Spatial Structure of Peer Networks and Academic Achievement*

Jimmy Chan♣, Shilan Feng♦, Ya Tan↑, Zhi Wang♥, Qinghua Zhang♠

February 16, 2022

Abstract

This paper investigates whether changing the seating arrangement in a classroom can facilitate positive spillovers from top-performing students to others, using a field experiment conducted in a Chinese high school. Within classrooms, teachers sort students into study groups, keeping gender composition and academic performance balanced across groups. Each group contains six students who are arranged in two rows of three seats. The treatment alters the spatial distribution of academic abilities around the non-top-performing students by assigning the two top students to seats in the spatial center of each group. The results suggest that in the treated groups, there are enhanced academic spillovers from the top students. The treatment especially benefits the two bottom-performing students in science subjects. In contrast, the treatment exerts negative effects on the test scores of the two middle-performing students in English and biology, and survey responses following the experiment support the disruption mechanism. The results suggest that the spatial layout of a peer network can have a significant impact on learning outcomes.

JEL Classification: I21; J13; R23

Keywords: Social interactions; Peer effects; Spatial structure; Test scores; Field experiment

---

* We thank Eleonora Patacchini, two anonymous referees, Xuejing Zuo, Huailu Li, and Kevin Lang for substantive comments and suggestions. We thank all the seminar participants at Peking University and Virtual Workshops on Topics in Regional and Urban Economics: China, Asia, and Beyond for their helpful comments and suggestions. Wang acknowledges the support from the National Natural Science Foundation of China (Grant # 72003036). Feng acknowledges the support from the National Social Science Foundation of China (Grant # 20CJL030). Tan and Feng acknowledge the support from the Key Laboratory of Mathematical Economics and Quantitative Finance, Ministry of Education (Peking University), China. RCT ID: AEARCTR-0007880.

♣ The Chinese University of Hong Kong. Email address: jimmyhingchan@cuhk.edu.hk.

♦ Peking University. Email address: carollan@pku.edu.cn.

↑ School of International Trade and Economics, University of International Business and Economics. Email address: tanya@pku.edu.cn.

♥ China Center for Economic Studies, Fudan University, and Shanghai Institute of International Finance and Economics. Email address: wangzhi@fudan.edu.cn.

♠ Guanghua School of Management, Peking University. Email address: zhangq@gsm.pku.edu.cn.

1 Introduction

In a well-defined peer group where peer interactions are forced or encouraged by institutions, low-performing students could benefit from the presence of high-performing students (Carrell, Fullerton, and West, 2009; Li et al., 2014). Any positive spillover from high-performing students may remain scant, however, because bottom-performing students (hereafter simply "bottom students") tend to interact mostly with other bottom students in their peer groups (Carrell, Sacerdote, and West, 2013). In light of the vast body of literature demonstrating the importance of physical proximity in enhancing effective social interactions and knowledge spillovers,[1] this paper asks whether changing the spatial layout of a peer group in an effort to reduce physical distance between top-performing (hereafter simply "top students") and non-top-performing students can enhance their interactions and improve educational outcomes for the latter.

In this paper, we take advantage of several features of Chinese classrooms to investigate whether (and how) changing the seating arrangement in a classroom can facilitate positive spillovers from top students to others. We conducted a field experiment with students in the tenth grade of a high school in Shanxi province in China. In Chinese high schools, students in the same grade are divided into classes, and students in the same class sit for most lessons in the same classroom. It is the practice of the high schools in the province to encourage students to study in groups. In the high school we observed, students in the same class are divided into study groups by the teacher at the beginning of the semester. Each group typically consists of six students who sit close to each other in two rows of three seats during regular classes and self-study sections over the course of the semester. The gender composition and academic

[1] For example, see Jaffe, Trajtenberg, and Henderson (1993), Rosenthal and Strange (2003, 2004), Arzaghi and Henderson (2008), and Bayer, Ross, and Topa (2008).

performance are kept balanced across groups.[2] As a result, students face similar peer environments at the group level. A student's actual peer environment, however, may vary depending on a group's seating arrangement, as it may be easier for a student to talk to someone who sits next to her than to someone who sits behind or in front of her.

Our experiment seeks to understand whether placing top students in the spatial center of each group affects the academic performances of the non-top students. About half of the groups were randomly selected as treated groups; the remaining groups were used as reference groups. In the *treated groups*, the two top students (as measured by previous test scores) were assigned to seats in the spatial center of each group block (i.e., in the middle of each row of three). As such, this treatment altered the spatial layout of a group by seating a top student next to each of the non-top students, including two bottom students and two middle-performing students (hereafter simply "middle students"). In the *reference groups*, seats were arranged based on students' preferences. Consistent with the findings of Carrell, Sacerdote, and West (2013), in the reference groups where students were allowed to choose their own seating arrangements, the bottom students were significantly less likely to sit next to a top student than they would be under a random assignment. This suggests that when social connections within a study group are endogenous, peer interactions between the top and bottom students may remain quite limited even though students are encouraged to learn as a group.

We show that our treatment alters the spatial distribution of academic abilities around the non-top students within a study group by significantly increasing the likelihood that they sit next to top students, relative to their counterparts in the reference groups. In addition, we show that the treatment status is uncorrelated with the spatial distribution of any other observed characteristics, including gender, age, and all the observed family characteristics. This ensures that the treatment effect on learning outcomes occurs only through its impact on the spatial

[2] For more details, please see section 2.

distribution of academic abilities.

We find that our treatment has significant impacts on learning outcomes for non-top students, but the treatment effect is heterogeneous. The treatment raises the total score of a bottom student by 0.165 standard deviations. This positive treatment effect is mainly the result of higher scores in science subjects. Their test scores in physics, chemistry, and biology increase by 0.257, 0.258, and 0.261 standard deviations, respectively, with all these results significant at the 5% level. In contrast, the treatment significantly reduces the total score of a middle student by 0.122 standard deviations and this negative treatment effect is concentrated in English and biology. For the top students, the treatment effects are small and insignificant.

Having the top students sit in the middle of each row appears to facilitate positive spillover to bottom students. Survey responses following the experiment show that the tendency to help other group members with study problems increases with academic ability and is highest among the top students. This is consistent with the key presumption of our experimental design: top students can generate strong positive spillovers within study groups. The responses show that the treatment significantly increases the likelihood that a top student offers academic help by 20.4 percentage points, suggesting enhanced academic interactions with top students in the treated groups.

For the middle students, however, it is possible that their learning environments may be disrupted to a greater extent in the treated groups, explaining the negative treatment effect. The survey results show that middle students in the treated groups chitchatted more often than their counterparts in the reference groups. Furthermore, we find evidence that the disruption can be mitigated when the two middle students sit on the same side. We also show that the survey responses do not lend support to either of two alternative explanations that have been discussed in the literature.

Our paper is related to five strands of the literature. First, educational interventions designed to enhance learning in schools have attracted great interest among researchers and policy makers (see Bishop (2006), Glewwe and Kremer (2006), and Glewwe and Muralidharan (2016) for detailed reviews). Our paper contributes to this literature by shedding light on the effectiveness of adjusting seating arrangements within classrooms, which is much less costly than most conventional education policies.

Second, a large strand of the literature has shown evidence that students' peer groups affect their educational outcomes.[3] In most research settings, students are assigned to or selected for a variety of peer groups (e.g., schools, classrooms, study groups, living environments, and social groups) and are, therefore, exposed to varying peer compositions (as measured by test scores, gender, race, friendship, and so forth) and receive varying educational inputs.[4] Our paper asks an additional question: in the same peer compositions, does the spatial distribution of students in the peer group affect the academic achievements of its members?

Third, previous research has found that the spatial configuration of economic activities plays an important role in shaping behavior and economic outcomes.[5] Our paper contributes to this strand of literature by showing that the spatial distribution of a peer group can have a significant effect on knowledge spillovers in a microscale learning environment.

Fourth, our results contrast with those of Lu and Anderson (2015), who, in a field experiment conducted in a Chinese middle school, find that there is no benefit to sitting next to

[3] See the review paper by Sacerdote (2011).

[4] See Hoxby (2000), Sacerdote (2001), Zax and Rees (2002), Hanushek et al. (2003), Whitmore (2005), Hoxby and Weingarth (2005), Vigdor and Nechyba (2007), Burke and Sass (2013), Carrell, Fullerton, and West (2009), Lavy and Schlosser (2011), Carman and Zhang (2012), Lu and Anderson (2015), Hahn et al. (2020), among others.

[5] See Rosenthal and Strange (2004) for a survey of empirical research on evidence of agglomeration economies, and Ross (2011) and Topa and Zenou (2015) for overviews of literature on neighborhoods and social networks and their effects on human behavior and economic outcomes.

a student with higher test scores. The difference may be caused in part by differences in the underlying learning environments. Study pressure in China is greater in high school than in middle school. In the high school we observe, as is common among the better schools in China, students are required to attend supervised study halls in their classrooms following regular classes. Students who sit next to each other in our study, therefore, may have more opportunities to interact than their counterparts in the study by Lu and Anderson (2015). Furthermore, insofar as students in our study are actively encouraged to learn as a group, top students may also be more highly motivated to help other group members.

Fifth, the effect of group incentives in classrooms is explored by Li et al. (2014), who find that providing group incentives can benefit spillovers from high-achieving students to low-achieving students. Our results provide an additional insight: the impact of group incentives can be enhanced by altering seating arrangements to facilitate peer interactions.

The rest of the paper is organized as follows. In section 2, we describe the research background. In section 3, we discuss the experimental design and the data. In section 4, we examine the seating arrangements in the treatment and reference groups. Section 5 presents the main regression specification and the results. Section 6 offers discussion and concludes.

2. Background

The high school where the experiment was conducted is located in the capital city of the province of Shanxi. Like other Chinese high schools, it includes three grades: 10 through 12. Admissions are based on standardized test scores and are highly competitive. Our experiment was conducted with tenth-grade students during the first semester of the school year. All students take the same subjects in that semester. Later, students are divided into a science stream and an art stream.

In 2017, when the experiment was conducted, the school admitted a total of 591 students to the tenth grade. These students were sorted into 11 classes according to their scores on the high school entrance exam and a prep exam conducted in the summer before the start of the semester. Each class contained an average of 54 students. Each class had one administrative teacher, and students in the same class took most of their classes in the same classroom.[6] The students with the highest test scores were assigned to a special class that prepares students for various academic tournaments, such as the Mathematical Olympiad, which are popular in China. Because this special class followed a unique curriculum and pursued a unique goal, we excluded it from our experiment. The remaining 536 students were divided into 10 classes and followed the same curriculum. Our experiment was conducted with students in nine of these 10 classes.[7] A typical school day for our subjects started at 7:20 a.m. Regular classes lasted from 7:45 a.m. to 5:30 p.m., with a half-hour break for physical exercise in the morning, a lunch break of two and a half hours in the afternoon, and an eight-minute break between each set of two regular classes. After dinner, students attended study halls in their home classrooms from 7:20 p.m. until 9:45 p.m.

In recent years, the Shanxi provincial government has adopted a policy of promoting group learning. This policy aims to improve students' skills in independent study and group coordination and to enhance academic spillovers from high-performing students to others. As a typical example of the implementation of this government policy, in our high school, students in grade 10 were divided into study groups. Most of the groups contained six students. To

[6] In Chinese high schools, an administrative teacher is a regular teacher with additional managerial responsibilities, including arranging seats, organizing class events, implementing school policies, communicating with parents, and so on.

[7] In the class that we do not include in our sample, students were not assigned to study groups because their administrative teacher had just arrived from another high school and was not familiar with the group-learning policy implemented in our observed high school. We excluded that class from our experiment.

assign students to groups, the usual practice is for administrative teachers to divide their students into six tiers of roughly equal numbers, ranging from those with high initial test scores to those with low initial test scores, and then to pick one student from each tier to form a group.[8] Thus, by design, the various groups are similar in overall academic abilities. Care is also taken to ensure that gender compositions are roughly the same across groups. While teachers did not explicitly tell students how the groups were formed, the same procedure was used in previous years and was not kept secret. The division of students into tiers was used only in the formation of groups. Students were never told to which tier they belonged. Nevertheless, some students may have been aware of their positions.

Study groups played a significant role in the classroom. There were group presentations and projects, and students were encouraged to study together and help each other. Each group of students sat together during all classes, including study hall, in a block consisting of two rows and three "columns." Figure 1 shows the spatial arrangement of a typical classroom. The groups were moved once every two weeks so that they all had opportunities to sit in multiple parts of the classroom. Given this seating rotation, peers from other groups who sat close to certain students tended to change every two weeks and therefore were less likely to generate a constant influence than were peers from the same study group. The seating arrangement within each group, however, was fixed for the entire semester.

3. Experimental design and data

3.1 Experimental design

The group study policy implemented in our high school aims in particular to enhance academic

[8] We did not directly observe the group-assignment process, but the data suggest that the procedure was followed closely. Sometimes, when a class had 55 or 56 students, one or two students would end up left out. The teacher added them to an already-formed group in a random fashion.

spillovers from high-performing students to others. The objective of our experiment was to understand whether (and how) changing the seating arrangement in a classroom can facilitate positive spillovers from top students to others. We explained our plans to the teachers in September 2017. The group assignment took place in early October after a baseline exam was conducted to assess the academic abilities of the students. Before the baseline exam, students sat in columns sorted by height. After the groups were formed, we randomly selected about half of the groups to become treated groups. The remaining groups were used as reference groups. After announcing the group assignments and before arranging the seats, the administrative teacher asked all the students to submit their seating preferences.[9] The teacher informed the students that their preferences would be considered (but not necessarily implemented) when arranging group seating. They were also informed that the teacher would keep their submitted preferences private.

The administrative teacher then arranged the seats in each group block mainly according to the students' preferences, as was commonly done in previous years. In the treated groups, however, the two students with the highest scores on the baseline exam were assigned to the middle seat of each row.[10] We can surmise that students were unlikely to notice this manipulation of the final seating arrangement.[11] As we shall see in section 4.2, the treatment significantly changed the spatial distributions of group members. The seating arrangements

[9] Typically, students informed the teacher of their seating preferences by identifying which of their peers they wanted to sit next to and which they would rather not sit next to.

[10] If two students ranked in a group earned similar baseline test scores, the teacher might also consult their scores for the high-school entrance exam and a prep exam conducted in the summer before the start of the semester.

[11] Although each student knew that the teacher would consider his/her preferences when arranging seat assignments, the students all understood that the teacher had other concerns (e.g., the other students' preferences) as well. According to the conversations between the researchers and the study students at the end of the semester, very few students questioned the teacher when final seating arrangements did not fulfill their wishes.

were fixed for the entire semester. We deliberately minimized our intervention regarding seating arrangements for two reasons. First, to identify the treatment effects, we needed to ensure that the students were unaware of the experiment; otherwise, this information alone might have changed their behavior. Second, investigating the seating arrangement patterns chosen by the students promised to be interesting in its own right as such an investigation might show us how students endogenously interact with each other in a microscale learning environment. We discuss the associated details in section 4.1.

In late October, we conducted a survey to collect information about students' individual and family characteristics and to ask their views about group learning.[12] Students sat for three more exams during the semester: the second exam was administered on November 20 (the midterm exam), the third one was administered on December 19, and the last one was administered on February 9, 2018 (the final exam). Figure 2 provides the timeline. Students did not sit with their groups during the exams; hence, there was no concern that students in the same group might have cheated by collaborating.

3.2 Data

The full sample comprised 481 students from nine classes who started grade 10 in the fall semester of 2017. We excluded students who belonged to groups with more than six students or for some reason were not assigned to any study group.[13] The study sample includes 444 students in 74 study groups. Among the 74 groups in our sample, 38 are treated groups. The data include individual-level demographic information and the test scores from all four exams

[12] In the survey, we purposely avoided asking any questions related to seating arrangements.

[13] In some rare cases, certain students required extra attention from teachers because of physical or mental conditions. These students were seated next to the teacher's desk so they could be better taken care of.

that took place in the fall semester. Table 1 reports the summary statistics for the students in the study sample. About half belong to a treated group. About 47.5% are females. The average age is 15.3 years. Nearly 70% are the only child in their families. The average annual income per family is 94,690 RMB. About 61% and 56% of their fathers and mothers, respectively, are college graduates.

Balanced group assignment

Although we did not directly observe the group assignment process, the data suggest that the procedure was followed closely. Figure 3 shows the academic and gender compositions of the 74 study groups in our sample. Panel A shows that for more than 40 of the groups, there is exactly one student from each of the six ranking tiers. Panel B shows that the majority of the groups include exactly three female students. Overall, the academic and gender compositions are fairly balanced across groups, which is largely consistent with the group assignment rule described in section 2.

Random treatment assignment

We next check whether the treatment status is correlated with any individual characteristics. The results reported in Table 2 show that the mean differences in various observed baseline characteristics between the treated and reference groups are small and insignificant. To obtain the results reported in Table 3, we regress the dummy that indicates membership in a treated group on all the observable baseline characteristics. For column 1, we use the total score for the baseline exam, and for column 2, we use the baseline scores of each subject separately. The results show no statistically significant difference between the treated groups and the reference groups. The joint F statistics are 0.732 and 0.791, with p-values of

0.646 and 0.658, respectively. For the bottom, middle, and top one-third of the students in our sample by baseline scores, we also test for balance in the baseline characteristics between the treated and reference groups. The results reported in Tables A1 and A2 in the appendix show that almost all the baseline characteristics are balanced for each of the subsamples.[14]

4. Seating arrangements

4.1 Endogenous seating arrangements in the reference groups

As described in section 3.1, the seating arrangements in the reference groups were based mainly on the students' preferences and therefore should not be random. To quantify the difference, we consider each pair of students who sit next to each other (hereafter referred to as a "neighboring pair"). For each group block in the reference group, we calculate the fraction of every possible type of neighboring pair: bottom-bottom, middle-middle, top-top, top-middle, top-bottom, and middle-bottom. Columns 1 and 2 in Table 4 show the means and standard deviations. For each pair type, we then test if the average fraction is statistically different from the theoretical probability under a random seating assignment (column 4).[15] In columns 5 and

[14] In Table A2, two coefficients are marginally significant. One is for the baseline score in English of the top one-third of students, and another is for the baseline score in biology of the bottom one-third of students. Our main results are robust to including these two variables as controls.

[15] The theoretical probability for each pair type under a random seating assignment is calculated as follows. Let us denote the six students in a group as h1, h2, m1, m2, b1, and b2. If they are randomly seated, then for any pair of neighboring students, there are 15 possible combinations (i.e., 6*5/2=15). Among these 15 combinations, the number of high-high combinations is only one, (h1, h2), so the high-high probability is 1/15 (about 6.67%). The same is true for bottom-bottom and middle-middle combinations. There are four possible high-middle combinations, (h1, m1), (h1, m2), (h2, m1), and (h2, m2), so the probability of high-middle is 4/15 (about 26.67%). Similarly, there are four high-bottom combinations, (h1, b1), (h1, b2), (h2, b1), and (h2, b2), so the probability of high-bottom is 4/15 (about 26.67%). Finally, for middle-bottom combinations, we have (m1, b1), (m1, b2), (m2, b1), and (m2, b2), so the probability of middle-bottom is 4/15 (about 26.67%).

6, we report the *t*-statistics and *p*-values. For top-bottom and bottom-bottom combinations, the mean differences are statistically significant at the 5% and 12% significance levels, with *p*-values of 0.011 and 0.118, respectively. When student preferences decide the seating arrangements, the bottom students are more likely to sit next to each other and the top students are less likely to sit next to a bottom student in the reference groups than they would be under a random assignment.[16] These findings have two important implications. First, although the size of a study group is small, it still matters who a given student sits next to. For instance, it may be easier for a student to talk to someone who sits next to her than to someone who sits behind or in front of her. Second, and more importantly, the results imply that when social connections are endogenously determined, peer interactions between top and bottom students in a well-defined peer group may be quite limited even though they are encouraged to learn as a group. This highlights the importance of adjusting seating arrangements to improve peer interactions.

4.2 Treatment effects on the spatial distributions of peer compositions

Our treatment alters the spatial distributions of study groups by seating a top student next to each of the non-top students. We next quantify how this treatment of the seating arrangement affects the spatial distributions of peer compositions as measured by the baseline test scores, gender, age, and the observed family characteristics. On the one hand, we expect to see that for bottom and middle students, the treatment will increase the likelihood that they sit next to top students. On the other hand, our identification requires that the spatial distributions of all the

[16] As an additional check, we ran 1,000 independent simulations, each of which randomly assigned seats to all the students in the reference groups within their group blocks. We find that in 93% of the simulations, the bottom-bottom share is lower than the actual share of that pair type (9.62%). And in 99% of the simulations, the top-bottom share is higher than the actual fraction of that pair type (17.3%).

other baseline characteristics (including gender, age, and family background) should be uncorrelated with the treatment status. To check this, we ran the following regression for the bottom and middle third of the students, separately:

$$z_{igc} = \alpha_1 Treat_g + x_i \alpha_2 + \delta_c + u_{igc} . \quad (1)$$

where z_{igc} represents the baseline characteristics of the group peer(s) who sat next to student i in group g in class c (hereafter referred to as the neighbor peer)[17]; $Treat_g$ is a dummy variable that equals one if group g is a treated group; x_i is a set of all of student i's baseline characteristics (the standardized baseline score, gender, age, single-child status, family income, and father's and mother's college education status); δ_c is classroom fixed effects; and u_{igc} is an error term. As the unit of randomization is a study group, standard errors are clustered at this level (Athey and Imbens, 2017) and reported in parentheses below the estimated coefficients. We also report the randomization inference *p*-values to address concerns related to the small sample size (Young, 2019) as well as the *p*-values adjusted for multiple hypothesis testing (Westfall and Young, 1993).

Our treatment significantly increases the likelihood that bottom and middle students sit next to top students (columns 1 and 2 in Table 5, panel A) and significantly raises the baseline scores of bottom and middle students' neighbor peers by 0.291 and 0.255 standard deviations, respectively (columns 4 and 5 in Table 5, panel A). All the other observed baseline characteristics of neighbor peers, however, such as gender, age, single-child status, family income, and parents' educational attainment, are uncorrelated with the treatment status (see panel B in Table 5). This ensures that the treatment effect on learning outcomes occurs only through its impact on the spatial distribution of academic abilities. For the top students, none

[17] If a given student sat in the middle column and thus had two neighbor peers, we used the average of the two peers' baseline attributes.

of the baseline characteristics of their neighbor peers is affected by the treatment (see columns 3 and 6 in panels A and B-3 in Table 5).[18]

5. Effects of spatial layout of academic abilities on learning outcomes

5.1 Main effects

Figure 4 separately plots the density distributions of the standardized baseline scores and the averages of the standardized scores on post-treatment exams for the top, middle, and bottom students.[19] The results reported in panels A to C in Figure 4 show that the baseline scores are balanced between the treated and reference groups, a finding that is consistent with the results shown in Tables 2 and 3. While the results reported in panel D suggest a positive treatment effect for the bottom students, those reported in panel E indicate that the treatment effect for the middle students is negative. For the main analysis, we ran separate regressions to investigate the treatment's effect on the academic achievements of the students in each third of the baseline score distribution. The regression specification is

$$y_{igc} = \beta_1 Treat_g + x_i \beta_2 + \kappa_c + \varepsilon_{igc} . \qquad (2)$$

The outcome y_{igc} represents the average of the standardized scores on all the post-treatment

[18] For a robustness check, we construct, for each student, a spatial average peer variable to characterize the spatial distribution of all his/her group peers in terms of each baseline attribute weighted by the physical distances between the students. As such, these spatial average peer variables capture not only the characteristics of the group peer members who sat next to each student, but also the characteristics of those who sat in front, behind, or farther away. We find that the spatial averages of baseline scores across group peers for the bottom and middle students are significantly higher in the treated groups than in the reference groups. The spatial layouts of all the other observed characteristics are uncorrelated with the treatment status.

[19] For each of the four exams conducted in the study semester, the standardized score represents the sum of the scores for all subjects, standardized at a mean of zero and a standard deviation of one.

exams for student i in group g in class c; x_i is a set of all of student i's baseline characteristics, including the standardized baseline score, gender, age, a single-child dummy, family income, and parents' educational attainment; κ_c are the classroom fixed effects, which control for differences in score distributions across classes; and ε_{igc} is an error term. As the unit of randomization is a study group, we cluster the standard errors at this level (Athey and Imbens, 2017). We also investigate the post-treatment scores of the various academic subjects as outcomes. To address the concern that our subgroup analysis based on initial test scores and subjects suffered from type-I errors, we compute the *p*-values adjusted for multiple hypothesis testing, following Westfall and Young (1993). To address the concern that the sample might be too small, we also compute the randomization inference *p*-values, following Young (2019). These adjusted *p*-values are reported in the rows below the clustered standard errors.

In equation (2), the coefficient of interest, β_1, captures the extent to which the standard deviations of the overall test scores are raised by the treatment relative to what occurs in the reference group. As shown in section 4.2 on the treatment altering the spatial distributions of peer academic abilities for the bottom and middle students, the estimate of β_1 tells us whether this change has significant impacts on learning outcomes. An unbiased estimation of β_1 requires that, conditional on the controls, $Treat_{cg}$ be uncorrelated with error term $\varepsilon_{i,cg}$. This identification assumption holds in the balanced-group-assignment and random-treatment-assignment settings, which are verified in section 3.

In Table 6, we report the results obtained with regression specification (2). The results reported in panels A to C indicate the estimates of the treatment effects for the bottom, middle, and top students separately. We find that the treatment has significant effects on learning outcomes for the non-top students, but the treatment effects are heterogeneous. The treatment raises the total test scores of the bottom students by 0.165 standard deviations (column 1 in

panel A) but reduces the total test scores of the middle students by 0.122 standard deviations (column 1 in panel B). The coefficients are statistically significant at the 10% and 5% levels, respectively. The treatment effect on the total scores of the top students is small and insignificant (column 1 in panel C).

In columns 2 to 7 in each panel in Table 6, we report the estimated treatment effects on scores by subject. For the bottom students, the positive treatment effect on the total score is the result mainly of score increases in science subjects, such as physics, chemistry, and biology. The score increments are 0.257, 0.258, and 0.261 standard deviations, respectively, and all are statistically significant at the 5% level. For the middle students, we find that sitting in a treatment group has negative and significant effects on their test scores in English and biology.[20]

5.2 Mechanism discussion

5.2.1 Enhanced positive spillovers from top students in treated groups

Are positive spillovers from top students enhanced by having them sit in the spatial center of each study group? To provide evidence, we investigate the students' views about within-group peer interactions using survey responses following the experiment. The survey item asked students to rate, on a scale ranging from one (strongly disagree) to five (strongly agree), the extent to which they agreed with the following statement: "I spent some time helping other group members with their study problems and considered this an opportunity to enhance learning for myself." Overall, about 43.84% of the top students selected "strongly agree,"

[20] For a robustness check, we run the main regressions by applying the Lasso method developed by Belloni, Chernozhukov, and Hansen (2014a, 2014b). The main results remain unchanged. In addition, we run the regressions using each of the three post-treatment exam results. The main results remain robust here as well.

compared with 38.37% of the middle students and 30.59% of the bottom students. After controlling for all the baseline characteristics, a top student is 18.1% more likely to help other group members with their study problems than a middle or bottom student (see column 1 in Table 7). This lends support to the key presumption of our experiment: within study groups, the top students can generate strong positive spillovers. The results reported in column 2 in Table 7 demonstrate that after controlling for all the baseline characteristics, top students in the treated groups are 20.4% more likely to offer academic help than their counterparts in the reference groups. The difference is statistically significant at the 10% level. These results imply that our seating adjustment can enhance academic interactions with top students.[21]

The positive spillovers from top students appear to be concentrated in science subjects, including physics, chemistry, and biology, as shown in panel A in Table 6. The top students may help bottom students better understand the core science concepts in the textbooks, extend these concepts into new areas, and apply them to solving problems.

5.2.2 More disruptions for middle students in treated groups

What explains the negative treatment effect for middle students? It is possible that middle students in the treated groups experienced greater disruption of the learning environment than their counterparts in the reference groups. The survey asked students how often in a given week they chatted with other group members about personal issues not related to study.[22] The results reported in column 1 in Table 8 suggest that the middle students in the treated groups

[21] The survey additionally asked students to rate the extent to which they agreed with the following statements: (1) “I often discussed academic problems with other group members,” (2) “I made a great effort to help my study group outperform other groups,” and (3) “I felt that group learning cost me too much time and energy.” For each statement, we find that the treatment effect on the likelihood of selecting “strongly agree” is statistically insignificant for each third of the students.

[22] The average frequency of chitchats per week is about four, as shown in Table 1.

chitchatted more often than their counterparts in the reference groups. For example, it is possible that in the treated groups, bottom students talked about study problems with top students who sat next to them and chatted about personal issues with middle students who sat in front of or behind them. Such disruption might be reduced by seating the two middle students on the same side so they can concentrate on their own studies (e.g., see Group 1 in Figure 1). As supporting evidence for this conjecture, we find that the positive treatment effect on chitchat for the middle students diminishes and becomes insignificant when they sat on the same side in the treated groups (column 4 in Table 8). Furthermore, for these middle students, the negative treatment effect on test scores becomes much smaller (-0.009 standard deviations) and statistically insignificant (column 1 in Table 9). As there are only seven treated groups where the middle students sat on the same side, we regard the results reported in columns 4 to 6 in Tables 8 and 9 as suggestive evidence given the small sample size.[23]

The heterogeneous treatment effects across academic subjects may further support the disruption mechanism. In particular, the negative treatment effect for middle students is concentrated largely in English and biology (see columns 4 and 7 in panel B in Table 6). In Chinese high schools, English exams are designed to evaluate students' proficiency in English grammar; biology exams evaluate students' understanding of extensive course notes. Because the ability to memorize rules and details is important to performing well on these exams, external distractions may have a larger disruptive influence on the learning outcomes. The negative treatment effect on middle students in English and biology disappears when they sit on the same side (see columns 4 and 7 in Table 9).

We also explore two alternative explanations discussed in the literature (Sacerdote, 2011).

[23] In Appendix Table A3, we report the treatment effects on academic outcomes for the bottom and top students by restricting the treated groups to this subsample. We find that the treatment significantly raises the total scores of the bottom and top students in this subsample by 0.345 and 0.131 standard deviations, respectively.

First, according to the invidious competition model, middle students may have been harmed by the presence of higher achieving peers in the treated groups. Our survey asked students to rate the extent to which they agreed with the following statement: "I felt that peer interactions between group members were mitigated by competition." The results reported in column 2 in Table 8 indicate that there is no evidence that middle students in the treated groups are more likely to select "strongly agree" than their counterparts in the reference groups. Second, the boutique model suggests that students perform better when surrounded by similar peers because similarity allows students to help each other more easily. Our treatment spatially separated two middle students and therefore may have precluded them from helping each other with study problems. There is no evidence, however, that the middle students in the treated groups were less likely to offer academic help (column 3 in Table 8).[24]

5.3 Further discussion

Our main results imply that there is a positive peer effect for bottom students when they sit next to top students. To compare our peer effect estimate with those from the literature, we estimate the extent to which bottom students' post-treatment test scores are affected by the baseline test scores of the peers who sat next to them. We use the treatment as an instrument because the ordinary least squares estimation is likely biased by the endogenous seating arrangement. We report the instrumental variable (IV) estimation results in Table A4. The peer effects are positive and significant (or marginally significant) for all the science subjects (physics, chemistry, and biology). In particular, a one-standard-deviation increase in the baseline score of the peer who sat next to a bottom student is associated with an increase in the

[24] In columns 5 and 6 in Table 8, we report the treatment effects on these two survey outcomes by investigating the treated groups in which the middle students sat on the same side and all the reference groups. The estimates are both small and insignificant.

test scores on the science subjects of about 0.88 standard deviations. It is larger than most estimates from the literature, which range from -0.12 to 1.9 (Sacerdote, 2011). The IV estimates are insignificantly positive for Chinese and math and insignificantly negative for English. These results echo the literature, where studies find that peer effects tend to be quite heterogeneous across academic subjects (e.g., Carman and Zhang, 2012).

We also explore whether the treatment effects on the bottom students vary with student characteristics. We report the results in Table A5 in the appendix. Although the bottom students on average benefit from sitting next to top students, the positive treatment effect attenuates as the gap in baseline scores between top and bottom students widens. We also find that the positive treatment effect is concentrated mostly in girls, which is consistent with findings reported in the literature indicating that girls are more likely to be nurtured by social influences than boys (e.g., Gong, Lu, and Song, 2018; Fletcher, Ross, and Zhang, 2020; Hahn et al., 2020). Finally, we find that the treatment effects are larger for students from wealthier socioeconomic backgrounds. We regard the results reported in Table A5 as suggestive evidence because of the limited sample size along the dimensions of the student characteristics.

6. Conclusion

This paper investigated whether changing seating arrangements in a classroom can facilitate positive spillovers from top students to others. We conducted a field experiment in a Chinese high school. Within classrooms, teachers sort students into study groups, keeping gender composition and academic performance balanced across the groups. By exploiting the data advantage whereby students face similar peer environments at the group level, we investigated the impacts of seating arrangements on test scores. Our treatment altered the spatial distribution of academic abilities by assigning the two top students to sit in the spatial center of

each group. We found evidence that there were enhanced interactions and positive spillovers between the top and bottom students in the treated groups. The spatial distribution of peer academic abilities has significant impacts on learning outcomes. The treatment benefits the bottom students, especially in science subjects. Their test scores in physics, chemistry, and biology increased by 0.257, 0.258, and 0.261 standard deviations, respectively, with all results significant at the 5% level. However, the performance of the middle students seemed to suffer from the treatment; the disruption mechanism may be the explanation for this finding.

The idea of studying in close proximity to academically higher performing peers has historical roots in China's education philosophy. In recent years, policy makers have begun promoting a group learning policy to improve students' abilities in independent thinking and group coordination and, more importantly, to enhance academic spillovers from high-scoring students to others. Our findings provide insights for educators and policy makers on the potential impacts of seating arrangements in a microscale learning environment on the learning outcomes of students with varying academic abilities. An interesting insight concerns the effect of group incentives on peer spillovers (Li et al., 2014) and endogenous social interactions within a peer group (Carrell, Sacerdote, and West, 2013). Because all students are encouraged to learn as a group, the top students should be motivated to help other group members. However, we found that when students are allowed to choose their own seating arrangements, high-performing students are less likely to sit next to low-performing students than they would be under random assignments. This implies that if teachers ignore seating arrangements, positive interactions between the top and bottom students may remain rare despite the presence of group learning incentives.

While this paper provides insights into the effects of the spatial structure of a peer network on students' academic outcomes, future research could further investigate several important questions. For example, it would be worthwhile to study whether seating

arrangements affect other outcomes, including mental health and social skills. It would also be interesting to examine the long-term effects on these outcomes as it may take years for the potential effects to emerge. In addition, future research could study whether the effects on academic outcomes decline or persist in later levels of education. Finally, it would be valuable to examine how spatial peer effects interact with teachers' characteristics and behaviors and other school-related input factors.

References

Arzaghi, Mohammad, and John Vernon Henderson. "Networking off Madison Avenue." *Review of Economic Studies* 75.4 (2008), 1011-1038.

Athey, Susan, and Guido Wilhelmus Imbens. "The Econometrics of Randomized Experiments." In Abhijit Vinayak Banerjee and Esther Duflo (eds.), *Handbook of Field Experiments*, Vol. 1. North-Holland, 2017. 73-140.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy* 116.6 (2008), 1150-1196.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28.2 (2014a), 29-50.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *Review of Economic Studies* 81.2 (2014b), 608-650.

Bishop, John. "Drinking from the Fountain of Knowledge: Students Incentive to Study and Learn.—Externalities, Information Problems and Peer Pressure." In E. Hanushek, F. Welch (eds.), *Handbook of the Economics of Education*, Vol. 2. Elsevier, 2006, 909-944.

Burke, Mary A., and Tim R. Sass. "Classroom Peer Effects and Student Achievement." *Journal of Labor Economics* 31.1 (2013), 51-82.

Carman, Katherine Grace, and Lei Zhang. "Classroom Peer Effects and Academic Achievement: Evidence from a Chinese Middle School." *China Economic Review* 23.2 (2012), 223-237.

Carrell, Scott E., Richard L. Fullerton, and James E. West. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." *Journal of Labor Economics* 27.3 (2009), 439-464.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. "From Natural Variation to Optimal

Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81.3 (2013), 855-882.

Fletcher, Jason M., Stephen L. Ross, and Yuxiu Zhang. "The Consequences of Friendships: Evidence on the Effect of Social Relationships in School on Academic Achievement." *Journal of Urban Economics* 116 (2020), 103241.

Glewwe, Paul, and Michael Kremer. "Schools, Teachers, and Education Outcomes in Developing Countries." In E. Hanushek, F. Welch (eds.), *Handbook of the Economics of Education*, Vol. 2, Elsevier, 2006, 945-1017.

Glewwe, Paul and Karthik Muralidharan. "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." In Eric A. Hanushek, Stephen Machin, Ludger Woessmann (eds.), *Handbook of the Economics of Education*, Vol. 5, Elsevier, 2016, 653-743.

Gong, Jie, Yi Lu, and Hong Song. "The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes." *Journal of Labor Economics* 36.3 (2018): 743-778.

Hahn, Youjin, Asadul Islam, Eleonora Patacchini, and Yves Zenou. "Friendship and Female Education: Evidence from a Field Experiment in Bangladeshi Primary Schools." *Economic Journal* 130.627 (2020), 740-764.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics* 18.5 (2003), 527-544.

Hoxby, Caroline M. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper 7867 (2000).

Hoxby, Caroline M., and Gretchen Weingarth. "Taking Race out of the Equation: School Reassignment and the Structure of Peer Effects." Unpublished manuscript, Department of Economics, Harvard University (2005).

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108.3 (1993), 577-598.

Lavy, Victor, and Analia Schlosser. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics* 3.2 (2011), 1-33.

Li, Tao, Li Han, Linxiu Zhang, and Scott Rozelle. "Encouraging Classroom Peer Interactions: Evidence from Chinese Migrant Schools." *Journal of Public Economics* 111 (2014), 29-45.

Lu, Fangwen, and Michael L. Anderson. "Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups." *Journal of Labor Economics* 33.1 (2015), 91-122.

Rosenthal, Stuart S., and William C. Strange. "Geography, Industrial Organization, and Agglomeration." *Review of Economics and Statistics* 85.2 (2003), 377-393.

Rosenthal, Stuart S., and William C. Strange. “Evidence on the Nature and Sources of Agglomeration Economies.” In John Vernon Henderson and Jacques-François Thisse (eds.), *Handbook of Regional and Urban Economics*, Vol. 4. Elsevier, 2004. 2119-2171.

Ross, Stephen L. “Social Interactions within Cities: Neighborhood Environments and Peer Relationships.” In Nancy Brooks, Kieran Donaghy, and Gerrit-Jan Knapp (eds.), *Handbook of Urban Economics and Planning*. Oxford University Press, 2011, Oxford, England.

Sacerdote, Bruce. “Peer Effects with Random Assignment: Results for Dartmouth Roommates.” *Quarterly Journal of Economics* 116.2 (2001), 681-704.

Sacerdote, Bruce. “Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?” In Eric A. Hanushek, Stephen Machin, and Ludger Woessmann (eds.) *Handbook of the Economics of Education*. Vol. 3. Elsevier, 2011. 249-277.

Topa, Giorgio, and Yves Zenou. “Neighborhood and Network Effects.” In G. Duranton, John Vernon Henderson, and William C. Strange (eds.), *Handbook of Regional and Urban Economics*. Vol. 5. Elsevier, 2015. 561-624.

Vigdor, Jacob, and Thomas Nechyba. “Peer Effects in North Carolina Public Schools.” In Ludger Woessmann and Paul E. Peterson (eds.), *Schools and the Equal Opportunity Problem*. 2007. 73-101.

Westfall, Peter H., and S. Stanley Young. “Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.” Vol. 279. John Wiley & Sons, 1993.

Whitmore, Diane. “Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment.” *American Economic Review* 95.2 (2005), 199-203.

Young, Alwyn. “Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results.” *Quarterly Journal of Economics* 134.2 (2019), 557-598.

Zax, Jeffrey S., and Daniel I. Rees. “IQ, Academic Performance, Environment, and Earnings.” *Review of Economics and Statistics* 84.4 (2002), 600-616.

Table 1: Summary statistics

	Mean	S.D.	Obs.	Min.	Max.
A. Baseline characteristics					
Dummy: treatment group	0.514	0.500	444	0.000	1.000
Standardized baseline score	0.000	1.000	422	-8.289	1.423
Dummy: female student	0.475	0.500	444	0.000	1.000
Age in months	15.287	0.392	440	13.500	17.000
Dummy: single child	0.698	0.460	440	0.000	1.000
Family annual income, 10,000 yuan	9.469	5.324	440	3.000	20.000
Dummy: father having a college degree	0.607	0.489	440	0.000	1.000
Dummy: mother having a college degree	0.561	0.497	440	0.000	1.000
B. Post-treatment characteristics					
Average standardized score on post-treatment exams	-0.045	1.027	441	-4.816	1.411
Dummy: help others with study problems	0.355	0.479	440	0.000	1.000
Frequency of chitchats during a week	4.302	3.452	440	0.000	10.000
Dummy: feel pressure from peer competition	0.020	0.142	440	0.000	1.000

Note: In panel B, the average standardized score on post-treatment exams represents the average of the standardized scores on the post-treatment exams, including the middle exam, the third exam, and the final exam.

Table 2: Baseline individual-level characteristics by treatment status

	Treatment group			Reference group			Treatment-Reference
	Mean	S.D.	Obs.	Mean	S.D.	Obs.	t stat
Standardized baseline total score	0.012	0.947	218	-0.012	1.055	204	0.700
Standardized baseline Chinese score	0.010	0.909	218	-0.011	1.091	204	0.0930
Standardized baseline math score	0.016	0.987	218	-0.018	1.016	204	-0.0421
Standardized baseline English score	-0.057	1.095	218	0.061	0.886	204	-0.683
Standardized baseline physics score	0.033	0.969	218	-0.035	1.033	204	0.924
Standardized baseline chemistry score	0.057	0.855	218	-0.061	1.133	204	1.406
Standardized baseline biology score	-0.022	0.971	218	0.024	1.032	204	0.583
Dummy: female student	0.487	0.501	228	0.463	0.500	216	0.765
Age in months	15.278	0.372	228	15.298	0.413	212	-0.487
Dummy: single child	0.667	0.472	228	0.731	0.444	212	-1.102
Family annual income, 10,000 yuan	9.579	5.508	228	9.351	5.129	212	0.520
Dummy: father having a college degree	0.588	0.493	228	0.627	0.485	212	-0.0991
Dummy: mother having a college degree	0.535	0.500	228	0.590	0.493	212	-0.837

Note: The *t*-statistic is for the null hypothesis that the difference between the means of the treatment and reference groups is zero. All *t*-statistics are adjusted for clustering at the study group level. In each test, we control for the classroom fixed effects.

Table 3: Random treatment assignment

Dependent variable: Dummy variable indicating the treatment group		
	(1)	(2)
Standardized baseline total score	0.015	
	(0.016)	
Standardized baseline Chinese score		0.000
		(0.020)
Standardized baseline math score		-0.021
		(0.032)
Standardized baseline English score		-0.021
		(0.023)
Standardized baseline physics score		0.014
		(0.028)
Standardized baseline chemistry score		0.038
		(0.036)
Standardized baseline biology score		0.001
		(0.028)
Dummy: female student	-0.007	-0.000
	(0.025)	(0.025)
Age in months	-0.018	-0.012
	(0.050)	(0.051)
Dummy: single child	-0.045	-0.043
	(0.032)	(0.033)
Family annual income, 10,000 yuan	0.003	0.004
	(0.005)	(0.005)
Dummy: father having a college degree	0.016	0.018
	(0.055)	(0.055)
Dummy: mother having a college degree	-0.049	-0.048
	(0.053)	(0.053)
Classroom fixed effects	Y	Y
Observations	418	418
R-squared	0.272	0.276
F-statistic	0.732	0.791
p-value	0.646	0.658

Note: We use the linear probability regression model. The standard errors are clustered at the study group level. The F-statistics are for testing the joint significance of all the observed individual characteristics.

Table 4: Endogenous seating arrangement in the reference groups

	Reference group			Random	*t*-stat	P(T>|*t*-stat|)
	Mean	S.D.	Obs.			
	(1)	(2)	(3)	(4)	(5)	(6)
Bottom-Bottom	9.62%	12.40%	26	6.67%	1.214	0.118
Middle-Middle	5.77%	10.70%	26	6.67%	-0.427	0.360
Top-Top	9.62%	12.40%	26	6.67%	1.214	0.118
Top-Middle	30.80%	17.80%	26	26.67%	1.184	0.124
Top-Bottom	17.30%	17.00%	26	26.67%	-2.809	0.011
Middle-Bottom	26.90%	18.60%	26	26.67%	0.064	0.475

Note: For each group block in the reference groups, we calculate the fraction of each type of neighboring pair: bottom-bottom, middle-middle, top-top, top-middle, top-bottom, and middle-bottom. Columns 1 and 2 show the means and standard deviations. Column 4 reports the theoretical probability of each pair type under a purely random arrangement. We test whether the average fraction is statistically different from the theoretical probability, and columns 5 and 6 report the *t* statistics and *p*-values. This exercise excludes 10 reference groups that are missing the baseline scores for some group members.

Table 5: Treatment effects on spatial layouts of peer characteristics

Panel A: Academic abilities of neighbor peer(s)

	Dummy: sitting next to top students			Baseline test score of neighbor peer(s)		
	Bottom (1)	Middle (2)	Top (3)	Bottom (4)	Middle (5)	Top (6)
Treatment	0.601***	0.464***	0.119	0.291***	0.255**	0.077
	(0.058)	(0.047)	(0.117)	(0.090)	(0.108)	(0.122)
Adjusted p-value (Westfall & Young, 1993)	0.000	0.000	0.3336	0.00061	0.0128	0.5381
Adjusted p-value (Young, 2019)	0.000	0.000	0.3850	0.0260	0.0670	0.5600
Observations	145	146	123	145	146	123
R-squared	0.530	0.370	0.211	0.464	0.342	0.476

Note: *** p<0.01, ** p<0.05, * p<0.1. In all the regressions, we control for a set of individual characteristics including each student's standardized baseline score, gender, age, a dummy variable indicating whether the student is a single child, the student's family's annual income, and parents' educational attainments. We also include the classroom fixed effects. The standard errors in parentheses are clustered at the study group level. When there are two neighbor peers, we use the average of the respective peer attributes.

Table 5: Treatment effects on spatial layouts of peer characteristics (cont'd)

Panel B: Other observed characteristics of neighbor peer(s)

	Panel B-1: Bottom-third students					
	(1)	(2)	(3)	(4)	(5)	(6)
	Female	Age	Single-child	Household income	College-educated father	College-educated mother
Treatment	-0.111	0.009	0.055	0.676	0.045	-0.034
	(0.107)	(0.100)	(0.080)	(1.098)	(0.096)	(0.109)
Adj. p-value (Westfall & Young, 1993)	0.3343	0.9675	0.4780	0.4673	0.6495	0.7372
Adj. p-value (Young, 2019)	0.2980	0.9320	0.5270	0.5300	0.6710	0.7300
Observations	147	147	147	147	147	147
R-squared	0.120	0.073	0.217	0.177	0.090	0.114
	Panel B-2: Middle-third students					
	(1)	(2)	(3)	(4)	(5)	(6)
	Female	Age	Single-child	Household income	College-educated father	College-educated mother
Treatment	-0.121	-0.012	0.035	0.804	-0.007	-0.070
	(0.080)	(0.069)	(0.077)	(0.958)	(0.085)	(0.102)
Adj. p-value (Westfall & Young, 1993)	0.1277	0.8169	0.6213	0.4020	0.9569	0.4686
Adj. p-value (Young, 2019)	0.2150	0.8780	0.6850	0.4470	0.9100	0.4510
Observations	148	147	147	147	147	147
R-squared	0.106	0.070	0.106	0.132	0.189	0.075
	Panel B-3: Top-third students					
	(1)	(2)	(3)	(4)	(5)	(6)
	Female	Age	Single-child	Household income	College-educated father	College-educated mother
Treatment	0.066	0.044	-0.110	-1.220	0.062	-0.048
	(0.085)	(0.053)	(0.084)	(0.937)	(0.100)	(0.090)
Adj. p-value (Westfall & Young, 1993)	0.4149	0.3167	0.1946	0.2153	0.4519	0.6152
Adj. p-value (Young, 2019)	0.4330	0.4810	0.1490	0.2380	0.5100	0.5250
Observations	125	124	124	124	124	124
R-squared	0.156	0.141	0.190	0.094	0.166	0.161

Note: *** p<0.01, ** p<0.05, * p<0.1. In all the regressions, we control for a set of individual characteristics including each student's standardized baseline score, gender, age, a dummy variable indicating whether the student is a single child, the student's family's annual income, and parents' educational attainments. We also include the classroom fixed effects. The standard errors in parentheses are clustered at the study group level. When there are two neighbor peers, we use the average of the respective peer attributes.

Table 6: Treatment effects on test scores

Dependent variable: Average standardized score of post-treatment exams							
	Panel A: Bottom-third students						
	Total score	Chinese	Math	English	Physics	Chemistry	Biology
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.165*	0.079	0.037	-0.234	0.257**	0.258**	0.261**
	(0.093)	(0.119)	(0.076)	(0.153)	(0.120)	(0.124)	(0.116)
Adj. p-value (Westfall & Young,1993)	0.0753	0.5065	0.6235	0.1273	0.0323	0.0377	0.0238
Adj. p-value(Young, 2019)	0.0850	0.5810	0.6780	0.1140	0.0530	0.0560	0.0380
Observations	144	145	145	145	144	145	145
R-squared	0.797	0.560	0.805	0.513	0.681	0.609	0.577
	Panel B: Middle-third students						
	Total score	Chinese	Math	English	Physics	Chemistry	Biology
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.122**	-0.111	-0.083	-0.163**	-0.025	-0.083	-0.169**
	(0.054)	(0.089)	(0.078)	(0.082)	(0.076)	(0.074)	(0.082)
Adj. p-value (Westfall & Young,1993)	0.0229	0.2141	0.2914	0.0457	0.7443	0.2642	0.4051
Adj. p-value(Young, 2019)	0.0520	0.3050	0.3180	0.1150	0.7550	0.3350	0.0560
Observations	147	147	147	147	147	147	147
R-squared	0.676	0.289	0.535	0.451	0.490	0.535	0.531
	Panel C: Top-third students						
	Total score	Chinese	Math	English	Physics	Chemistry	Biology
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.047	0.070	0.120	0.055	0.052	-0.059	-0.001
	(0.045)	(0.095)	(0.076)	(0.08)	(0.059)	(0.046)	(0.072)
Adj. p-value (Westfall & Young,1993)	0.2914	0.4638	0.1125	0.4917	0.3770	0.1974	0.9891
Adj. p-value(Young, 2019)	0.3480	0.4580	0.1140	0.5010	0.4480	0.3600	0.9820
Observations	124	124	124	124	124	124	124
R-squared	0.700	0.295	0.477	0.465	0.520	0.583	0.452

Note: ***p<0.01, ** p<0.05, * p<0.1. In all the regressions, we control for a set of individual characteristics including each student's standardized baseline score, gender, age, a dummy variable indicating whether the student is a single child, the student's family's annual income, and parents' educational attainments. We also include the classroom fixed effects. The standard errors in parentheses are clustered at the study group level.

Table 7: Positive spillovers from top students

	Dummy: help others with study problems	
	Full sample	Top-third students
	(1)	(2)
Dummy: middle third	0.0564	
	(0.0533)	
Dummy: top third	0.181***	
	(0.0631)	
Treatment		0.204*
		(0.103)
Adjusted p-value (Young, 2019)		0.072
Observations	418	124
R-squared	0.057	0.134

Note: *** p<0.01, ** p<0.05, * p<0.1. The survey following the experiment asked a given student to rate, on a scale from one (strongly disagree) to five (strongly agree), how much he or she agreed with the following statement: "I spent some time helping other group members with their study problems and considered this an opportunity to enhance learning for myself." The dependent variable in this table is a dummy variable indicating whether the student selected "strongly agree." See Table 6 for notes.

Table 8: Disruptions for middle students in treated groups

	Middle students from all treated and reference groups			Middle students from special treated groups and all reference groups		
	Chitchats	Competition	Help others with study problems	Chitchats	Competition	Help others with study problems
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	1.087*	-0.025	-0.018	0.454	-0.034	0.090
	(0.635)	(0.019)	(0.086)	(1.457)	(0.037)	(0.183)
Adjusted p-value (Young, 2019)	0.1690	0.2580	0.8480	0.8410	0.9200	0.7810
Observations	147	147	147	85	85	85
R-squared	0.099	0.103	0.156	0.159	0.153	0.274

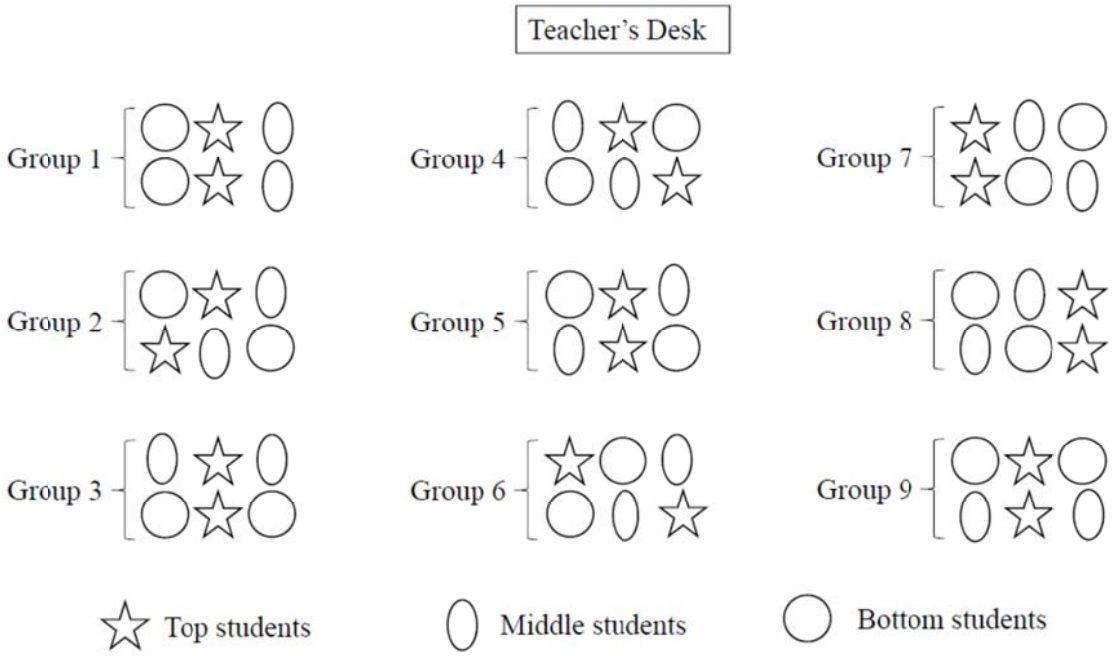
Note: *** p<0.01, ** p<0.05, * p<0.1. The dependent variables in columns 1 and 4 are the times in a week a given student chatted with other group members about personal issues not related to study. The dependent variables in columns 2 and 5 are a dummy variable indicating whether the student strongly agrees that "I felt that peer interactions between group members were mitigated by competition." The dependent variables in columns 3 and 6 are the dummy variable indicating whether the student strongly agrees that "I spent some time helping other group members with their study problems and considered this an opportunity to enhance learning for myself." The regressions in columns 4 to 6 use the middle students from all the reference groups and those from the treated groups where they sat on the same side. See Table 6 for additional notes.

Table 9: Treated groups with middle students sitting on the same side

	Middle-third students						
	Total score	Chinese	Math	English	Physics	Chemistry	Biology
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.009	-0.008	0.030	-0.038	0.112	-0.115	-0.059
	(0.090)	(0.166)	(0.115)	(0.191)	(0.099)	(0.120)	(0.101)
Adj. p value (Westfall & Young, 1993)	0.9241	0.9625	0.7969	0.8440	0.2569	0.3387	0.5626
Adj. p value (Young, 2019)	0.9390	0.9600	0.8300	0.9030	0.5750	0.5960	0.7060
Observations	85	85	85	85	85	85	85
R-squared	0.762	0.280	0.660	0.450	0.591	0.602	0.663

Note: The regressions in the table use the middle students from all the reference groups and those from the treated groups where they sat on the same side. See Table 6 for notes.

Figure 1: Spatial layout of a classroom

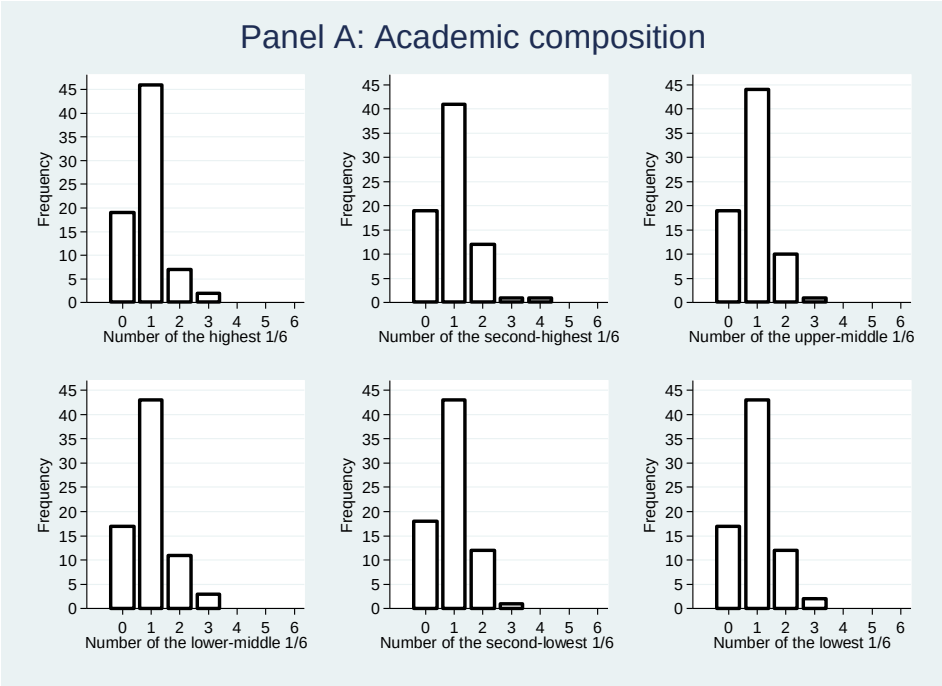


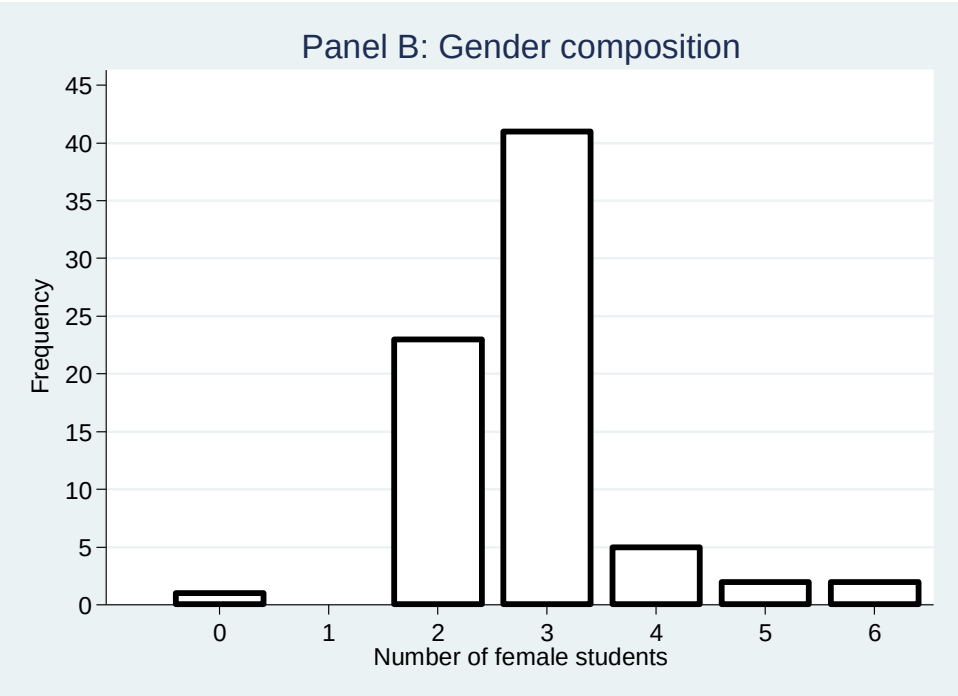
Note: Top students, represented by the stars, are the students who belong to the top one-third in the distribution of the baseline exam score. Middle students, represented by the ovals, are the students who belong to the middle third. Bottom students, represented by the circles, are the students who belong to the lowest third. Group 1, Group 3, Group 5, and Group 9 are the treated groups. The remaining groups are the reference groups.

Figure 2: Study timeline

	4 weeks		1 week	4 weeks	4 weeks	8 weeks
	11-Sep	9-Oct ~ 13-Oct	20-Oct	20-Nov	19-Dec	9-Feb-18
Researchers communicated with teachers and provided instructions.	※					
Students took the baseline exam. Teachers assigned groups. Teachers announced group assignments and asked students for their seat preferences. Seating arrangements were formed.		※				
Researchers conducted the survey.			※			
Students took the second exam (the midterm exam).				※		
Students took the third exam.					※	
Students took the last exam (the final exam).						※

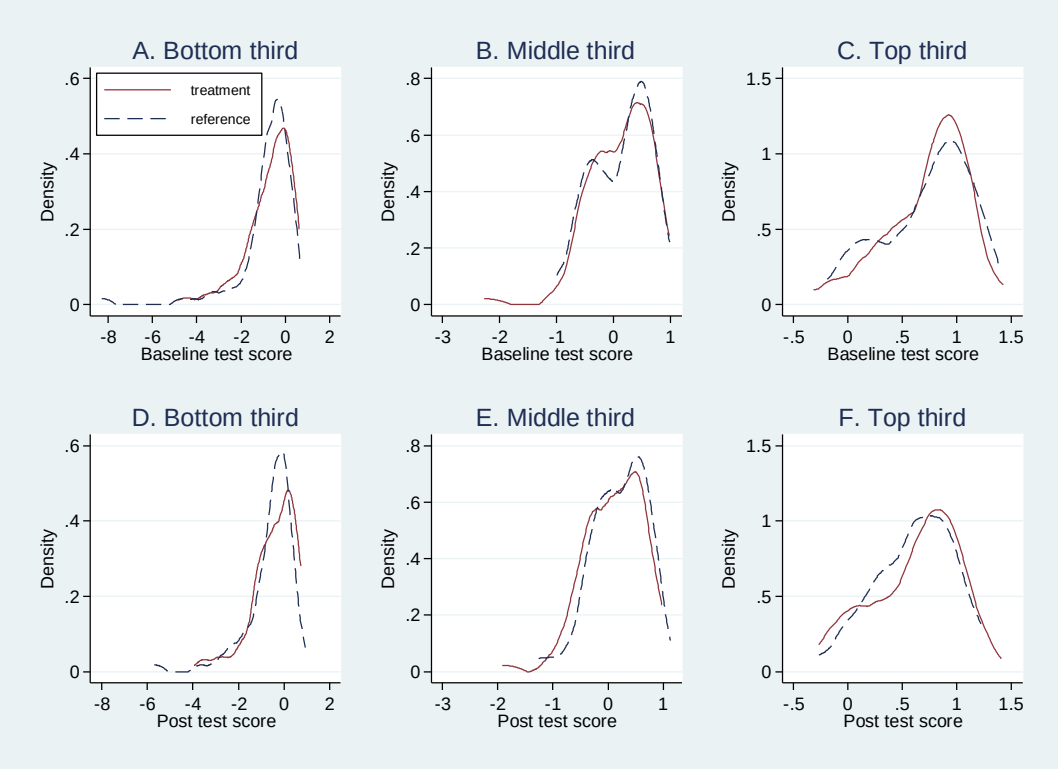
Figure 3: Peer compositions of study groups





Note: The graphs use all of the 74 study groups in our sample.

Figure 4: Density distributions of test scores



Note: In Panels D to F, post test score represents the average of the standardized scores on the post-treatment exams. The solid lines represent the treatment group. The dashed lines represent the reference group.

Supplementary Tables

Table A1: Baseline individual-level characteristics by treatment status for each one-third of students by baseline score

Panel A: Bottom-third students

	Treatment group			Reference group			Treatment-Reference
	Mean	S.D.	Obs.	Mean	S.D.	Obs.	t stat
Standardized baseline total score	-0.728	1.070	76	-0.783	1.307	72	0.376
Standardized baseline Chinese score	-0.321	0.927	76	-0.466	1.380	72	0.416
Standardized baseline math score	-0.716	1.183	76	-0.611	1.249	72	-0.737
Standardized baseline English score	-0.615	1.280	76	-0.427	0.983	72	-1.078
Standardized baseline physics score	-0.605	1.154	76	-0.665	1.266	72	0.144
Standardized baseline chemistry score	-0.501	0.984	76	-0.685	1.501	72	0.893
Standardized baseline biology score	-0.580	1.037	76	-0.718	1.165	72	1.465
Dummy: female student	0.487	0.503	76	0.569	0.499	72	-1.236
Age in months	15.321	0.414	76	15.271	0.431	71	1.363
Dummy: single child	0.645	0.482	76	0.746	0.438	71	-0.541
Family annual income, 10,000 yuan	9.020	5.535	76	9.937	5.186	71	0.0917
Dummy: father having a college degree	0.592	0.495	76	0.676	0.471	71	-0.878
Dummy: mother having a college degree	0.513	0.503	76	0.620	0.489	71	-1.037

Note: The *t*-statistic is for the null hypothesis that the difference between the means of the treatment and reference groups is zero. All *t*-statistics are adjusted for clustering at the study group level. In each test, we control for the classroom fixed effects.

Table A1: Baseline individual-level characteristics by treatment status for each one-third of students by baseline score (cont'd)

Panel B: Middle-third students

	Treatment group			Reference group			Treatment-Reference
	Mean	S.D.	Obs.	Mean	S.D.	Obs.	t stat
Standardized baseline total score	0.125	0.548	76	0.136	0.504	71	0.151
Standardized baseline Chinese score	-0.029	0.803	76	0.058	0.863	71	-0.226
Standardized baseline math score	0.237	0.553	76	0.067	0.688	71	0.803
Standardized baseline English score	-0.008	0.842	76	0.227	0.738	71	-1.202
Standardized baseline physics score	0.182	0.635	76	0.090	0.698	71	1.227
Standardized baseline chemistry score	0.149	0.572	76	0.052	0.700	71	0.715
Standardized baseline biology score	-0.022	0.854	76	0.168	0.674	71	-1.201
Dummy: female student	0.487	0.503	76	0.417	0.496	72	0.997
Age in months	15.288	0.310	76	15.297	0.363	72	-0.799
Dummy: single child	0.684	0.468	76	0.708	0.458	72	-0.996
Family annual income, 10,000 yuan	9.928	5.177	76	8.611	4.987	72	0.737
Dummy: father having a college degree	0.539	0.502	76	0.542	0.502	72	1.386
Dummy: mother having a college degree	0.526	0.503	76	0.597	0.494	72	0.0285

Panel C: Top-third students

	Treatment group			Reference group			Treatment-Reference
	Mean	S.D.	Obs.	Mean	S.D.	Obs.	t stat
Standardized baseline total score	0.732	0.384	66	0.725	0.399	61	0.917
Standardized baseline Chinese score	0.437	0.842	66	0.447	0.659	61	-0.155
Standardized baseline math score	0.605	0.495	66	0.585	0.545	61	0.345
Standardized baseline English score	0.530	0.759	66	0.443	0.641	61	1.908
Standardized baseline physics score	0.595	0.556	66	0.563	0.538	61	0.333
Standardized baseline chemistry score	0.592	0.531	66	0.546	0.501	61	0.260
Standardized baseline biology score	0.619	0.532	66	0.731	0.524	61	-0.530
Dummy: female student	0.455	0.502	66	0.419	0.497	62	0.576
Age in months	15.191	0.326	66	15.311	0.419	59	-1.349
Dummy: single child	0.667	0.475	66	0.814	0.393	59	-0.614
Family annual income, 10,000 yuan	9.879	5.602	66	9.280	5.156	59	0.180
Dummy: father having a college degree	0.636	0.485	66	0.729	0.448	59	-0.728
Dummy: mother having a college degree	0.561	0.500	66	0.593	0.495	59	-0.615

Note: The *t*-statistic is for the null hypothesis that the difference between the means of the treatment and reference groups is zero. All *t*-statistics are adjusted for clustering at the study group level. In each test, we control for the classroom fixed effects.

Table A2: Random treatment assignment for each one-third of students by baseline score

Dependent variable: Dummy variable indicating the treatment group								
	All students		Bottom		Middle		Top	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standardized baseline total score	0.015		0.026		-0.015		0.178	
	(0.016)		(0.036)		(0.096)		(0.166)	
Standardized baseline Chinese score		0.000		0.018		-0.000		0.030
		(0.020)		(0.036)		(0.053)		(0.056)
Standardized baseline math score		-0.021		-0.065		0.064		0.073
		(0.032)		(0.046)		(0.058)		(0.097)
Standardized baseline English score		-0.021		-0.044		-0.073		0.110*
		(0.023)		(0.039)		(0.052)		(0.057)
Standardized baseline physics score		0.014		-0.024		0.086		-0.008
		(0.028)		(0.043)		(0.080)		(0.089)
Standardized baseline chemistry score		0.038		0.047		0.043		0.074
		(0.036)		(0.050)		(0.073)		(0.101)
Standardized baseline biology score		0.001		0.098*		-0.083		-0.072
		(0.028)		(0.051)		(0.050)		(0.107)
Dummy: female student	-0.007	-0.000	-0.082	-0.078	0.062	0.103	0.020	0.014
	(0.025)	(0.025)	(0.082)	(0.093)	(0.069)	(0.069)	(0.082)	(0.081)
Age in months	-0.018	-0.012	0.113	0.127	-0.098	-0.121	-0.164	-0.164
	(0.050)	(0.051)	(0.095)	(0.108)	(0.111)	(0.108)	(0.148)	(0.152)
Dummy: single child	-0.045	-0.043	0.005	-0.011	-0.069	-0.077	-0.035	-0.010
	(0.032)	(0.033)	(0.078)	(0.075)	(0.066)	(0.078)	(0.081)	(0.088)
Family annual income, 10,000 yuan	0.003	0.004	0.004	0.008	0.004	0.007	0.004	0.003
	(0.005)	(0.005)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)
Dummy: father having a college degree	0.016	0.018	-0.041	-0.045	0.105	0.089	-0.063	-0.074
	(0.055)	(0.055)	(0.101)	(0.103)	(0.075)	(0.079)	(0.100)	(0.108)
Dummy: mother having a college degree	-0.049	-0.048	-0.045	-0.076	-0.045	-0.032	-0.063	-0.063
	(0.053)	(0.053)	(0.098)	(0.102)	(0.082)	(0.084)	(0.095)	(0.096)
Classroom fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Observations	418	418	147	147	147	147	124	124
R-squared	0.272	0.276	0.294	0.333	0.274	0.310	0.316	0.336
F-statistic	0.732	0.791	0.886	1.449	0.648	1.111	0.826	0.833
p-value	0.646	0.658	0.522	0.164	0.715	0.365	0.569	0.617

Note: We use the linear probability regression model. The standard errors are clustered at the study group level. The *F*-statistics are for testing the joint significance of all the observed individual characteristics.

Table A3: Treated groups with middle students sitting on the same side, additional results

	Panel A: Bottom-third students						
	Total score	Chinese	Math	English	Physics	Chemistry	Biology
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.345***	0.090	0.043	-0.343*	0.627***	0.496***	0.501***
	(0.102)	(0.162)	(0.124)	(0.199)	(0.181)	(0.183)	(0.152)
Adj. p value (Westfall & Young, 1993)	0.00068	0.5782	0.7270	0.0856	0.0005	0.0067	0.0010
Adj. p value (Young, 2019)	0.2060	0.6920	0.7520	0.4510	0.3660	0.3660	0.2010
Observations	82	83	83	83	82	83	83
R-squared	0.808	0.737	0.790	0.527	0.697	0.664	0.596

	Panel B: Top-third students						
	Total score	Chinese	Math	English	Physics	Chemistry	Biology
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.131*	-0.019	0.276**	-0.092	0.233**	0.102*	0.039
	(0.068)	(0.142)	(0.130)	(0.110)	(0.103)	(0.060)	(0.094)
Adj. p value (Westfall & Young, 1993)	0.0553	0.8928	0.0341	0.4056	0.0234	0.0895	0.6754
Adj. p value (Young, 2019)	0.2830	0.9250	0.3010	0.5610	0.3150	0.2060	0.7210
Observations	72	72	72	72	72	72	72
R-squared	0.682	0.314	0.578	0.375	0.552	0.596	0.421

Note: The regressions use the middle students from all the reference groups and those from the treated groups where they sit on the same side. See Table 6 for notes.

Table A4: OLS and IV estimation of peer effect for bottom students

				Panel A: OLS estimation			
Dependent variable: Average standardized score on post-treatment exams							
	Total score	Chinese	Math	English	Physics	Chemistry	Biology
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peer test score	-0.043	-0.053	-0.028	-0.092	-0.032	-0.041	0.049
	(0.076)	(0.099)	(0.085)	(0.099)	(0.110)	(0.098)	(0.098)
Adjusted p value (Westfall & Young,1993)	0.5674	0.5929	0.7405	0.3521	0.7671	0.6744	0.6167
Observations	142	143	143	143	142	143	143
R-squared	0.793	0.567	0.806	0.506	0.670	0.599	0.566
				Panel B: IV estimation			
Dependent variable: Average standardized score on post-treatment exams							
	Total score	Chinese	Math	English	Physics	Chemistry	Biology
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peer test score	0.557	0.268	0.126	-0.802	0.870*	0.878	0.887*
	(0.381)	(0.446)	(0.269)	(0.547)	(0.495)	(0.532)	(0.469)
Adjusted p value (Westfall & Young,1993)	0.1434	0.5478	0.6390	0.1426	0.0789	0.0986	0.0584
Observations	142	143	143	143	142	143	143
R-squared	0.529	0.384	0.624	-0.001	0.324	0.233	0.190
First stage F-stat	9.94	9.92	9.92	9.92	9.94	9.92	9.92

Note: See Table 6 for notes.

Table A5: Heterogeneous treatment effects for bottom students

Dependent variable: Average standardized score on post-treatment exams, bottom-third students							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.354***	0.211**	0.009	0.066	0.038	0.098	0.099
	(0.124)	(0.094)	(0.113)	(0.125)	(0.170)	(0.145)	(0.223)
Treatment*score gap	-0.179**						-0.151*
	(0.076)						(0.081)
Treatment*gender gap		-0.088					-0.064
		(0.102)					(0.115)
Treatment*dummy: female student			0.321*				0.325*
			(0.163)				(0.172)
Treatment*dummy: above median family income				0.241*			0.223
				(0.136)			(0.162)
Treatment*dummy: having a college-educated father					0.193		0.072
					(0.178)		(0.218)
Treatment*dummy: having a college-educated mother						0.114	-0.066
						(0.167)	(0.194)
Observations	142	144	144	144	144	144	142
R-squared	0.805	0.798	0.802	0.800	0.799	0.797	0.813

Note: See Table 6 for notes.