 | 0.033 | 0.707 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents estimates for the English (panel a), Physics (panel b) and Chemistry (panel c) scores. Each column represents a different dependent variable covering in order the standardize scores on the HSEE test, the average of 7th to 9th grade end of year exams, 7th grade, 8th grade and 9th grade end of year exams. The columns are adjusted based on the grades in which these subjects are covered The rows in the main section of each panel present the estimates on the tercile mean admissions test scores for the student's assigned class and for the sibling class of the student's assigned class. The bottom section presents significance levels for F-tests associated with the null hypothesis that the coefficients on the three own class tercile means are jointly zero and that the coefficient on the three sibling class tercile means are jointly zero. All regressions contain controls for pre-test rank and admissions test score plus class letter, classroom number and cohort FE's. Standard errors are clustered at the class level.

### (b) Effects on Physics Score

| | Intent-to-treat Peers | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | HSEE Physics | g7-g9 FN Mean Physics | g8 FN Mean Physics | g9 FN Mean Physics |
| 1 Tercile pseudo-class peer mean | 0.0644 | 0.137** | 0.236** | 0.149* |
| | (0.0738) | (0.0589) | (0.108) | (0.0843) |
| 2 Tercile pseudo-class peer mean | -0.111 | -0.109 | -0.1000 | -0.191 |
| | (0.133) | (0.107) | (0.179) | (0.171) |
| 3 Tercile pseudo-class peer mean | 0.0506 | -0.0670 | -0.183 | -0.000679 |
| | (0.138) | (0.107) | (0.194) | (0.160) |
| 1 Tercile pseudo-sibling-class mean | 0.158 | 0.155** | 0.236** | 0.245** |
| | (0.0970) | (0.0672) | (0.111) | (0.102) |
| 2 Tercile pseudo-sibling-class mean | -0.150 | 0.0328 | 0.188 | -0.106 |
| | (0.131) | (0.0944) | (0.170) | (0.138) |
| 3 Tercile pseudo-sibling-class mean | 0.0929 | -0.123 | -0.395* | 0.0321 |
| | (0.143) | (0.108) | (0.205) | (0.147) |
| Observations | 6115 | 6052 | 6096 | 6080 |
| Adjusted $R^2$ | 0.402 | 0.507 | 0.497 | 0.453 |
| Clustered s.e. | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes |
| Student Attributes | yes | yes | yes | yes |
| Teacher Attributes | yes | yes | yes | yes |
| Joint Peer Effect F-test p-value | 0.795 | 0.095 | 0.130 | 0.328 |
| Joint Sibling Effect F-test p-value | 0.436 | 0.029 | 0.013 | 0.108 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

### (c) Effects on Chemistry Score

| | Intent-to-treat Peers | |
|---|---|---|
| | (1) | (2) |
| | HSEE Chemistry | g9 FN Mean Chemistry |
| 1 Tercile pseudo-class peer mean | 0.0350 | 0.120 |
| | (0.0897) | (0.0935) |
| 2 Tercile pseudo-class peer mean | -0.194 | -0.274 |
| | (0.150) | (0.168) |
| 3 Tercile pseudo-class peer mean | 0.0486 | 0.0700 |
| | (0.178) | (0.173) |
| 1 Tercile pseudo-sibling-class mean | 0.104 | 0.0843 |
| | (0.111) | (0.0989) |
| 2 Tercile pseudo-sibling-class mean | -0.0829 | -0.0708 |
| | (0.182) | (0.171) |
| 3 Tercile pseudo-sibling-class mean | 0.111 | 0.113 |
| | (0.194) | (0.195) |
| Observations | 6115 | 6082 |
| Adjusted $R^2$ | 0.363 | 0.431 |
| Clustered s.e. | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes |
| Student Attributes | yes | yes |
| Teacher Attributes | yes | yes |
| Joint Peer Effect F-test p-value | 0.580 | 0.291 |
| Joint Sibling Effect F-test p-value | 0.789 | 0.797 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## Table 10: Gender Differences in Average Peer and Sibling Effects (intent-to-treat)

### (a) Effects on Chinese+Math Score (Boys)

| | Intent-to-treat Peers (Boys) | | | | |
|---|---|---|---|---|---|
| | (1)<br>HSEE CHN&MTH | (2)<br>g7-g9 FN Mean CHN&MTH | (3)<br>g7 FN Mean CHN&MTH | (4)<br>g8 FN Mean CHN&MTH | (5)<br>g9 FN Mean CHN&MTH |
| 1 Tercile pseudo-class peer mean | 0.197* | 0.221** | 0.204** | 0.177 | 0.248** |
| | (0.118) | (0.0916) | (0.0912) | (0.110) | (0.115) |
| 2 Tercile pseudo-class peer mean | -0.371* | -0.219 | -0.0444 | -0.324 | -0.280 |
| | (0.218) | (0.166) | (0.161) | (0.200) | (0.197) |
| 3 Tercile pseudo-class peer mean | -0.0439 | -0.269 | -0.321* | -0.287 | -0.146 |
| | (0.210) | (0.171) | (0.164) | (0.200) | (0.204) |
| 1 Tercile pseudo-sibling-class mean | 0.526*** | 0.362*** | 0.261** | 0.372*** | 0.453*** |
| | (0.134) | (0.107) | (0.0998) | (0.122) | (0.129) |
| 2 Tercile pseudo-sibling-class mean | -0.634*** | -0.369** | -0.138 | -0.317 | -0.583*** |
| | (0.199) | (0.160) | (0.151) | (0.194) | (0.189) |
| 3 Tercile pseudo-sibling-class mean | -0.0232 | -0.115 | -0.200 | -0.152 | -0.0167 |
| | (0.216) | (0.198) | (0.190) | (0.216) | (0.233) |
| Observations | 3126 | 3080 | 3126 | 3119 | 3105 |
| Adjusted $R^2$ | 0.461 | 0.610 | 0.610 | 0.551 | 0.513 |
| Clustered s.e. | class | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes | yes |
| Student Attributes | yes | yes | yes | yes | yes |
| Teacher Attributes | yes | yes | yes | yes | yes |
| Sub-sample | Boys | Boys | Boys | Boys | Boys |
| Joint Peer Effect F-test p-value | 0.113 | 0.006 | 0.009 | 0.008 | 0.053 |
| Joint Sibling Effect F-test p-value | 0.000 | 0.001 | 0.014 | 0.006 | 0.001 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents estimates separately for subsamples of boys and girls. Each column represents a different dependent variable covering in order the standardize scores on the combined Chinese and Math portions of the HSEE test, the average of 7th to 9th grade end of year exams, 7th grade, 8th grade and 9th grade end of year exams . The rows in the main section of each panel present the estimates on the tercile mean admissions test scores for the student's assigned class and for the sibling class of the student's assigned class. The bottom section presents significance levels for F-tests associated with the null hypothesis that the coefficients on the three own class tercile means are jointly zero and that the coefficient on the three sibling class tercile means are jointly zero. All regressions contain controls for pre-test rank and admissions test score plus class letter, classroom number and cohort FE's. Standard errors are clustered at the class level.

(b) Effects on Chinese+Math Score (Girls)

| | Intent-to-treat Peers (Girls) | | | | |
|---|---|---|---|---|---|
| | (1) HSEE CHN&MTH | (2) g7-g9 FN Mean CHN&MTH | (3) g7 FN Mean CHN&MTH | (4) g8 FN Mean CHN&MTH | (5) g9 FN Mean CHN&MTH |
| 1 Tercile pseudo-class peer mean | 0.158 | 0.205** | 0.245*** | 0.180 | 0.162* |
| | (0.0985) | (0.0865) | (0.0894) | (0.109) | (0.0923) |
| 2 Tercile pseudo-class peer mean | -0.00773 | 0.0154 | 0.0817 | -0.0950 | 0.0923 |
| | (0.187) | (0.161) | (0.168) | (0.182) | (0.170) |
| 3 Tercile pseudo-class peer mean | -0.157 | -0.221 | -0.288** | -0.146 | -0.197 |
| | (0.171) | (0.140) | (0.139) | (0.165) | (0.159) |
| 1 Tercile pseudo-sibling-class mean | 0.0546 | 0.142 | 0.193** | 0.132 | 0.104 |
| | (0.108) | (0.0871) | (0.0971) | (0.0991) | (0.101) |
| 2 Tercile pseudo-sibling-class mean | 0.156 | 0.0431 | -0.0107 | 0.0191 | 0.118 |
| | (0.171) | (0.148) | (0.148) | (0.178) | (0.169) |
| 3 Tercile pseudo-sibling-class mean | -0.348** | -0.225 | -0.122 | -0.260 | -0.302* |
| | (0.174) | (0.158) | (0.174) | (0.186) | (0.173) |
| Observations | 2989 | 2952 | 2982 | 2976 | 2973 |
| Adjusted $R^2$ | 0.427 | 0.580 | 0.582 | 0.527 | 0.482 |
| Clustered s.e. | class | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes | yes |
| Student Attributes | yes | yes | yes | yes | yes |
| Teacher Attributes | yes | yes | yes | yes | yes |
| Sub-sample | Girls | Girls | Girls | Girls | Girls |
| Joint Peer Effect F-test p-value | 0.362 | 0.031 | 0.002 | 0.266 | 0.174 |
| Joint Sibling Effect F-test p-value | 0.192 | 0.066 | 0.145 | 0.141 | 0.123 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Gender Differences in Average Peer and Sibling Effects (intent-to-treat)

(a) Effects on Math Score (Boys)

| | Intent-to-treat Peers (Boys) | | | | |
|---|---|---|---|---|---|
| | (1) HSEE Math | (2) g7-g9 FN Mean Math | (3) g7 FN Mean Math | (4) g8 FN Mean Math | (5) g9 FN Mean Math |
| Admin-test pseudo-class peer-girl 1Tercile mean std. score | 0.170 | 0.121 | 0.0800 | 0.0890 | 0.202* |
| | (0.111) | (0.101) | (0.0985) | (0.115) | (0.113) |
| Admin-test pseudo-class peer-girl 2Tercile mean std. score | 0.0937 | 0.121 | 0.0776 | 0.0902 | 0.169 |
| | (0.175) | (0.162) | (0.158) | (0.185) | (0.177) |
| Admin-test pseudo-class peer-girl 3Tercile mean std. score | -0.277 | -0.261* | -0.238 | -0.160 | -0.313* |
| | (0.169) | (0.156) | (0.154) | (0.184) | (0.167) |
| Admin-test pseudo-class peer-boy 1Tercile mean std. score | 0.242** | 0.194* | 0.130 | 0.236* | 0.210* |
| | (0.103) | (0.109) | (0.102) | (0.135) | (0.112) |
| Admin-test pseudo-class peer-boy 2Tercile mean std. score | -0.304 | -0.292 | -0.109 | -0.452** | -0.303 |
| | (0.201) | (0.196) | (0.190) | (0.225) | (0.207) |
| Admin-test pseudo-class peer-boy 3Tercile mean std. score | -0.0205 | -0.148 | -0.191 | -0.199 | -0.101 |
| | (0.182) | (0.154) | (0.152) | (0.178) | (0.179) |
| Admin-test pseudo-sibling-class girl 1Tercile mean std. score | 0.179 | 0.140 | 0.116 | 0.151 | 0.157 |
| | (0.111) | (0.103) | (0.0940) | (0.126) | (0.112) |
| Admin-test pseudo-sibling-class girl 2Tercile mean std. score | 0.0378 | -0.0915 | -0.131 | -0.0600 | -0.0541 |
| | (0.179) | (0.155) | (0.163) | (0.176) | (0.178) |
| Admin-test pseudo-sibling-class girl 3Tercile mean std. score | -0.189 | -0.159 | -0.240 | -0.0813 | -0.184 |
| | (0.189) | (0.177) | (0.170) | (0.204) | (0.186) |
| Admin-test pseudo-sibling-class boy 1Tercile mean std. score | 0.371*** | 0.299*** | 0.204** | 0.326*** | 0.352*** |
| | (0.104) | (0.0907) | (0.0909) | (0.100) | (0.105) |
| Admin-test pseudo-sibling-class boy 2Tercile mean std. score | -0.410** | -0.133 | -0.0135 | -0.113 | -0.261 |
| | (0.192) | (0.173) | (0.167) | (0.204) | (0.204) |
| Admin-test pseudo-sibling-class boy 3Tercile mean std. score | -0.116 | -0.0768 | 0.104 | -0.111 | -0.184 |
| | (0.201) | (0.177) | (0.170) | (0.194) | (0.204) |
| Observations | 3126 | 3083 | 3126 | 3120 | 3107 |
| Adjusted $R^2$ | 0.376 | 0.563 | 0.562 | 0.495 | 0.451 |
| Clustered s.e. | class | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes | yes |
| Student Attributes | yes | yes | yes | yes | yes |
| Teacher Attributes | yes | yes | yes | yes | yes |
| Sub-sample | Boys | Boys | Boys | Boys | Boys |
| Joint Peer Girl Effect F-test p-value | 0.163 | 0.234 | 0.367 | 0.659 | 0.081 |
| Joint Sibling Girl Effect F-test p-value | 0.353 | 0.405 | 0.201 | 0.661 | 0.362 |
| Joint Peer Boy Effect F-test p-value | 0.078 | 0.164 | 0.381 | 0.086 | 0.185 |
| Joint Sibling Boy Effect F-test p-value | 0.002 | 0.013 | 0.122 | 0.013 | 0.007 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: In this table, the tercile peer means in own and sibling classes are estimated separately for boys and girls. Each column represents a different dependent variable covering in order the standardize scores on the combined Chinese and Math portions of the HSEE test, the average of 7th to 9th grade end of year exams, 7th grade, 8th grade and 9th grade end of year exams . Panel a presents the estimates based on the subsample of boys, and panel b presents the estimates based on the girls subsample. The rows in the main section of each panel present the estimates on the tercile mean admissions test scores separately for the boys and girls in the student's assigned class and for their sibling class. The bottom section presents significance levels for F-tests associated with the null hypothesis that the coefficients on the three own class tercile means for boys or girls are jointly zero and that the coefficient on the three sibling class tercile means for boys or girls are jointly zero. All regressions contain controls for pre-test rank and admissions test score plus class letter, classroom number and cohort FE's. Standard errors are clustered at the class level.

(b) Effects on Math Score (Girls)

| | Intent-to-treat Peers (Girls) | | | | |
|---|---|---|---|---|---|
| | (1) HSEE Math | (2) g7-g9 FN Mean Math | (3) g7 FN Mean Math | (4) g8 FN Mean Math | (5) g9 FN Mean Math |
| Admin-test pseudo-class peer-girl 1Tercile mean std. score | -0.139 | -0.0580 | 0.0533 | -0.118 | -0.126 |
| | (0.0980) | (0.0886) | (0.0899) | (0.107) | (0.0904) |
| Admin-test pseudo-class peer-girl 2Tercile mean std. score | 0.138 | -0.0107 | -0.0608 | -0.0525 | 0.105 |
| | (0.166) | (0.164) | (0.175) | (0.188) | (0.167) |
| Admin-test pseudo-class peer-girl 3Tercile mean std. score | -0.296* | -0.304** | -0.314* | -0.174 | -0.354** |
| | (0.160) | (0.151) | (0.161) | (0.166) | (0.153) |
| Admin-test pseudo-class peer-boy 1Tercile mean std. score | 0.292*** | 0.262** | 0.0950 | 0.384*** | 0.291*** |
| | (0.0961) | (0.108) | (0.110) | (0.119) | (0.105) |
| Admin-test pseudo-class peer-boy 2Tercile mean std. score | 0.169 | 0.182 | 0.260 | 0.0428 | 0.243 |
| | (0.185) | (0.178) | (0.194) | (0.202) | (0.172) |
| Admin-test pseudo-class peer-boy 3Tercile mean std. score | 0.121 | 0.0975 | 0.0615 | 0.0956 | 0.0910 |
| | (0.145) | (0.149) | (0.167) | (0.161) | (0.149) |
| Admin-test pseudo-sibling-class girl 1Tercile mean std. score | 0.0511 | 0.0525 | 0.0416 | 0.0746 | 0.0606 |
| | (0.0953) | (0.0887) | (0.101) | (0.103) | (0.0843) |
| Admin-test pseudo-sibling-class girl 2Tercile mean std. score | 0.0192 | -0.144 | -0.253 | -0.116 | -0.0699 |
| | (0.174) | (0.162) | (0.183) | (0.186) | (0.161) |
| Admin-test pseudo-sibling-class girl 3Tercile mean std. score | -0.0838 | -0.0514 | -0.176 | 0.0380 | -0.0374 |
| | (0.188) | (0.191) | (0.195) | (0.212) | (0.186) |
| Admin-test pseudo-sibling-class boy 1Tercile mean std. score | 0.0375 | 0.101 | 0.0488 | 0.180* | 0.0694 |
| | (0.0777) | (0.0839) | (0.0938) | (0.0952) | (0.0800) |
| Admin-test pseudo-sibling-class boy 2Tercile mean std. score | 0.0850 | 0.0916 | 0.290* | -0.0562 | 0.0636 |
| | (0.154) | (0.148) | (0.161) | (0.165) | (0.163) |
| Admin-test pseudo-sibling-class boy 3Tercile mean std. score | -0.123 | -0.0550 | 0.0979 | -0.106 | -0.149 |
| | (0.176) | (0.164) | (0.177) | (0.186) | (0.164) |
| Observations | 2989 | 2953 | 2982 | 2976 | 2974 |
| Adjusted $R^2$ | 0.378 | 0.547 | 0.539 | 0.486 | 0.446 |
| Clustered s.e. | class | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | yes | yes | yes | yes | yes |
| Student Attributes | yes | yes | yes | yes | yes |
| Teacher Attributes | yes | yes | yes | yes | yes |
| Sub-sample | Girls | Girls | Girls | Girls | Girls |
| Joint Peer Girl Effect F-test p-value | 0.150 | 0.130 | 0.140 | 0.359 | 0.056 |
| Joint Sibling Girl Effect F-test p-value | 0.925 | 0.754 | 0.251 | 0.874 | 0.882 |
| Joint Peer Boy Effect F-test p-value | 0.004 | 0.020 | 0.296 | 0.008 | 0.004 |
| Joint Sibling Boy Effect F-test p-value | 0.776 | 0.477 | 0.162 | 0.302 | 0.604 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## Table 12: Peer and Sibling Effects by Head Teacher Subjects (intent-to-treat)

| | Row Number | Intent-to-treat Peers (1) HSEE Math | (2) g7-g9 FN Mean Math | (3) g7 FN Mean Math | (4) g8 FN Mean Math | (5) g9 FN Mean Math |
|---|---|---|---|---|---|---|
| 1 Tercile pseudo-class peer mean | 1 | 0.802*** | 0.571*** | 0.519** | 0.514** | 0.696*** |
| | | (0.206) | (0.199) | (0.220) | (0.222) | (0.206) |
| 2 Tercile pseudo-class peer mean | 2 | -0.563* | -0.491 | -0.374 | -0.751** | -0.348 |
| | | (0.330) | (0.318) | (0.338) | (0.339) | (0.330) |
| 3 Tercile pseudo-class peer mean | 3 | -0.335 | -0.325 | -0.207 | -0.531 | -0.203 |
| | | (0.333) | (0.328) | (0.337) | (0.364) | (0.339) |
| 1 Tercile pseudo-sibling-class mean | 4 | 0.155 | 0.0615 | -0.205 | 0.139 | 0.210 |
| | | (0.193) | (0.199) | (0.235) | (0.218) | (0.206) |
| 2 Tercile pseudo-sibling-class mean | 5 | -0.159 | -0.182 | 0.0902 | -0.320 | -0.279 |
| | | (0.256) | (0.263) | (0.300) | (0.299) | (0.269) |
| 3 Tercile pseudo-sibling-class mean | 6 | -0.0722 | 0.205 | -0.0710 | 0.650 | -0.00472 |
| | | (0.389) | (0.363) | (0.342) | (0.408) | (0.413) |
| assigned MTH Tchr not head Tchr=1 × 1 Tercile pseudo-class peer mean | 7 | -0.666*** | -0.362 | -0.312 | -0.322 | -0.500* |
| | | (0.249) | (0.243) | (0.256) | (0.268) | (0.259) |
| assigned MTH Tchr not head Tchr=1 × 2 Tercile pseudo-class peer mean | 8 | 0.551 | 0.713* | 0.833** | 0.849** | 0.526 |
| | | (0.386) | (0.383) | (0.395) | (0.425) | (0.399) |
| assigned MTH Tchr not head Tchr=1 × 3 Tercile pseudo-class peer mean | 9 | -0.248 | -0.458 | -0.551 | -0.212 | -0.567 |
| | | (0.349) | (0.340) | (0.355) | (0.377) | (0.355) |
| assigned MTH Tchr not head Tchr=1 × 1 Tercile pseudo-sibling-class mean | 10 | -0.0779 | 0.0278 | 0.247 | -0.000817 | -0.162 |
| | | (0.218) | (0.218) | (0.249) | (0.242) | (0.232) |
| assigned MTH Tchr not head Tchr=1 × 2 Tercile pseudo-sibling-class mean | 11 | -0.0156 | 0.0993 | -0.0250 | 0.168 | 0.137 |
| | | (0.277) | (0.290) | (0.324) | (0.319) | (0.296) |
| assigned MTH Tchr not head Tchr=1 × 3 Tercile pseudo-sibling-class mean | 12 | -0.559 | -0.855** | -0.529 | -1.318*** | -0.708* |
| | | (0.396) | (0.376) | (0.354) | (0.418) | (0.423) |
| assigned MTH Tchr sibling-class head Tchr=1 × 1 Tercile pseudo-class peer mean | 13 | -0.732** | -0.375 | -0.317 | -0.284 | -0.518* |
| | | (0.289) | (0.263) | (0.258) | (0.305) | (0.294) |
| assigned MTH Tchr sibling-class head Tchr=1 × 2 Tercile pseudo-class peer mean | 14 | 0.560 | 0.477 | 0.529 | 0.476 | 0.370 |
| | | (0.429) | (0.393) | (0.408) | (0.424) | (0.411) |
| assigned MTH Tchr sibling-class head Tchr=1 × 3 Tercile pseudo-class peer mean | 15 | 1.032** | 0.569 | 0.121 | 1.023** | 0.482 |
| | | (0.492) | (0.442) | (0.420) | (0.510) | (0.477) |
| assigned MTH Tchr sibling-class head Tchr=1 × 1 Tercile pseudo-sibling-class mean | 16 | 0.578** | 0.389 | 0.350 | 0.351 | 0.484 |
| | | (0.278) | (0.289) | (0.327) | (0.287) | (0.304) |
| assigned MTH Tchr sibling-class head Tchr=1 × 2 Tercile pseudo-sibling-class mean | 17 | 0.129 | 0.0341 | 0.0177 | 0.126 | -0.00131 |
| | | (0.452) | (0.444) | (0.473) | (0.474) | (0.469) |
| assigned MTH Tchr sibling-class head Tchr=1 × 3 Tercile pseudo-sibling-class mean | 18 | -0.717 | -0.365 | 0.456 | -1.125** | -0.395 |
| | | (0.550) | (0.501) | (0.466) | (0.547) | (0.569) |
| Observations | | 6114 | 6035 | 6107 | 6095 | 6080 |
| Adjusted $R^2$ | | 0.370 | 0.548 | 0.541 | 0.489 | 0.443 |
| Clustered s.e. | | class | class | class | class | class |
| Cohort&Class#&Class Letter Fixed Effects | | yes | yes | yes | yes | yes |
| Student Attributes | | yes | yes | yes | yes | yes |
| Teacher Attributes | | yes | yes | yes | yes | yes |
| H0: Hd-Tchr Peer Effect Jointly 0 (row 1-3) | | 0.001 | 0.016 | 0.100 | 0.002 | 0.012 |
| H0: Hd-Tchr Sibling Effect Jointly 0 (row 4-6) | | 0.866 | 0.893 | 0.818 | 0.408 | 0.734 |
| H0: Non-Hd-Tchr Peer Effects Jointly Identical to Hd-Tchr (row 7-9) | | 0.038 | 0.128 | 0.138 | 0.142 | 0.071 |
| H0: Non-Hd-Tchr Sibling Effects Jointly Identical to Hd-Tchr (row 10-12) | | 0.418 | 0.148 | 0.291 | 0.019 | 0.346 |
| H0: Sib-Hd-Tchr Peer Effect Jointly Identical to Hd-Tchr (13-15) | | 0.009 | 0.168 | 0.443 | 0.057 | 0.251 |
| H0: Sib-Hd-Tchr Sibling Effect Jointly Identical to Hd-Tchr (16-18) | | 0.030 | 0.263 | 0.384 | 0.058 | 0.163 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table interacts own and sibling tercile means with dummy variables for whether the math teacher is not a head teacher and whether the math teacher is the head teacher of the sibling class (omitted category is math teacher is student's head teacher). Each column represents a different dependent variable covering in order the standardize scores on the combined Chinese and Math portions of the HSEE test, the average of 7th to 9th grade end of year exams, 7th grade, 8th grade and 9th grade end of year exams. The first six rows present the estimates on the own and sibling class tercile peer variables. The next six rows present the estimates on the interaction of those variables with math teacher not a head teacher, and the last six rows in the main section present the estimates on the interaction with math teacher head teacher for sibling class. The bottom section presents significance levels for F-tests where the F-tests in order test whether a set of estimates on three tercile variables are jointly zero so that the first F-test is associated with rows 1-3, the second F-test with rows 4-6, etc. All regressions contain controls for pre-test rank and admissions test score plus class letter, classroom number and cohort FE's. Standard errors are clustered at the class level.