Complete Mixed-Seating, Incentivized Study Help, and Peer Effects: Evidence from a Randomized Experiment in Elementary Schools

Jia Wu[1], Junsen Zhang[2], and Chunchao Wang[3]

ABSTRACT

This study analyzes experimental data from a carefully designed seating arrangement in elementary schools in China to estimate the effects of a complete mixed-seating classroom and incentivized study help. We randomly set up two-thirds of classes as complete mixed-seating classrooms (MS) in which low- and high-ability are randomly paired as deskmates and sat together for 20 weeks. A tournament-type incentive based on the improved score of lower-track students is set among upper-track students in half of the MS classes (MSR). Being assigned in the MSR classes boosted lower-track students' mathematics scores by 0.24 standard deviations and improved lower- and upper-track students' extraversion and agreeableness by 5%–8%. The cognitive and non-cognitive performance of students in the MS classes showed minimal changes. The positive treatment effect can be attributed to the deskmate-level peer effects, which are associated with network structure changes: the incentivized complete mixed-seating classes promote social ties between deskmates and separate links among lower-track students.

Keywords: education, complete mixed-seating classroom, peer effects, friend networks
JEL codes: I20, I21, J24

1. Jia Wu, Department of Economics, Jinan University, Guangzhou, China. 2. Junsen Zhang, Corresponding Author, jszhang@cuhk.edu.hk, Department of Economics, 9/F, Esther Lee Building, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. 3 Chunchao Wang, Department of Economics, Jinan University; Jinan University–University of Birmingham Joint Institute, Guangzhou, China. The experiment in this paper is registered in the AEA RCT Registry (RCT ID: AEARCTR-0006127). The authors wish to thank many colleagues and friends for their helpful comments. The authors are grateful for the funding support provided by the National Natural Science Foundation of China (Grant No. 71703055), the National Social Science Fund of China (Grant No. 18ZDA081), and Fundamental Research Funds for the Central Universities (Grant Nos. 19JNQM18 and 17JNQN011). The views expressed in this article are solely the authors'.

1. Introduction

This study presents new evidence on the efficacy of a complete mixed-seating (or negatively assortative matching) classroom with incentives for students by using experimental data obtained from elementary schools in China. The achievement gap in students or adults is significant and has caused severe social concerns (Dobbie and Fryer, 2011; Clotfelter, Ladd, and Vigdor, 2009; Young, 2013). Exploring the effectiveness of peer effects on student performance is promising because simply reorganizing peer groups can benefit low-performing students by draining fewer resources than many educational programs. Burke and Sass (2013) show that the effectiveness of learning from peers may depend on their relative ability. Moreover, peer effects are strong if peers are close to one another and have more interactions (Bobonis and Finan, 2009; Feld and Zölitz, 2017). For these reasons, a mixed pairing of low and high achievers at the desk level may be an effective arrangement to enhance the performance of low achieving students. To improve the interaction between high- and low-achieving students, rewards can be provided for the former as an incentive to help the latter. Li, Han, Zhang, and Rozelle (2014) is the only study looking at this arrangement and finding a positive effect of the mixed pairing on the test score of low-achieving students. The current paper attempts to reassess the issue with an improved design owing to the importance of this new promising arrangement. Perhaps more importantly, we depart from Li et al. (2014) by carefully analyzing students' non-cognitive skills and the mechanisms or channels through which a mixed seating arrangement may or may not be effective in improving students' performance.

The current study brings this idea into an experiment in elementary schools in China. We develop a **complete** mixed-seating environment with two distinct schemes. In the first scheme, low- and high-ability students are divided into two tracks according to the median score of a previous examination. Students in the lower and upper tracks are randomly paired as deskmates and sit together for 20 weeks (Mixed-seating [MS] treatment group). In the second scheme, half of the complete mixed-seating classes are randomly selected, and a tournament-type incentive is provided for the upper-track students. To motivate high-achieving students to help low-achieving students, upper-track students are rewarded with a monetary award and a praise certificate if their deskmates' improved score ranks to reach the top 10% among lower-track students in class in the mid-term or final examinations. We refer to these classes as "mixed-seating with reward (MSR)."[1] In the control classes, students' seats are randomly assigned. We show that MSR intervention promotes lower-track students' mathematics scores and upper- and lower-track students' social skills.

Although tracking is widely used in the United States, Europe, and China (Betts, 2011),[2] an important concern on tracking is that gains by students in upper-track classes may be offset by losses of students in lower-track classes. We thus design a promising method to group students by pairing students with different abilities as deskmates to study whether students' achievement can be affected by those who sit next to them. The fixed-seat system in China's elementary schools, which requires students to sit beside the same deskmates throughout the semester, provides a good opportunity to explore this area with scant research (Li et al., 2014). We also provide incentives for high-ability students to help low-ability students in half of the complete mixed-seating classes.

Our inspiration for experimental design of mixing high- and low-ability students within a classroom comes from Li et al. (2014). We improve their experimental design from two aspects. First, Li et al. (2014) uses a **partial** pairing of approximately 10 low-achieving students in the bottom quartile and 10 high-achieving students in the top quartile, and show that incentivized low-achievers benefit from the paired high-achievers.[3] We use a **complete** mixed-sitting setting. Second, Li et al. (2014) award the top three progressive low-ability students, along with their deskmates. In comparison, we remove the incentives of low-ability students, which is useful to discern the effect of peer-to-peer interaction because such an approach can avoid the money-

[1] In this study, "complete mixed-seating," and "negatively assortative matching" are used interchangeably. Both denote classes with intervention (i.e., MS and MSR classes). "Lower track," "low performing," and "low ability" are also used interchangeably. Similarly, "upper track," "high performing," and "high ability" are used interchangeably.

[2] Tracking groups students with similar abilities together. Students are usually sorted into different classrooms within a school (within-school tracking) in the United States and into different types of schools (across-school tracking) in Europe and China. For example, public secondary schools in the United States divide students into high, middle, and low ability tracks. In Germany, students' achievements in the last four years of primary school determine the types of secondary school that students will be permitted to attend. In China, the scores of junior middle school graduates determine whether they can be admitted to elite or regular high schools.

[3] Li et al. (2014) design two treatments to study peer interaction. In the first treatment, the top three progressive low-ability students (i.e. low-ability students who achieve the greatest increase in scores between the baseline and endline tests) are rewarded. In the second treatment, the top three progressive low-ability students and their paired high-achieving deskmates are rewarded with an equivalent cash prize to study together. They design a partial paring process, in which 10 low-achieving students in the bottom quartile and 10 high-achieving students in the top quartile are paired as deskmates. Students in the two middle quartiles are not considered.

seeking motivation of low-ability students. In terms of empirical analysis, apart from assessing students' academic performance, we also examine the non-cognitive ability that is becoming increasingly important for both scholars and educational practitioners (Kautz et al., 2014; Akee, Copeland, Costello, and Simeonova, 2018). Most importantly, we collect data on friend networks and analyze the mechanism of peer effects by looking into the network. Jackson (2008) conjectures that networks of relationships play central roles in human interaction, which generates peer effects. However, empirical analysis remains lacking.[4] To complement the existing literature, we asked each student to name his/her friends in two rounds of surveys and identify the friends' network structure in each class in a survey. By examining the network data, we provide evidence to better understand the deskmate-level peer effects through the network ties between deskmates.

Our main empirical analyses and findings can be summarized as follows. First, after the treatment, students in the pure complete mixed-seating class without external incentives do not perform differently from those in the control class. However, when high-ability students are offered rewards as incentives to help their low-ability deskmates in the MSR classes, the mathematics scores of low-ability students improve by 0.24 standard deviations in the final examination. Results from a non-linear specification show that students whose initial mathematics scores are at the bottom or close to the median score distribution benefit most from the MSR intervention. We also find that high-ability students are nearly unaffected in the MS and MSR classes.

Second, we estimate the effect of MS and MSR interventions on students' non-cognitive skills, which are measured by the "Big Five" personality traits. We find that being assigned to the MSR classes improves lower- and upper-track students' extraversion by 6% and 5%, respective, and agreeableness by 8% and 7%, respectively. Given that extraversion and agreeableness are related to social skills, these findings indicate that the MSR intervention leads students to be more sociable. The level of conscientiousness, which relates to "organized and hardworking," is also promoted for lower-track students in the MSR classes, which is consistent with their score gain.

Third, we find that peer effects at the deskmate level underlie the mechanism of our MSR

[4] A possible explanation is that collecting network information is costly and time-consuming.

treatment. Estimates from a linear-in-mean peer effect model show that being assigned with incentivized strong deskmates in the MSR classes increases the lower-track students' test scores by 0.09 standard deviations. Lower- and upper-track students' sociability increases by 0.11–0.23 points when assigned with more sociable deskmates in the MSR classes. However, no peer effects are observed among students in the MS classes. These results imply that the rewards offered to upper-track students are important in generating peer effects.[5]

Fourth and most importantly, we present evidence that the network structure of students changes in the MSR classes. By matching the name of the nominated friends for each individual and the name of his/her classmates, we construct one network for each class. We find that at the end of the 20-week experiment, the MSR treatment increases the likelihood of deskmates to nominate each other as friends by 8.4–9.6 percentage points. Moreover, MSR treatment induces lower-track students to select less low-ability friends, and they nominate more high-ability friends instead. In parallel, upper-track students in the MSR classes select more low-ability friends. Overall, these findings present a plausible mechanism of the deskmate-level peer effects: the incentivized complete mixed-seating arrangement promotes between-track friendship formation. As friends, low-ability students can interact more easily with high-ability deskmates.

Finally, we present several tests to exclude the alternative explanations of the treatment effects as well as peer effects. (1) To prove that our treatment effects do not rely on teacher effort, we compare students' assessments of their teachers and find them unchanged before and after the intervention. To exclude the possibility of teacher quality factors in the detected peer effects in the MSR classes, we compare peer effects in classes with high- and low-educated teachers. We find that teacher quality does not matter. (2) To further rule out alternative explanations of the treatment effects, we show that the gender composition within a desk is not related to the test score gain of low-ability students in the MSR classes. (3) We also exclude the possibility that our results are driven by deskmate's collusion to cheat in the exam.

---

[5] We complement the literature on micro-level peer effects as evidence on peer effects at the deskmate level is limited. Lu and Anderson (2015) appear to be the only one study on the relation between gender composition of neighboring-seating students and students' score. They randomly assign students' seats for the entire class and find that being surrounded by five females rather than five males increases female students' test scores. Different from Lu and Anderson's (2015) work, we examine the complete mixed-seating classroom with and without incentivized high-ability students to explore the score change of students.

This paper contributes to a burgeoning literature that focuses attention on sorting students within a classroom (Burke and Sass, 2013; Booij, Leuven and Oosterbeek, 2017; Li et al., 2014). If relative ability matters for peer effects, a mixed classroom may be a promising way to engender peer benefits for low-ability students. Is mixed-seating classroom helpful in narrowing score gaps among students? How to prompt high-ability students to help low-ability students? Is one's non-cognitive skill affected by the characteristics of peers who sit next to him/her? These questions to date have no clear answers. Hence, by designing a complete mixed-seating classroom with and without incentivized high-ability students, this study sheds some light on these questions. This study will also be valuable to policymakers who are concerned about egalitarianism in education.

In our second contribution, we provide empirical evidence on the shrouded mechanism behind the peer effects by exploring the student network. Although the literature suggests that the influence of friends on behavior can be substantial (Jackson, 2013), there are limited studies to associate peer effects with social relationships. Carrell et al. (2013) indicate that low-ability students are more likely to have low predicted GPA partners and friends, which can explain the negative peer effects on low-achievers. Zárate (2019) shows that being assigned to more sociable/high-achieving peers increases the number of sociable/high-achieving friends. The two studies emphasize network homophily. Two theoretical studies also focus on this issue. Calvó-Armengol, Patacchini, and Zenou (2009) show that the link between network centrality and student outcome exists at the Nash equilibrium. Bramoullé, Djebbari, and Fortin (2009) discuss necessary, sufficient conditions for identification of the endogenous and exogenous peer effects under a network interaction.[6] [7] Different from these studies, the present study searches for empirical evidence in an environment with exogenous interventions and focuses on the evolution of networks in a class. We show that the MSR intervention promotes friendship formation between deskmates with different abilities. Moreover, being assigned with incentivized high-ability deskmates helps separate the connection among low-ability students (homophily).

[6] Several studies investigate the effect of social relationships based on the quasi-randomness of network. For example, Fletcher and Ross (2018) find that an individual's smoking and drinking behavior responds to the behavior of his/her friends. Lavy and Sand (2018) show that the number of friends has effects on students' educational outcomes depending on the type of friends. However, these studies do not discuss the relation between the network and peer effects.

[7] The MSR intervention may stimulate the network ties between randomly assigned deskmates because it shortens the spatial distance between them. Deskmates can conveniently interact with and help each other.

The remainder of this paper is structured as follows. Section 2 introduces the background of primary school education in China and the county in which our experiment was implemented. Section 3 describes our experiment design. Section 4 presents the model specifications and the main results. Section 5 examines the possible mechanisms of the treatment effect and peer effects. Section 6 discusses several related issues. Section 7 presents robustness checks. Section 8 compares our findings with the related literature, and the last section concludes the paper.

2. Background

China implements a nine-year compulsory education system based on the compulsory education law enforced in 1986. Any child who has reached six years old should enter primary school (Grades 1–6). In areas where conditions are unfavorable, the starting age of the child to receive primary education can be postponed to seven years old. Most teachers in primary schools are hired by the local government (at least in our surveyed county). All primary school graduates should enter nearby middle schools (Grades 7–9) without taking entrance exams. Therefore, students face weak competition in schools due to the absence of a selection process for primary school graduates. In this context, high-ability students would suffer no loss in offering help to their low-ability classmates; thus, low-ability students can easily seek help from high-achieving students in their class. After finishing their nine-year compulsory education, about one-third of graduates either enroll in vocational high schools or enter the labor market directly. Students who seek to continue high school (Grades 10–12) must pass an entrance exam to be accepted into the school.

A school year in primary education is divided into two semesters. A semester comprises 19 weeks with an additional week for exams. The curriculum of primary schools is designated by the Ministry of Education and is uniform across the entire nation. All students in primary schools use uniform textbooks. Chinese language and mathematics are the two major courses, which are usually taught five days a week.[8] Students need to take end-of-term exams and earn scores at the end of each semester. Grades in other subjects, such as history, art, and music, are assigned by teachers without the need for examinations. In primary schools, student seats in one class are determined by the class head teacher. Typically, students' seats are arranged according to height,

[8] In some locations, English is also a major subject.

that is, short students sit in front of tall students in a class. Class head teachers have the discretion to adjust student seats in special cases. For example, pupils who are short-sighted or have amplyacousia are assigned to sit in the front of the classroom. Students can neither decide where they want to sit nor select their deskmates. After the arrangement, the seats are fixed for a long time (usually one or more than a year), except for two columns rolling at the same time per week. A graph of a typical desk is shown in the Appendix (Figure A1).[9] The rolling columns do not change a student's deskmate because the two columns roll together. Students study on their assigned seats in the class, which provides us an opportunity to investigate the effect of their deskmates.

Our experiment was implemented in four elementary schools in Longhui County, Hunan Province, China. Longhui is a relatively poor rural area. According to the National Bureau of Statistics of China, the county's per capita GDP was 12,466 CNY (1,945 USD) in 2016, which was less than a quarter of the national level (53,980 CNY or 8,487 USD). However, the educational resource of the county is close to the national average. For example, according to the Longhui Statistical Yearbook 2015, the enrollment rate was universal (100%) for primary school and almost universal (99.2%) for middle school due to the nine-year compulsory education in China. About three-quarters of students (74.7%) proceeded to study in high school in Longhui. The student–fulltime teacher ratio is 17.2, indicating that an average of 17 students are instructed by one teacher, which is similar to the national level (17.05). In 2016, the average public expenditure per incumbent student in Longhui was 5,770 CNY (907 USD), which was higher than the average level in China (4,408 CNY, or 693 USD). For the college admission rate, approximately 12.5% of high school graduates enrolled in tier 1 universities, which is similar to the national average (12.9%). In summary, the education resource and teaching conditions of primary schools in our experimental county are similar to the national averages.

[9] Class sizes usually range from 20 to 60 students depending on the number of students in a specific cohort and the school size.

3. Experiment

3.1 Study design

In collaboration with the Longhui Educational Bureau, we selected four primary schools to participate in our experiment in the first semester of 2015 (September 2015–February 2016). The experiment lasted for 20 weeks. The four schools were not selected randomly. We required each selected school to have at least three classes in each grade from Grades 3–5 because two types of treatment were introduced, and randomization was implemented at the school-by-grade level. All selected schools are commuter public schools. At the beginning of the first semester, we randomly selected three classes from each grade in each school as our target classes. A total of 36 classes were selected. We selected two classes randomly in each grade in each school as the treated classes (24 treated classes). These classes were assigned as the complete mixed-seating classrooms, in which low- and high-ability students were randomly paired as deskmates. From the 24 complete mixed-seating classes, half were assigned the MSR strategy, and the other half were the pure MS classes. The remaining 12 classes (one class each from Grades 3–5 of the four schools) were the control classes. Our initial sample includes 695 students in the MS classes, 672 students in the MSR classes, and 667 students in the control classes.

At the beginning of the semester, the seats of students in each of the selected 24 treated classes were reassigned to form the complete mixed-seating environment. The basic rules for assigning seats were as follows. First, students in a class were divided into two tracks according to the median score of their average performance on the Chinese and mathematics in the previous end-of-term exam. If a student's average score on the two courses was higher than the median score of the class, then he/she will be assigned to the upper track. Otherwise, he/she will be assigned to the lower track. Second, students in a class were sorted from the shortest to the tallest within each track and then categorized into tertiles. Specifically, the top 30%, the middle 40%, and the bottom 30% height groups were generated for the upper- and lower-track students. We considered student height in our setting to prevent tall students seated near the blackboard from blocking the view of the short students seated behind them. Finally, we implemented a random sequence of pairing students. We started by assigning lower- and upper-track students in the short-height groups. We randomly selected one student from the upper track and one student from the lower track as deskmates, and the pair was assigned a desk in the front left area of the classroom. Then, we randomly selected the next pair of students, one from each of the two tracks, and assigned them with a desk next to the prior pair of students. Figure 1 illustrates the desk arrangement in a typical classroom. The order of the seating arrangement in the physical space of the classroom for each pair is from left to right and from front to back. After the seats of students in the short-height group were assigned, we assigned seats for students in the middle-height group. Finally, we assigned seats for students in the tall-height group. In case a class had an odd number of students, we picked the tallest student in the lower track and assigned this single student to a desk in the last row of the classroom.[10] In our design, taller students mostly sat behind shorter students.[11]

Next, we selected one of the two treated classes in each grade in each school to be the MSR treated class, leaving the 12 classes as pure MS classes. Thus, we had 12 MSR and 12 pure MS classes. In the 12 MSR classes, we offered monetary awards and praise certificates as incentives for the 30 upper-track students in the middle and the end of the semester. We set award targets on the basis of the learning progress of the grade quantiles of the average scores that their low-ability deskmates achieved. Specifically, an upper-track student was offered 100 CNY if his/her deskmate's improved score ranked in the top 10% in the midterm or final exams among lower-track students in the class. Given that the average class size is 50, a 10% award rate implies that 2.5 students in an MSR class can receive rewards for each exam (in total, 12×2.5=30 upper-track

[10] Several students were unsatisfied with the seating assignment and requested to change seats after the seating arrangement. In our sample, 37 students were not in favor of the original seating arrangement (half of them were in the upper track, and the other half were in the lower track). This number accounts for 3% of the treated sample. We allowed the class head teacher to make minor adjustments to the seats for students who requested to change seats. An unsatisfied student would be re-paired with a new deskmate who was in the same track and same height group as his/her original assigned deskmate. Table A1 of the Appendix describes non-compliers and their original deskmates. We also report the non-compliers' new deskmates and students who eventually sit with non-compliers' original deskmates in detail (we call them the "extended non-compliers"). The 37 students have slightly higher baseline scores and comprise a higher fraction of girls than students who accepted the seating arrangement. Furthermore, they score a high level of neuroticism in the "Big Five" personality traits. In our main analysis on the treatment effects, we treat the 37 non-compliers as normal students because they account for a very small fraction of the treated sample. Furthermore, an unsatisfied student is only allowed to exchange a new deskmate in the same track and in the height group with his/her original assigned deskmate, and thus, this scenario does not affect the complete mixed-seating arrangement of the class. Even if we consider the non-compliers, their original deskmates, and the extended non-compliers together, the number of non-randomly assigned students only accounts for 8% of the overall sample. We use two methods to deal with these non-randomly assigned students. First, we drop them from the treatment classes. Second, we use two-stage least squares to estimate the treatment effects. We instrument the acceptance of the seating arrangement with the initial assigned seat. Both alternatives qualitatively yield the same results, which are available from the corresponding author upon request.

[11] The average heights for the top 30%, middle 40%, and bottom 30% students are 52.4 inches, 50.8 inches and 50 inches, respectively.

students are offered rewards in each exam). Midterm and final exams were conducted during the experiment. Therefore, an upper-track student may receive a packaged benefit of 200 CNY in total (or approximately 30 USD when using the nominal exchange rate to transfer). This amount can be considered a financial incentive for upper-track students to help their low-ability deskmates on study.[12] The reward plan was announced in the MSR classes after the completion of the seating arrangement, but no students in the MS and control classes were informed about the reward plan. The total expense of our program is 6,000 RMB (or 900USD), which is a relatively low cost to motivate high-ability students to offer their help to low-ability students.

After the seats were arranged in the MS and MSR classes, each pair of students who shared the same desk was fixed during the experiment. To prevent students from being assigned at the edge of the classroom for the whole semester and developing short-sightedness, columns were rolled during the experiment without changing the row. Specifically, two columns as a unit were rotated clockwise every week. Therefore, the deskmate for each student was fixed throughout the experiment.

Students stayed in school from 8:30 A.M. to 4:10 P.M. daily during weekdays. Each day, six lessons are conducted by a teacher, and one is done through self-study. Each lesson lasts 40 minutes followed by 10-minute breaks. Students must stay in their seats during the lessons and self-study courses. Although students in the same desk are not permitted to talk loudly during lectures, they can communicate easily with their deskmates in a soft voice. In the self-study lesson, students can communicate with their deskmates conveniently. Students cannot easily communicate with students across the aisles, because each of the two columns is bordered by aisles,. Students can talk to other students seated in the front or behind them, but this scenario is not as convenient as communicating with their deskmates because a student needs to tap or turn around to talk with those in front or behind him/her. In summary, we concentrate on deskmates, because students sharing the same desk are the most conveniently interactive peer group.

[12] The average tuition fee in our selected schools is between 100 CNY (15 USD) and 300 CNY (45 USD) per semester. In another survey conducted in the same county, the data show that a pupil receives an allowance of approximately 350 CNY (52.5 USD) per year from their parents. From the allowances, a pupil spends 102, 85, and 163 CNY (15.3, 12.8, and 24.5 USD) on extracurricular books, stationery, and other things, respectively.

The protocol for the control group is simple, and no intervention is carried out. As in MS or MSR classes, students sat with the same deskmate for the entire semester. Similarly, two columns of desks were rotated together in the control classes every week.

3.2 Data

Two rounds of surveys and three rounds of examinations were conducted. The baseline tests on Chinese and mathematics were conducted in June 2015. The average scores of Chinese and mathematics in the baseline survey were used to separate the students into lower- and upper-track groups. The midterm and final exams were conducted in November 2015 and February 2016, respectively. The tests for the two subjects in the final exam were uniform across the entire county.[13] The students gave their answers on paper sheets. We recruited a teacher to proctor the exam process rigorously in each of the selected 36 classes. The students sat separately and were not allowed to look at the answer sheets of the other students.

Chinese and mathematics scores are used to measure students' cognitive abilities. We convert the original Chinese and mathematics scores into standardized scores (z-score). Figure A2 compares the density distributions of the baseline average z-score of the students, where the dashed line refers to the control group, and the solid lines refer to the MS or MSR group. The graph shows that the baseline test scores are balanced between the control and treatment groups in our experimental design. Additionally, the Wilcoxon rank-sum test fails to reject the null hypothesis that the MS/MSR treatment and control samples are random draws from a single population (*p*-values equal to 0.69 and 0.87, respectively). The middle and lower parts of Figure A2 compare the baseline score distribution between each of the two treatment and control groups for the upper- and lower-track students, respectively. Similarly, we find no systematic differences in the observed baseline scores for the upper- and lower-track students between the treatment and control groups.

The baseline survey was conducted in September 2015, which was the beginning of a new semester. The MS and MSR interventions began at the same time and lasted for the entire semester

[13] The mid-term tests are generally assigned by subject teachers. Thus, these tests are not uniform across the county. Therefore, we use the final exam scores to measure students' achievement.

(20 weeks). A follow-up survey with similar questions was conducted in February 2016 after the final exam. In two rounds of the survey, we used a well-accepted model, namely, the “Big Five” personality test, to measure non-cognitive abilities, which are important elements of human capital and successful lives (Kautz et al., 2014). The “Big Five” includes five domains, namely, openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. The “Big Five” model has also been widely used to test personality traits in China (e.g., Zhang, 2011). The “Big Five” personality test employed in this study comprises 60-item questions. Each respondent described how satisfied he/she is in each question according to five levels, that is, from “very dissatisfied” to “very satisfied.” An answer of “very dissatisfied” is encoded as 1, and every increase in the satisfaction level incurs an additional score of 1. An answer of “very satisfied” is encoded as 5. The score of each domain is calculated by summarizing the items in the domain. We collected students’ “Big Five” personality traits in two rounds of the survey.

In two rounds of the survey, we also designed a questionnaire asking for the student’s personal information, such as gender, age, height, weight, interest in each subject, and so on. We surveyed the parents or guardians of the students to collect information on parents’ characteristics, such as age, education level, income level, and time inputs on the students. Figure 2 illustrates the timeline of our data collection and experiment.

Among the students registered in the 36 classes, 2,034 students with baseline scores were collected in September 2015. A total of 109 students who might transfer to other schools or have grade retention issues did not participate in the midterm or final exams. Students without deskmates or with missing values of the key variables were excluded from our sample (123 students). Finally, we obtain a sample of 1,802 students.[14] Each track has 901 students.

Table 1 reports the means and differences by the treatment status for variables measuring student cognitive and non-cognitive achievement, individual characteristics, and family background. Each variable is regressed on two dummies indicating the MS and MSR interventions. We add school-by-grade fixed effects because the random assignment is conducted at each grade in each school. The standard error is clustered at the class level. Column 1 reports the coefficients

[14] We show that sample attrition is not incurred by treatments in Table 1.

on the constant, which represent the mean of the control group. Columns 2 and 4 present the coefficients on the treatment dummies, indicating the respective difference between each of the treatment and control groups.

Among the 1,802 students in the sample, the MS/MSR and control groups are statistically indistinguishable in terms of the differences between the baseline characteristics. First, the key outcome variables are similar in the three groups. The average Chinese and mathematics raw test scores of the MS/MSR group do not differ significantly from those of the control group (rows 1–2). The non-cognitive abilities, measured by the "Big Five" personality traits, are also similar among students in the treatment and control classes (rows 3–7). Second, no significant differences by treatment status can be observed in students' characteristics, parents' education level, and household income. For example, slightly more than half of the students are male. Most of the students are non-minority and without urban *hukou*, as suggested by the fact that 1% is the minority, and 95% of the students do not have urban *hukou*. The average number of years of education of the students' parents is nearly 9.5, representing a middle school education experience. Approximately 28% of households have computers, and 18% of households have a car. These variables do not differ statistically between the treatment and control groups.

One selection issue is the likelihood of completing the survey. As previously described, among the 2,034 students selected in our study in June 2015, 232 students are dropped from the main analysis because of transferring to other schools, grade retention, or missing test scores. As a result, 574 students are included in the control classes, and 634 and 594 students are included in the MS and MSR classes, respectively (last row, Table 1). The mixed-seating classroom program may affect the likelihood of completing the survey by increasing the probability of school transfer. For example, if students did not favor their assigned deskmates, they may transfer to another school. In view of this concern, the penultimate row of Table 1 reports the treatment effects on the probability of completing the survey. The estimates reported in the table comprise coefficients when the number of missing students is regressed on two treatment dummies with the standard error cluster at the class level. There is no evidence of the difference in the missing students between the two treatment groups and the control group. Hence, the missing students were not related to our treatment. For the most part, the differences between the MS/MSR and the control groups are minor, indicating that the design of the MS/MSR is clean.

As aforementioned, the general rule for class head teachers to assign seats in a classroom is on the basis of student height. To visualize this arrangement, we plot the average height of students for each row, separated for the control, MS, and MSR groups in Figure 3 (the solid line). The figure shows that student height increases with the row number in control classes. Taller students basically sit in front of shorter students in the MS and MSR classes, but no strictly positive correlation can be observed between student height and row number because students in the lower and upper tracks are randomly paired (without considering their height) as deskmates in each of the height groups in the class. We also superimpose the average score of students for each row, classified by treatment status (the dashed line). We find that student's scores are typically unrelated to the row number and student height in the control, MS, and MSR classes.

We further examine whether the distinction between two students in a class is a determining factor for them to be assigned as deskmates. For any pair of students in a class, we generate a dummy that takes a value of 1 if two students in a class are deskmates, and takes a value of 0 otherwise. We then employ a linear probability model to regress this dummy on the absolute difference in characteristics between a student and his/her potential deskmate. We conduct our regressions in the control, MS, and MSR groups, separately. We present the results in Table A2 in the Appendix. The reporting coefficients on student characteristics are small and insignificant, indicating that the distinction between students has no predictive power on the assigned deskmates. In comparison, the absolute difference in percentile rank is positively associated with the probability of students to be deskmates in the MS and MSR classes, implying that the treated classes are negatively pairing students with different performance to be deskmates. In summary, the seats are approximately randomly assigned in the control group, which is important for a clear identification of the MS and MSR treatment effects.

Results from the seating arrangement of students in the treatment and control groups are presented graphically in Figure 4. The figure shows histograms of the deskmates' baseline z-scores for low-ability students in a class by treatment status. The left graph shows that the MS classes assign lower-track students with deskmates who have much higher average scores than those in the control classes. The right graph shows a similar pattern in the MSR classes. These results confirm that our mixed-seating classrooms pair low- and high-ability students as deskmates.[15]

4. Empirical model specification and main estimation results

4.1 Empirical model specification

The basic regression model to identify the treatment effect is a school-by-grade fixed-effect model, which regresses student *i*'s endline achievement on the dummies indicating MS and MSR treatment statuses and the student's baseline achievement:

$$y_{ijg}^{endline} = \beta_0 + \beta_1 y_{ijg}^{baseline} + \gamma_1 MS_{ijg} + \gamma_2 MSR_{ijg} + \eta_g + \varepsilon_{ijg}, \qquad (1)$$

where $y_{ijg}^{endline}$ and $y_{ijg}^{baseline}$ denote the endline and baseline standardized test scores/non-cognitive skills of student i from class j and grade g, respectively. MS_{ijg} is a dummy variable that equals 1 if student i is in the MS class or equals 0 otherwise. Likewise, MSR_{ijg} is a dummy variable that equals 1 if student i is in the MSR class or equals 0 otherwise. In addition, η_g is a set of dummies controlling for time-invariant school-by-grade fixed effects, and ε_{ijg} is an error term consisting of unobserved parts that affect the endline test scores. The estimates for γ_1 and γ_2 are the average within-grade effects of the MS and MSR interventions, respectively. Considering the possible correlation in students' performance within the classroom, we cluster standard errors at the class level. Following Duflo, Dupas, and Kremer (2011) and Alan and Ertac (2019), we include the lagged dependent variable in the right hand of the regression equation.[16]

Under the conditional exogenous assumption of the two treatment dummies (MS and MSR), the estimates of the average treatment effects γ_1 and γ_2 are consistent. To improve the precision of the estimation, another specification is to include student and parent characteristics (gender, age,

[15] We also examine the correlation of baseline characteristics between deskmates for the control, MS, and MSR groups. The correlation coefficient on the percentile rank of baseline average achievement between deskmates is 0.49 for the control group, compared with −0.09 and −0.06 for the MS and MSR groups, respectively, implying that the treatment effectively creates a negative ability assortative pairing class. Positive relations are observed on deskmates' ages, heights, and weights in the control, MS, and MSR classes. This finding is expected because our sample of students belong to close birth cohorts. For other individual characteristics and family background, we find a rather weak correlation between deskmates in the control and treatment classes.

[16] An alternative specification is to use the first difference of the outcome as the dependent variable. We find that the main results from this specification are almost the same as those from Equation (1).

height, weight, *hukou* register status, race, father's education, mother's education, a dummy indicating whether the student's household has a computer, and a dummy indicating whether the student's household has a car). The model can be written as:

$$y_{ijg}^{endline} = \beta_0 + \beta_1 y_{ijg}^{baseline} + \gamma_1 MS_{ijg} + \gamma_2 MSR_{ijg} + X_{ijg}\Gamma + \eta_g + \varepsilon_{ijg}, \qquad (2)$$

where X_{ijg} is the set of student and parent characteristics observed in the baseline survey.

As low- or high-ability students are randomly paired as deskmates in the MS and MSR classes. We use reduced form linear-in-mean peer effect estimation to determine the effect of the deskmate's ability on student's endline achievement. The peer's baseline average achievement is used as a proxy for student ability:

$$y_{ijg}^{endline} = \beta_0 + \beta_1 y_{ijg}^{baseline} + \theta_1 y_{-i,jg}^{baseline} + X_{ijg}\Gamma + h_{mj} + \varepsilon_{ijg}, \qquad (3)$$

where $y_{-i,jg}^{baseline}$ refers to the deskmate's baseline z-score or non-cognitive skills. Note that the random variation in a peer's ability is obtained for students in each track, and thus, we separately estimate Equation (3) for each track of students. Given that the random assignment of seats takes place within the height group in each class, a class-by-height group fixed-effect h_{mj} is controlled. In this specification, θ_1 reflects the casual effect of the deskmate's baseline score on students' endline achievement.

4.2 Main estimation results

4.2.1 *Effects of the MS and MSR programs on student academic achievement*. We start by examining the effects of the MS and MSR on the performance of students in the lower and upper tracks separately. We employ Equations (1) and (2) to estimate the treatment effects. Given that students are arranged into upper and lower tracks according to the median scores of a class, each track has 901 students. We expect that the MSR treatment benefits lower-track students because they are likely to be helped by their incentivized upper-track deskmates on study.

Panel A of Table 2 reports the estimated results for lower-track students. Column 1 shows that the effect of MS is weak and statistically insignificant (0.01 with s.e.=0.122). The MSR effect is positive although not significant (0.14 with s.e.=0.109). Column 2 shows that when individual controls are added, the precision of the estimate is improved. The estimated coefficient of MSR is 0.14σ (s.e.=0.065), suggesting that low-ability students benefit from incentivized high-ability deskmates in the MSR classes. By contrast, the effect of MS treatment on lower-track students' average z-score is small and not statistically different from 0.

Different courses may be a concern (Zimmerman, 2003; Brunello, De Paola, and Scoppa, 2010). Columns 3 to 6 present the estimated results by course, in which Chinese and mathematics z-scores are examined separately. Columns 3–4 show that the MS or MSR intervention has no effect on low-ability students' Chinese scores. Column 6 shows that, when all controls are added, the MSR intervention promotes lower-track students' mathematics scores by 0.24σ (s.e.=0.097). Such an effect is statistically different from zero at the 5% significance level. In comparison, the MS intervention has a weak effect on the mathematics scores. We also test whether the MS effect is the same as the MSR effect. A small *p*-value of the *F*-test shows that the MSR effect is significantly different from the MS effect on students' mathematics scores.

One may be concerned that our treatments may harm upper-track students' academic performance. Panel B of Table 2 reports the estimated results for upper-track students. Columns 1–2 show that the estimated effects of MS and MSR are both minor and do not significantly differ from 0. When analyzing Chinese and mathematics separately, columns 3–6 show similar results. The MS and MSR interventions have no significant impact on students' Chinese and mathematics scores.[17]

4.2.2 Effects of the MS and MSR programs on student non-cognitive skills. In this section, we examine whether the complete mixed-seating classes changes the psychological status or sociability of students by estimating the effects of MS and MSR intervention on student non-cognitive skills. This is an important issue. For example, pairing mixed-ability students provides additional opportunities for students to communicate with peers in different abilities and may induce the extrovert personality and students to become more sociable. Offering rewards in the

[17] The analysis of complex nonlinear treatment effects is shown in Table A3. We separate the entire sample students into quartiles according to the baseline score in each class and compare the treatment effects on students in each quartile. We find that the MSR is effective for mathematics and the bottom three quartiles of students.

MSR classes is similar to a tournament, which increases the level of competition among students in the upper track and induces the level of anxiety among students. However, the extent to which the interventions affect personality traits is unclear to us. This section answers these questions.

The estimated equation is specified by Equation (2). Table 3 reports the estimated results. Several features can be drawn from the table. Students in the MSR classes have high scores on the extraversion scale (column 1). Specifically, being assigned in the MSR classes promotes lower- and higher-track students' extraversion by 2.34 (s.e.=0.764) and 1.91 (s.e.=0.612) points, which are equivalent to 6% and 5% increase at the mean value of extraversion of students in the control classes. Extraversion measures one's interests and energies to the external world rather than the internal world. Extraverted people appear to be sociable, and they like to interact with others. The MS intervention has no effect on the extraversion trait.

Column 2 shows that the MSR intervention significantly increases low- and high-students' agreeableness by 2.61 (s.e.=1.248) and 2.33 (s.e.=1.176) points, which are equivalent to 8% and 7% increase at the mean value of agreeableness of students in the control classes. Agreeableness reflects the cooperative and unselfish manner of people. Moreover, agreeable individuals value getting along with others. Our result indicates that the MSR treatment may improve the quality of relations among students. The MS intervention has no effect on the agreeableness trait. Note that these two "Big Five" personality traits are related to social skills. Thus, the results may imply that the MSR intervention increases social ties among students. We investigate this issue by exploiting the students' friend networks in Section 5.2.

Openness, which measures one's openness to new aesthetic, cultural, intellectual experiences, and curious thinking, does not change among students in the complete mixed-seating classes (column 3). Although the level of competition may increase among high-ability students in the MSR classes, the level of student anxiety does not increase. The MS and MSR treatments have no effect on the neuroticism scale (column 4). Column 5 shows that being assigned in the MSR classes promotes the level of conscientiousness of lower-track students by 1.36 points (s.e.=0.781). This scale measures "organized and hardworking" skills, implying that lower-track students in the MSR classes work harder than those in the control classes. This finding reconciles our results that lower-track students in the MSR classes earn scores in the previous analysis.

5. Mechanism Analysis

Peer effects are the possible channel through which lower-track students benefit from the mixed-ability environment. In this section, we examine peer effects at the deskmate level in the treatment classes. Then, we uncover a plausible mechanism behind peer effects by sharpening our analysis on the student network.

5.1 Peer effects

In our two treatment schemes, as students are randomly assigned with new deskmates, we can estimate the peer effects of deskmates. Table 4 presents the estimate of the influence of peer baseline cognitive and non-cognitive abilities on high- and low-ability students' endline performance, and in the MS and MSR classes separately by Equation (3). We additionally control the peer's gender because Lavy, Silva, and Weinhardt (2012) indicate that the gender of peer should be a concern.

Panel A shows the estimated results for lower-track students. Columns 1–6 present results for students in the MS classes. We find that being assigned to peers with high baseline z-scores or high performance in non-cognitive skills causes no change in performance for lower-track students in the MS classes. All estimated coefficients are small and indifferent from 0. Columns 7–12 in Panel A employ the same specification to estimate peer effects for students in the MSR classes. Low-ability students' cognitive achievement benefits from incentivized stronger peers: the coefficient on peer's baseline z-score is 0.09σ with a standard error of 0.049 (column 7). Columns 8 and 9 show that being assigned to deskmates with high extraversion and agreeableness scales increases lower-track students' corresponding traits by 0.23 (s.e.=0.124) and 0.11 (s.e.=0.053) points, implying that being assigned to more sociable peers promote students' sociability. This finding is consistent with Zárate (2019).[18]

Panel B shows the estimated results for upper-track students. Columns 1–6 suggest that for upper-track students in the MS classes, peer's cognitive and non-cognitive abilities nearly have

[18] Zárate (2019) shows that having more sociable peers affects a student's non-cognitive skills.

no impact. All estimated coefficients are small and insignificant, except for neuroticism, which is marginally significant. When we move to students in the MSR class, column 9 indicates that being assigned to more agreeable lower-track deskmates improves upper-track students' agreeableness by 0.17 (s.e.=0.099) points. It implies that upper-track students are more likely to get along with others if they are assigned with more agreeable deskmates.

Overall, our results on peer effects are twofold. First, the analysis of cognitive abilities indicates that being with high achievers promotes lower-track students' test scores in the MSR classes. Second, the analysis of non-cognitive abilities indicates that being with more sociable deskmates increases both lower- and upper-track students' social skills. The two findings underlie low-ability students' score gain in the MSR classes: students may be more likely to interact with their deskmates and help lower-track students improve their scores. In the next section, we discuss the potential mechanism of peer effects by exploring student network.

5.2 Network structure

Networks of relationships play an important role in human interaction, which may generate peer effects. However, empirical evidence is lacking perhaps because of the endogeneity of friendship formation and the difficulty of collecting network information.

Homogeneous students endogenously group together (Carrell, Sacerdote, and West, 2013). Such a case may appear in the control and pre-treatment classes in our setting and affect students' achievement. For example, if the initial students in the lower track group together and behave similarly, then their achievement may be negatively affected by the interacting peers (reflection effect). The mixed-seating arrangement is supposed to lower the probability of being assigned to a deskmate with a similar ability. Therefore, the original friendship connections can be reduced due to the physical separation of students with similar abilities. Furthermore, given that our findings show that the MSR intervention promotes students' sociability, students may form new friendships with their assigned deskmates. Hence, students were surveyed on their friendships in their class to explore the degree to which our experiment alters selected friends. In two rounds of the survey, we asked students to name six best classmates (three boys and three girls) with whom they were friends.[19] For each student, his/her friends are identified.[20] By matching the name of the nominated friends of each individual and the name of his/her classmates, we construct the network for each class,[21] where a link exists between students if any student nominates another student as friends. Students communicate reasonably with their friends in daily life, so this is the network of potential interactions. We analyze the change in the friendship structure before and after the experiment. Combined with our randomized seat assignment, our identification avoids endogenous friendship formation, which is the major concern in the empirical application (Hsieh and Lee, 2016).

We analyze three effects that may be affected by the intervention. First, we test whether our treatments affect the formation of friendships between deskmates. Second, we examine whether low-ability students make more connections with upper-track students by exploring the fraction of high-ability students in a student's network. Finally, we ask whether our interventions crowd out within-track friendships or generate new between-track friendships by comparing the number of lower- and upper-track friends before and after the treatments. We provide an underlying mechanism to the deskmate-level peer effects and attempt to understand the existence of the MSR treatment effect through these analyses.

5.2.1. *Are students more likely to nominate their deskmates as friends?* As Table 4 shows deskmate-level peer effects in the MSR classes, the most interesting issue is whether the intervention promotes friendship formation between deskmates. To answer this question, we use similar specifications as Equation (2), except that we use the change in the friendship status between deskmates before and after the interventions as the dependent variable and exclude lagged outcomes. All individual controls and school-by-grade fixed effects are included. The standard errors are clustered at the class level. Column 1 of Table 5 presents the difference in treatment versus controls in terms of whether the status of a deskmate nominated as a friend

[19] In the data, 22 and 18 students nominated six friends in the first and second round of surveys, respectively. Therefore, the truncated number of friends at six is not a serious concern.

[20] If A nominates B as his/her friend, then we define B as A's friend regardless whether or not B nominates A as friend.

[21] We match the nominated friends' names to students' names using Chinese characteristics in the same class. For unmatched cases, if the pronunciation of a reported friend's name is the same as that of a student's name in the class, we also match them together because students may incorrectly write his/her friend's name. This process obtains a nearly perfect matching between reported friends' and students' names in the same class. A total of 96% reported friends can be matched with students in the same class.

changes before and after the intervention. Column 1 in Panel A shows that the MSR intervention significantly promotes the within-deskmate friendship formation for lower-track students. Specifically, pairing with incentivized high-ability students increases the probability by 8.3 percentage points (s.e.=0.048) for low-ability students to nominate their deskmates as friends. In comparison, the MS intervention has a small and statistically insignificant effect.

Column 1 in Panel B shows that MSR intervention also promotes high-ability students to form friendships with their low-ability deskmates. The estimated effect is 0.1 (s.e.=0.05), indicating that incentivized high-ability students are 10 percentage points more likely to nominate lower-ability deskmates as friends after a 20-week treatment in the MSR classes. Again, the effect of MS is small and statistically insignificant. Taken together, students in the MSR classes tend to nominate their between-track deskmates as friends than those in the control classes.

5.2.2. *Do low-ability students form more network ties with high-ability students?* We first measure the fraction of high-ability students in each student's network. Then, we regress the change of this fraction before and after the treatments on two treatment variables, individual controls, and school-by-grade fixed effects. Column 2 of Table 5 shows the estimated effect. Column 2 in Panel A shows that the MSR intervention significantly increases the inclination for lower-track students to form social ties with upper-track students. The estimated coefficient of MSR treatment is 0.06 (s.e.=0.024). The MS intervention has a relatively small effect, which is half that of the MSR intervention. The antepenultimate row of column 2 in Panel A reports the p-value of testing the equality of the estimated coefficients between MS and MSR interventions. The small p-value indicates that the MSR effect differs significantly from the MS effect.

Column 2 in Panel B reports the effects of our interventions on the fraction of high-ability friends of students in the upper track. The result shows that the estimated coefficient on the MSR treatment is 0.04 (s.e.=0.022), indicating that the MSR intervention increases the fraction of high-ability friends for high-ability students. However, the magnitude of these estimates is smaller than that in column 1. In comparison, the coefficient on the MS treatment is 0.02 (s.e.=0.023) and is statistically insignificant. Overall, Panel B of Table 5 shows that lower-track students make more friends with high-ability students in the MSR classes. Although the MS intervention also prompts social ties with high-ability students, its effect is relatively small.

Table A5 in the Appendix provides a robustness check of the results in column 2 in Panel B of Table 5. We separate the two students sharing the same desk into relatively low- and high-achieving students according to their average baseline scores.[22] This specification allows us to compare student network formation between treatment and control classes at the deskmate level. We run regressions for the relatively low- and high-achieving samples separately. The results are entirely consistent with those in column 2, Panel B of Table 5.

5.2.3. *Do connections among low-ability students decline?* A related question is whether our interventions crowd out within-track friendships, or they simply generate new between-track friendships. We examine the number of friends to verify this issue. We use the change in the number of upper-track (or lower-track) friends before and after the interventions as outcomes. We add individual controls and school-by-grade fixed effects. Columns 3–4 in Table 5 report the results. Column 3 in Panel A shows that the MS intervention reduces the number of lower-track friends. Specifically, the selected number of lower-track friends for low-ability students decreases by 0.05 in the MS classes (s.e.=0.027). The magnitude of the estimated effect is tripled in the MSR classes: MSR intervention reduces the selected number of low-track friends by 0.19 for low-ability students, which amounts to a 35% decrease at the mean value in the control classes (–0.53). Column 3 in Panel B shows that for high-ability students, the MSR intervention increases the number of their lower-track friends by 0.09 (s.e.=0.042), which offsets nearly 30% of the negative mean of control classes (–0.29).

Column 4 in Panel A shows that low-ability students prefer to make friends with upper-track students in the treatment classes. This scenario is more evident in the MSR classes than in the MS classes. Specifically, MS and MSR interventions increase the number of upper-track friends by 0.14 (s.e.=0.086) and 0.26 (s.e.=0.097). Although upper-track students generate more new friends who are in the same track as themselves, these estimates are not significant, which can be seen in column 4 in Panel B. In summary, our results show that the complete mixed-seating environment

[22] We use the word "relative" to refer to two students seated at the same desk with different baseline scores. Notably, the relatively high-achieving (low-achieving) students in the MS and MSR classes are also the upper-track (lower-track) students, but they may not be so in the control classes.

crowds out within-track friendships for lower-track students. Such an effect is more evident in the MSR classes than in the MS classes.

Figure A3 presents a picture of the friend networks between low- and high-ability students in a selected MSR class to illustrate the results in Table 5. Clear evidence can be found that students with similar abilities are likely to group together before the treatment. In comparison, low-ability students are more likely to select high-ability students as friends after the treatment. More details of this figure are presented in the Appendix. As a formal test of whether ties among lower-track students declined in the complete mixed-seating class, we follow Grund and Densley (2012) and run a dyad-level regression. We run endline friendship status between two students on a lower-track dummy and its interactions with MS and MSR. Individual characteristics surveyed in the baseline are controlled.[23] Standard errors need to be adjusted because the error terms in the regression are correlated across observations. To perform statistical inference, we permute the networks 500 times to obtain the p-value as the error terms. Table A4 reports the results. We find that students in the lower-track are more likely to connect (0.012 with p-value=0.02). Our variable of interest, the interaction term "MSR×Lower track," is significantly negative, implying that MSR helps break the connections among lower-track students (−0.05 with p-value=0.01). MS intervention has smaller effects than the MSR with the coefficient of "MS×Lower track" being 0.02 (p-value=0.01). We also find that friendship patterns are characterized by homophily (Moody, 2001): two students are more likely to be friends when they have the same age, same mother's education, are both interested in Chinese, and have computers.

In conclusion, Tables 6–7 suggest that the MSR intervention induces the separation of students in the lower track, who prefer to form friendships with newly assigned high-ability deskmates in the class. In addition, the incentivized high-ability students are more likely to select low-ability

[23] Using the friendship information, we construct the adjacency matrix, which is a 1,802×1,802 matrix. We then turn our data into the dyad level combing the adjacency matrix. Each observation of the data represents a pair of students. The estimation model takes the form: $Friend_{pq} = \beta_0 + +\beta_1 lower\ track_{pq} + \beta_2 lower\ track_{pq} \times MS_{pq} + \beta_3 lower\ track_{pq} \times MSR_{pq} + X'_{pq}\Gamma + v_{pq}$. The dependent variable, $Friend_{pq}$, is a dummy that takes a value of 1 if a pair of students (p, q) are friends (including p nominates q as friends or vice versa). $lower\ track_{pq}$ is a dummy indicating whether p and q both belong to the lower track. MS_{pq} (MSR_{pq}) is an indicator representing that p and q are both in the MS (MSR) class. X_{pq} is a set of dummies. Each dummy takes a value of 1 if p and q have the same gender, race, *hukou* registration status, parental educational level, both have a computer or a car, and both are interested in Chinese or mathematics, or takes a value of 0 otherwise. v_{pq} is the error term. The estimates are obtained by OLS.

deskmates as their friends. These findings reconcile the evidence that students are more sociable in the MSR classes. Moreover, social ties constructed between deskmates reasonably improve their interactions, and lower-track students could learn from their deskmates. Results in Tables 6–7 provide a critical piece of evidence in understanding the mechanism of the deskmate-level peer effects in the MSR classes. The peer effects and within-desk friendship formation provide important pieces of evidence in understanding the outcome of our treatment effects.

5.2.4 Does student centrality in a network matter? In a network, some nodes have more connections with other nodes in a network and are easier to obtain and propagate information. Centrality is a widely used metric to measure the influence of a node. One may ask whether more central lower-track students benefit to a greater extent from the mixed-seating environment because they are more capable of communicating with their linked high-ability friends. We use in-degree centrality to measure the influence of a node in the network. In-degree centrality of student i is a count of the number of students nominated i as a friend in the baseline survey. Relative to the out-degree centrality, in-degree centrality better reflects the node's popularity and information access.[24] Student's endline cognitive and non-cognitive performance are regressed on student in-degree centrality, two treatment variables, and their interactions. Student's baseline scores, characteristics, and school-by-grade fixed effects are also included. Table A6 reports the estimated results. We find that being nominated by more students as a friend does not lead to any significant treatment effect, both for cognitive and non-cognitive abilities. Our results suggest that our treatment effects do not depend on one's centrality in the network.

6. Further Analyses

6.1 Treatment effect by gender and gender composition between deskmates

In this section, we examine the heterogeneous treatment effect through two aspects, namely, by gender and gender composition of deskmates. We mainly concentrate on student academic achievement. First, boys and girls may have a comparative advantage in one specific course. For example, high mathematics scores are prevalent among boys, whereas high Chinese scores are usually found among girls. Second, one may worry that the positive treatment effect of MSR

[24] We also consider out-degree centrality. The estimated results are similar to those using the in-degree centrality measure.

arising from our random assignment increases the possibility of mixed gender within a desk. Given that we paired the top and bottom students as deskmates according to students' baseline test scores, boys were more likely to match with girls if girls generally performed better in the baseline score than boys or vice versa.[25] Though a prima facie fact shows that the proportion of different gender deskmates is almost the same between the treatment and control classes,[26] we need to test formally that the MSR effect is not due to the mixed-gender effect. To disentangle the two effects, we separately estimate these effects in the sub-samples divided by student gender or gender composition of deskmates. The full set of controls is used in the specification, and the standard errors are clustered at the class level. For brevity, we only report the estimated coefficients on the MSR intervention (the MS effects are small and insignificant in all regressions).

Panel A of Table 7 presents the regression results by student gender, where columns 1–4 report the estimated coefficients of MSR for lower-track boys and girls, and columns 6–8 report the corresponding results for upper-track boys and girls. Column 1 presents the sample size for boys and girls in the lower track: 546 boys and 355 girls' baseline average scores are below the median score of the class. Column 2 shows that lower-track boys assigned to the MSR group earned a grade increase of 0.18σ relative to the control group, and the effect is statistically significant at the 5% level. In comparison, girls assigned to the MSR group earned a grade increase of 0.08σ, which is not statistically different from zero. We cannot reject the hypothesis that boys have the same effect as girls in the MSR group (the *p*-value of *F*-test, which tests the equality of the MSR coefficients on boys and girls, is 0.469). Column 3 shows that, for Chinese test scores, low-ability boys or girls have small and insignificant score gains. Column 4 shows that, for mathematics scores, low-ability boys in the MSR classes acquired 0.26σ score gain (s.e.=0.087). Low-ability girls in the MSR classes gained 0.21σ in their endline mathematics scores (s.e.= 0.137). We notice that the estimated results for girls are less significant than for boys, which may be due to the smaller sample size for female students. Furthermore, the large *p*-value of the *F*-test indicates that the coefficients between boys and girls are not significantly different from each other.

[25] In our sample, the baseline average raw score of girls is 6.4% higher than that of boys.

[26] Approximately 52.9% of students sharing the same desk are of different genders in the MS and MSR classes and 50.7% of deskmates are of different genders in the control classes.

Columns 5–8 in Panel A of Table 7 report the effects of the MSR intervention on the upper-track male and female students. A total of 473 boys and 428 girls are present in the upper track (column 5). Column 6 shows that MSR intervention has a small and insignificant effect on high-ability boys' and girls' average z-score. Similarly, the MSR effect on high-ability boys' and girls' Chinese scores is small and insignificant (column 7). Although the magnitude of the MSR effect on high-ability boys' and girls' mathematics scores is larger than that in Chinese, none of these estimates is statistically different from zero. Overall, the results in Panel A of Table 7 show little evidence that the MSR effects vary significantly between genders.

Panel B of Table 7 separately estimates the MSR effects for subsamples divided by the gender composition of deskmates.[27] In the table, the same gender means that the gender of two students sharing a desk are the same, whereas different gender indicates that the gender of two students sharing a desk are different. Pairs of same-gender and different-gender deskmates total 428 (355) and 473 (546) in the lower track (upper track), as shown in columns 1 and 5. Columns 2–4 report the estimated MSR effects for the lower-track students. The measured effect of the MSR intervention on the average score is 0.17σ for lower-track students assigned with incentivized deskmates with the same gender and 0.08σ for those assigned incentivized deskmates of the opposite gender (column 2). The two estimated coefficients bound the estimate of the effect of the MSR intervention on the lower-track students (recall that the MSR effect for lower-track students is 0.14σ). Column 3 shows that the MSR treatment has an insignificant effect on the Chinese scores of lower-track students with deskmates of the same or different gender. Column 4 shows that the estimated MSR effect on mathematics scores is positive for lower-track students regardless of the gender composition of the deskmates. From the statistical language, the large *p*-values from the *F*-test indicate that these differences are not significant between the two groups. Columns 6–8 show the estimated MSR effect for the upper-track students. The positive MSR effect exists only in the mathematics scores and among students with deskmates of the same gender. The estimated coefficient is 0.19 (s.e.=0.1).

[27] Subsamples are divided according to the gender composition of deskmates, which is unlikely to be endogenous. We find that the proportion of deskmates who are of different genders in the treatment and control classes are similar (52.9% vs. 50.7%).

Taken together, the results in Panel B of Table 7 indicate that the score gain for low-ability students under the MSR intervention is not due to the mixed gender between deskmates. In addition, upper-track students can only benefit from the MSR program if they are assigned a deskmate of the same gender.

6.2 Role of teachers

In the complete mixed-seating classroom, a teacher can adjust the instructional level when the composition of students' abilities changes within the class. Teachers may vary systematically with peer composition because they would match the instruction closely to student needs (Burke and Sass, 2013). This section reports on tests of explanations related to the role of teachers. Subsection 6.2.1 argues that our treatment does not alter the teacher's effort. Subsection 6.2.2 presents evidence that the detected peer effects in the MSR classes do not vary by teacher quality, and thus, teacher quality is not the mechanism of the peer effects.

6.2.1 *Treatment effects on teacher effort*. Each student was asked to evaluate the role of the class head teacher through an evaluation questionnaire in two rounds of surveys. The evaluation questionnaire is an 18-item questionnaire with a 1–4 scale response. The questionnaire consists of three domains that measure different components, namely, the teacher's (1) learning instruction, (2) daily care for students, and (3) fairness to students. Additional details of this questionnaire are presented in the Appendix (Table A7). For our analysis, we adjust the reverse-scored item to ensure that a high value indicates a positive evaluation of the teacher. We aggregate all items and divide the summation by the standard deviation in each domain. Next, we use OLS to regress each domain on the treatment status by adding individual controls and school-by-grade dummies.

Table A8 presents the estimated results. We find that students in treatment classes provide similar evaluations on their teachers compared with those in the control classes, indicating that teacher's function may not be the driving force of the treatment effect of the MSR intervention.[28]

6.2.2 *Peer effects and teacher quality*. High-quality teachers may tailor their teaching methods to

[28] We do not conclude that the teacher has no contribution to the MSR treatment effect because class head teachers acquire some information about the treatment: They arranged students' seats and announced the rewards to the upper-track students in the MSR classes. If they adjusted their behaviors in an implicit way that was not captured by the 18 questions, then the improvement of students' score may be associated with teacher effects. However, even if this did occur, our conservative conclusion is that the incentivized mixed-seating class, along with teacher effects, helps improve lower-track students' scores and lower- and upper-track students' non-cognitive abilities.

meet students' needs in mixed-seating classes. In this sense, the estimated peer effects may include a teacher effect. Table A9 presents results to show that this is not the case. Our measure of teacher quality is a dummy, that takes a value of 1 if the class head teacher received a bachelor's degree, and takes a value of 0 otherwise. We employ a similar specification to Equation (3), except that we further add an interaction term between the deskmate's baseline z-score and the teacher quality.[29] The coefficient on this interaction term can tell us whether peer effects vary by teacher quality. We find that none of the coefficients of the "Deskmate's baseline z-score×Teacher with a bachelor degree" is statistically different from 0. Moreover, the estimated effects of peer quality are very close to those in Table 4. These results imply that our detected peer effects in the MSR classes do not vary by teacher quality. Therefore, we can conclude that teacher quality is not an important factor in the deskmate-level peer effects detected in the MSR classes.

7. Robustness

7.1 Correction of the small treatment-control variations

Our randomization is implemented at the class level with 36 classes. For statistical inference, we only have 36 observations for the identification of the treatment effect. To deal with this issue, we use a two-step method to make the small sample correction to the standard error. First, we apply the Moulton procedure, as proposed by Angrist and Pischke (2009), to calculate the Moulton standard error. Second, we follow Cameron, Miller, and Gelbach (2008) to correct the Moulton standard errors by magnifying the residuals by $\sqrt{K/(K-1)}$, where K is the number of classes. In this case, our statistical inference is based on *t*-distribution with K–2 degrees of freedom. For a convenient view of the significance of the estimated coefficients, a 95% confidence interval is reported. Table A10 reports the estimated treatment effects. Panel A shows that for students in the lower track, the effects of the MSR intervention on lower-track students' mathematics scores and extraversion scale are still significant at the 5% level (columns 3 and 4).[30] Panel B shows that for students in the upper track, the effects of MSR intervention on extraversion and agreeableness scales are significant at the 5% level (columns 4 and 5). These results are consistent with the

[29] Teacher quality is absorbed in the class-by-height group fixed effect.
[30] The results of using the corrected standard errors show that the effects of the MSR intervention on students' average endline score and agreeableness are significant at the 10% level.

findings in Tables 2 and 3, indicating that our estimates are robust.

7.2 Deskmates' collusion to cheat in the test

Finally, a plausible alternative interpretation of the test score gain for the lower-track students should be discussed: Deskmates may collude to earn the rewards. For example, upper-track students may help their low-ability deskmates cheat in the test, which would contaminate our estimated treatment effects. We argue that this would not be the case for three reasons. First, students' seats in exam venues were arranged according to their student numbers, and deskmates were less likely to sit together during the exam. Second, each student sat individually and was separated by a vacant seat on either side (i.e., each student took the exam in a separate table). Therefore, lower-track students had difficulty seeing other students' answer sheets in such an environment. Third, we recruited a teacher to proctor the exam rigorously in each of the 36 classes to prevent students from cheating.

8. Comparison with the literature

In this section, we compare our estimated effects with related literature. In a partial paring process, Li et al. (2014) find that pairing high and low achieving classmates as benchmates and offering them rewards improves low achieving students' average test scores by 0.265 standard deviations. In our setting, the estimated MSR effect on lower-track students' average z-score is 0.14 standard deviations. The difference may arise from that Li et al. (2014) incentivize both high and low achievers. In comparison, only upper-track students are rewarded in our setting. Therefore, the relatively large effect detected by Li et al. (2014) may rise from the incentive of low achievers to obtain money.

It is helpful to place the magnitude of the mixed-seating effects in a context by comparing it to the tracking effects. We compare our results reported in Table 2 with the within-school tracking effects in the literature.[31] In the early literature, Hoffer (1992) finds that students placed in the high-ability group in schools outperform otherwise similar students in schools without tracking. The size of the effect for high-ability students ranges from 0.18–0.26 standard deviations in

[31] Within-school tracking sorts students into different classrooms in accordance with student ability within a school. The complete mixed-seating intervention is a practice that mixes students of different abilities in a class. Thus, we compare the literature on within-school tracking.

tracking schools. Argys, Rees, and Brewer (1996) report that tracking boosts scores of students in high-ability classes by 8.4% (or nearly 0.25 standard deviations). In recent literature, Duflo et al. (2011) provide a clean estimate of within-school tracking in 121 elementary schools in Kenya. They show that after 18 months in tracking schools, students in the top half of the pre-assignment score distribution gain 0.19 standard deviations, whereas those in the bottom half gain 0.16 standard deviations. Card and Giuliano (2016) find that placement in a fourth-grade gifted/high-achieving class increases the scores of high-achieving minority students by 0.5 standard deviations. These findings confirm that tracking mainly benefits high-ability students.[32]

In our findings, 20 weeks of studying in the complete mixed-seating environment with incentivized high-ability students increased the endline mathematics scores of low-ability students in the classes by 0.24 standard deviations. These effects are relatively small compared with the effect in Card and Giuliano (2016), but they are close to the magnitude estimated by Hoffer (1992) and Duflo et al. (2011). Our estimates imply that the complete mixed-seating classroom with incentives benefits students, especially those who belong to the bottom of the class. Therefore, within-class mixed-seating is similar to an egalitarian social policy, which is preferable for education administrators who are concerned about equality of education. Combined with the findings in tracking literature, we may draw the conclusion that tracking is economically efficient to talents, whereas our MSR class decreases inequality among students with different abilities.

Along similar lines, we can put the magnitude of our estimates in the context of existing papers on peer effects. Here, we concentrate on the effect of peers' quality on low-ability students. Table 4 shows that for lower-track students, the estimated peer effects on test scores are 0.09 standard deviations when they are assigned with upper-track incentivized peers. Our estimates are smaller than the estimated peer effects in the literature. Burke and Sass (2013) find that low-ability students can reap benefits by 0.23–0.28 standard deviations from middle-ability classroom peers. Feld and Zölitz (2017) show that the increased fraction of middle-ability peers (relative to low-

[32] Another thread of literature argues that tracking harms low-ability students and increases inequality among students. For example, Garlick (2018) shows that tracking at the dormitory decreases the scores of low-achieving students by 0.24 standard deviations compared with students in randomly assigned dormitories. With respect to tracking and inequality, the Organization for Economic Co-operation and Development (OECD) (2003) shows that countries applying early tracking systems have high achievement variance in student's scores.

or high-ability peers) boosts the scores of low-ability students by 0.13–0.15 standard deviations in their endline GPAs. Thus, our findings imply that having an incentivized high-quality deskmates has roughly one-third to two-thirds of these effects. As for peer effects on non-cognitive abilities, Zárate (2019) shows that being assigned with more sociable peers increases students' extraversion and agreeableness scale by 0.067 and 0.066 standard deviations, respectively. Our estimated results in Table 4 show that a more extraverted or agreeable peer can improve the extraversion and agreeableness of lower-track students by 0.225 and 0.11 points, respectively, which are equivalent to 0.035 and 0.017 standard deviations, respectively. Therefore, our results are one-quarter to one-half of the effects estimated by Zárate (2019). It is natural that different results are obtained in different contexts. Peer effects may be related to group size. Burke and Sass (2013) and Feld and Zölitz (2017) examine the peer effects in the same class or squadron, which has 30–50 students. In Zárate's (2019) study, the number of students in a dormitory ranges from 3 to 6. Contrary to these studies, we analyze peer effects at the smallest level, which only involves two students.

9. Conclusion

A main challenge of education in developing countries is that students' performance is highly diverse. Peer effects play an important role in policy debates on how to improve low-ability students' performance at a low cost. In this paper, we design a promising method that groups students by pairing deskmates with different abilities to study whether students' cognitive and non-cognitive performance can be affected by those who sit next to them.

We set complete mixed-seating classes in four selected elementary schools, where low- and high-ability students were randomly paired as deskmates. To motivate high-ability students to help their low-ability deskmates on study, we offer a monetary award and a praise certificate to high-ability students if their deskmates' improved score to be in the top 10% progress among low-ability students in the class. We find that low-ability students achieve higher scores in mathematics in MSR classes than those in control classes, while high-ability students are nearly unaffected. In terms of the non-cognitive abilities, low- and high-ability students become more sociable in the MSR classes.

Beyond estimating the treatment effects, we also seek to explore the mechanism. We detect deskmate-level peer effects in the MSR classes. We find that being assigned with high achievers or more sociable peers boosts students' corresponding performances. Based on new data including granular network information of students, we also provide empirical evidence that students' network connection underlies a possible mechanism of peer effects. We find that MSR promotes friendship formation between deskmates and separates the links among lower-track students, who now form more social ties with high-ability students.

The main lesson of this study is that the design of the incentivized complete mixed-seating classes provides a low-cost method for improving low-ability students' achievement. Although after-school tutoring can help low-ability students effectively, this solution is costly. Tracking may help high-ability students, but it increases inequality. Policymakers who are concerned with improving student performance and equality of education can consider our designed method to narrow the score disparity between high- and low-ability students. Our findings on the non-cognitive skills indicate that the sociability of students can be fostered.

It is an open question of whether similar results can be obtained in different contexts. For example, ethnicity may be a concern if students of different ethnicities are reluctant to be interactive. Given that our data contain few minority students, we cannot say if any difference would occur when the number of minority students increases in the class. We believe that the incentivized complete mixed-seating classes may be likely to generate similar effects in other developing countries in South Asia, where student abilities are diverse, the non-minority constitutes a large proportion, and seats are relatively stable over time. The complete mixed-seating classes may not apply to the United States or other countries if students' seats are not fixed over time. However, our analysis of the friend networks is still helpful to gain a better understanding of the peer effects in the educational context even in these countries.

References

Akee, Randall, William Copeland, E. Jane Costello, and Emilia Simeonova, 2018. "How does household income affect child personality traits and behaviors?" *American Economic Review* 108(3): 775–827.

Alan, Sule and Seda Ertac, 2019. "Fostering patience in the classroom: Results from randomized educational intervention," *Journal of Political Economy* 126(5): 1856–1911.

Angrist, Joshua D. and Jörn-Steffen Pischke, 2009. Mostly harmless econometrics: An empiricist's companion, Princeton, NJ. Princeton University Press.

Argys, Laura M., Daniel I. Rees, and Dominic J. Brewer, 1996. "Detracking America's schools: Equity at zero cost?" *Journal of Policy analysis and Management* 15(4): 623–645.

Betts, Julian R. 2011. The economics of tracking in education, in Handbook of the Economics of Education 3: 341–381. Elsevier.

Booij, Adam S., Edwin Leuven, and Hessel Oosterbeek, 2017. "Ability peer effects in university: Evidence from a randomized experiment," *Review of Economic Studies* 84(2): 547–78.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, 2009. "Identification of peer effects through social networks," *Journal of Econometrics* 150: 41–55.

Brunello, Giorgio, Maria De Paola, and Vincenzo Scoppa, 2010. "Peer effects in higher education: Does the field of study matter?" *Economic Inquiry* 48(3): 621–634.

Bobonis, Gustavo J. and Frederico Finan, 2009. "Neighborhood peer effects in secondary school enrollment decisions," *The Review of Economics and Statistics* 91(4): 695–716.

Burke, Mary A. and Tim R. Sass, 2013. "Classroom peer effects and student achievement." *Journal of Labor Economics* 31(1): 51–82.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, 2009. "Peer effects and social networks in education," *The Review of Economic Studies* 76: 1239–1267.

Cameron, A. Colin, Douglas L. Miller, and Jonah B. Gelbach, 2008. "Bootstrapped-based improvements for inference with clustered errors," *Review of Economics and Statistics* 90(3): 414–427.

Card, David and Laura Giuliano, 2016. "Can tracking raise the test scores of high-ability minority students?" *American Economic Review* 106 (10): 2783–2816.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West, 2013. "From natural variation to optimal policy? The importance of endogenous peer group formation," *Econometrica* 81(3): 855–882.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor, 2009. "The academic achievement gap in grades 3 to 8," *The Review of Economics and Statistics* 91(2): 398–419.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, 2011. "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya," *American Economic Review* 101(5) 1739–1774.

Dobbie, Will and Roland G. Fryer Jr., 2011. "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone," *American Economic Journal: Applied Economics* 3(3):158–187.

Feld, Jan and Ulf Zölitz, 2017. "Understanding peer effects: On the nature, estimation, and channels of peer effects," *Journal of Labor Economics* 35(2): 387–428.

Fletcher, Jason M. and Stephen L. Ross, 2018. "Estimating the effects of friends on health behaviors of adolescents," *Health Economics* 27(10): 1450–1483.

Garlick, Robert. 2018. "Academic peer effects with different group assignment policies: Residential tracking versus random assignment," *American Economic Journal: Applied Economics* 10(3): 345–369.

Grund, Thomas U. and James A. Densley, 2012. "Ethnic heterogeneity in the activity and structure of a black street gang," *European Journal of Criminology* 9(3): 388–406.

Hoffer, Thomas B. 1992. "Middle school ability grouping and student achievement in science and mathematics," *Educational Evaluation and Policy Analysis* 14(3): 205–227.

Hsieh, Chih-Sheng and Lung-Fei Lee, 2016. "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics* 31(2): 301–319.

Jackson, Matthew. O., 2008. Social and economic networks, Princeton, NJ, Princeton University Press.

Jackson, Matthew. O., 2013. "Unraveling peers and peer effects: Comments on Goldsmith-Pinkham and Imbens' 'Social Networks and the Identification of peer effects'," Unpublished version.

Kautz, Tim, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans, 2014. "Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success," OECD Education Working Papers, ISSN: 1993-9019 (online), http://dx.doi.org/10.1787/19939019.

Lavy, Victor and Edith Sand, 2018. "The effect of social networks on students' academic and noncognitive behavioural outcomes: evidence from conditional random assignment of friends in school," *Economic Journal* 129(617): 439–480.

Lavy, Victor, Olmo Silva, and Felix Weinhardt, 2012. "The good, the bad, and the average: Evidence on ability peer effects in schools," *Journal of Labor Economics* 30(2): 367–414.

Li, Tao, Li Han, Linxiu Zhang, and Scott Rozelle, 2014. "Encouraging classroom peer interactions: Evidence from Chinese migrant schools," *Journal of Public Economics* 111: 29–45.

Lu, Fangwen and Michael L. Anderson, 2015. "Peer effects in microenvironments: The benefits of homogeneous classroom groups," *Journal of Labor Economics* 33(1): 91–122.

Moody, James, 2001. "Race, school integration, and friendship segregation in America," *American Journal of Sociology* 107(3): 679–716.

OECD (Organization for Economic Cooperation and Development), 2003. Learning for tomorrow's world: First results from PISA 2003. Paris: OECD.

Young, Alwyn, 2013. "Inequality, the urban-rural gap, and migration," *The Quarterly Journal of Economics* 128(4): 1727–1785.

Zárate, Román Andrés, 2019. "Social and cognitive peer effects: Experimental evidence from selective high schools in Peru," Working paper.

Zimmerman, David J., 2003. "Peer effects in academic outcomes: Evidence from a natural experiment," *Review of Economics and Statistics* 85(1): 9–23.

Zhang, Li-fang, 2011. "Hardiness and the Big Five personality traits among Chinese university students," *Learning and Individual Differences* 21(1):109–113.

Table 1. Difference in Characteristics between the Treatment and Control Groups in the Baseline Survey

	Control's mean	Diff. between MS and control		Diff. between MSR and control	
	(1)	(2)	(3)	(4)	(5)
Chinese test raw score	66.87	0.17	[2.58]	−0.54	[2.62]
Mathematics test raw score	57.66	−1.54	[3.52]	−0.28	[3.20]
Extraversion	38.13	1.26	[1.16]	0.05	[1.55]
Agreeableness	34.45	1.14	[1.36]	0.07	[2.15]
Openness	40.20	0.99	[0.77]	0.35	[0.97]
Neuroticism	38.75	0.59	[0.92]	–0.11	[1.31]
Conscientiousness	38.30	0.92	[1.08]	0.13	[1.46]
Male	0.57	−0.02	[0.03]	0.01	[0.02]
Age	8.55	0.17	[0.34]	0.10	[0.30]
Height (inch)	52.29	−0.27	[1.23]	−0.44	[0.96]
Weight (kg)	25.94	−0.25	[1.50]	−0.46	[1.23]
Non-minority (yes=1)	0.99	−0.00	[0.01]	0.01	[0.01]
Hukou registration status (urban *hukou*=1)	0.05	−0.01	[0.03]	0.01	[0.03]
Interest in Chinese (yes=1)	0.60	0.07	[0.07]	−0.06	[0.07]
Interest in Mathematics (yes=1)	0.55	0.03	[0.07]	−0.01	[0.07]
Father's education (# of years)	9.78	−0.20	[0.42]	0.05	[0.40]
Mother's education (# of years)	9.27	−0.38	[0.40]	−0.02	[0.38]
Student's household possesses a computer (yes=1)	0.28	0.02	[0.08]	0.02	[0.08]
Student's household has a car (yes=1)	0.18	−0.01	[0.06]	−0.02	[0.06]
Key variables with any missing values	0.17	−0.05	[0.03]	0.02	[0.03]
Number of individuals	574	634		594	

Notes: Column 1 presents the mean values for the control group. Columns 2 and 3 present the differences between each of the two treatment groups (MS and MSR) and the control group, respectively. Brackets contain standard errors clustered at the class level.
* Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

Table 2. Effects of the Interventions on Students' Endline Test Scores

	Average score		Chinese score		Mathematics score	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Lower-track students						
MS	0.006	0.013	−0.046	−0.020	0.014	−0.004
	[0.122]	[0.085]	[0.105]	[0.070]	[0.161]	[0.106]
MSR	0.136	0.139*	−0.006	−0.009	0.246*	0.242**
	[0.109]	[0.080]	[0.091]	[0.060]	[0.145]	[0.097]
Controls	No	Yes	No	Yes	No	Yes
p-value of effect (MS=MSR)	0.309	0.065	0.713	0.856	0.175	0.006
Observations	901	901	901	901	901	901
Panel B. Upper-track students						
MS	−0.079	−0.063	−0.052	−0.051	−0.054	−0.031
	[0.095]	[0.057]	[0.067]	[0.042]	[0.147]	[0.087]
MSR	0.022	0.045	−0.003	0.004	0.077	0.109
	[0.075]	[0.057]	[0.082]	[0.043]	[0.100]	[0.080]
Controls	No	Yes	No	Yes	No	Yes
p-value of effect (MS=MSR)	0.294	0.127	0.585	0.256	0.350	0.154
Observations	901	901	901	901	901	901

Notes: This table reports the treatment effects of the MS and MSR interventions on students' endline average scores and Chinese and mathematics test scores. The estimated equations are specified by Equation (1) for odd columns and Equation (2) for even columns. Panel A reports estimated results for lower-track students and Panel B reports estimated results for upper-track students. Each track contains 901 students. Controls include the corresponding baseline score, gender, age, height, *hukou* registration status, minority, father's education, mother's education, and whether the student's household has a computer or a car. Robust standard errors clustered at the class level are presented in brackets.

* Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

Table 3. Effects of Interventions on the "Big Five" Personality Traits

	Extraversion	Agreeableness	Openness	Neuroticism	Conscientio-usness
	(1)	(2)	(3)	(4)	(5)
Panel A. Lower-track students					
MS	0.575	1.095	−0.379	0.963	0.883
	[0.819]	[1.387]	[0.604]	[0.737]	[0.913]
MSR	2.341***	2.610**	0.805	0.550	1.358*
	[0.764]	[1.248]	[0.742]	[0.634]	[0.781]
Controls	Yes	Yes	Yes	Yes	Yes
p-value of intervention in the lower-track (MS=MSR)	0.085	0.308	0.150	0.613	0.640
Mean of the dependent variable for students in control classes	37.05	31.82	39.71	37.52	37.18
Observations	901	901	901	901	901
Panel B. Upper-track students					
MS	−0.086	0.260	−0.176	0.483	−0.129
	[0.696]	[1.162]	[0.431]	[0.642]	[0.083]
MSR	1.911***	2.327**	−0.659	−0.016	0.766
	[0.612]	[1.176]	[0.592]	[0.605]	[0.729]
Controls	Yes	Yes	Yes	Yes	Yes
p-value of intervention in the upper track (MS=MSR)	0.099	0.069	0.560	0.492	0.241
Mean of the dependent variable for students in control classes	37.82	32.77	39.76	38.18	38.31
Observations	901	901	901	901	901
p-value of MS effect (lower track=upper track)	0.219	0.268	0.654	0.247	0.168
p-value of MSR effect (lower track=upper track)	0.447	0.386	0.018	0.273	0.336

Notes: This table reports the regression estimates of treatment effects of MS and MSR interventions on student personality traits measured using the "Big Five" model. The dependent variable is each of the five domains following the "Big Five" model surveyed in the endline questionnaire. The estimated equation is Equation (2). Panel A reports the estimated results for lower-track students and Panel B reports estimated results for upper-track students. Controls include the corresponding measured outcome variable in the baseline survey, gender, age, height, *hukou* registration status, minority, father's education, mother's education, and whether the student's household has a computer or a car. Robust standard errors clustered at the class level are presented in brackets.

* Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

Table 4. Peer Effects in the Complete Mixed-Seating Classes

	Estimation for MS classes						Estimation for MSR classes					
	Average z-scores	Extra-version	Agree-ableness	Openness	Neuro-ticism	Conscientio-usness	Average z-scores	Extra-version	Agree-ableness	Openness	Neuro-ticism	Conscientio-usness
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. Lower-track students												
Deskmate's baseline performance	−0.022	0.054	−0.059	0.042	0.026	0.044	0.092*	0.225*	0.110**	0.108	−0.075	−0.022
	[0.106]	[0.044]	[0.038]	[0.057]	[0.051]	[0.049]	[0.049]	[0.124]	[0.053]	[0.061]	[0.101]	[0.086]
Class-by-height group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	317	317	317	317	317	317	297	297	297	297	297	297
Panel B. Upper-track students												
Deskmate's baseline performance	−0.032	0.054	0.084	0.055	0.095*	0.074	0.033	0.152	0.168*	−0.089	0.058	−0.021
	[0.034]	[0.044]	[0.054]	[0.078]	[0.052]	[0.092]	[0.039]	[0.112]	[0.099]	[0.067]	[0.059]	[0.077]
Class-by-height group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	317	317	317	317	317	317	297	297	297	297	297	297

Notes: This table reports the regression estimates of the effect of peer quality by examining the effect of peer baseline cognitive and non-cognitive abilities on student's endline cognitive and non-cognitive abilities. A peer is defined as one's deskmate. The outcome variables are the student's endline average scores (columns 1 and 7) or "Big Five" personality traits (columns 2–6 and 8–12). Regressions are run for upper- and lower-track students in the MS and MSR classes separately. The estimated equation is Equation (3). Controls include gender, age, height, *hukou* registration status, minority, father's education, mother's education, and whether the student's household has a computer or a car. Robust standard errors clustered at the class level are presented in brackets.

* Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

Table 5. Effects of the Interventions on Student Friend Networks

	Whether a deskmate is nominated as a friend (1)	Fraction of upper-track students in one's network (2)	Number of lower-track friends (3)	Number of upper-track friends (4)
Panel A. Lower-track students				
MS	0.041	0.033*	−0.051*	0.141
	[0.039]	[0.019]	[0.027]	[0.086]
MSR	0.083*	0.064***	−0.188***	0.255***
	[0.048]	[0.024]	[0.036]	[0.097]
Controls	Yes	Yes	Yes	Yes
p-value of the effect (MS=MSR)	0.263	0.091	0.001	0.152
Mean of the dependent variable for students in control classes	–0.105	–0.063	−0.532	0.331
Observations	901	901	901	901
Panel B. Upper-track students				
MS	0.047	0.021	−0.053	0.133
	[0.047]	[0.023]	[0.037]	[0.111]
MSR	0.096*	0.038*	0.085**	0.142
	[0.050]	[0.022]	[0.042]	[0.097]
Controls	Yes	Yes	Yes	Yes
p-value of the effect (MS=MSR)	0.168	0.432	0.001	0.93
Mean of the dependent variable for students in control classes	–0.063	–0.040	−0.290	0.355
Observations	901	901	901	901

Notes: This table reports the regression estimates of treatment effects of MS and MSR interventions on student friend networks. The dependent variables are the first-difference of whether a deskmate is nominated as a friend before and after the intervention (Column 1), the first-difference of the fraction of upper-track students in one’s friend network (Column 2), the number of lower-track friends (column 3), and the number of upper-track friends (column 4). Panel A reports estimated results for lower-track students and Panel B reports estimated results for upper-track students. All specifications include school-by-grade fixed effects. Controls include gender, age, height, *hukou* registration status, minority, father’s education, mother’s education, and whether the student’s household has a computer or a car. Robust standard errors clustered at the class level are presented in brackets.

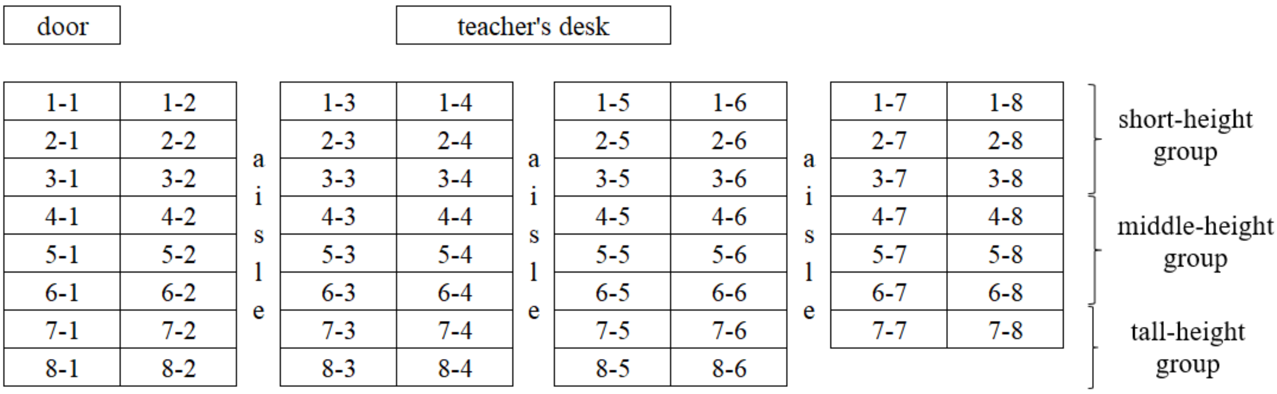
* Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

Table 6. MSR Effect by Gender and Gender Composition between Deskmates

	Obser-vations	Average score	Chinese	Mathematics	Obser-vations	Average score	Chinese	Mathematics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. By gender								
	Lower-track students				Upper-track students			
Boys	546	0.176**	0.033	0.256***	473	0.070	0.066	0.124
		[0.073]	[0.061]	[0.087]		[0.063]	[0.042]	[0.092]
Girls	355	0.079	−0.061	0.206	428	0.018	−0.038	0.080
		[0.112]	[0.092]	[0.137]		[0.056]	[0.051]	[0.079]
p-value (boys=girls)		0.469	0.328	0.748		0.521	0.263	0.642
Panel B. By gender composition within a desk								
	Lower-track students				Upper-track students			
Same gender	428	0.174*	−0.035	0.252**	355	0.104	−0.011	0.194**
		[0.099]	[0.081]	[0.112]		[0.082]	[0.069]	[0.100]
Different gender	473	0.079	−0.017	0.202**	546	0.012	0.027	0.033
		[0.078]	[0.057]	[0.104]		[0.060]	[0.045]	[0.085]
p-value (same gender =different gender)		0.623	0.257	0.871		0.339	0.310	0.079

Notes: This table reports the regression estimates of treatment effects of the MSR intervention on students' endline average scores and Chinese and mathematics test scores. In Panel A, the sample is divided into subsamples by student gender and by track. In Panel B, the sample is divided into subsamples by gender composition of deskmates and by track. Controls are added in all regressions. The controls include the corresponding baseline score, gender, age, height, *hukou* registration status, minority, father's education, mother's education, and whether the student's household has a computer or a car. Robust standard errors clustered at the class level are presented in brackets.

* Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.



Notes: Each rectangle indicates a desk; "*x*-*y*" means a student in *x* row and *y* column.

Figure 1. Desk Arrangement of a Typical Classroom

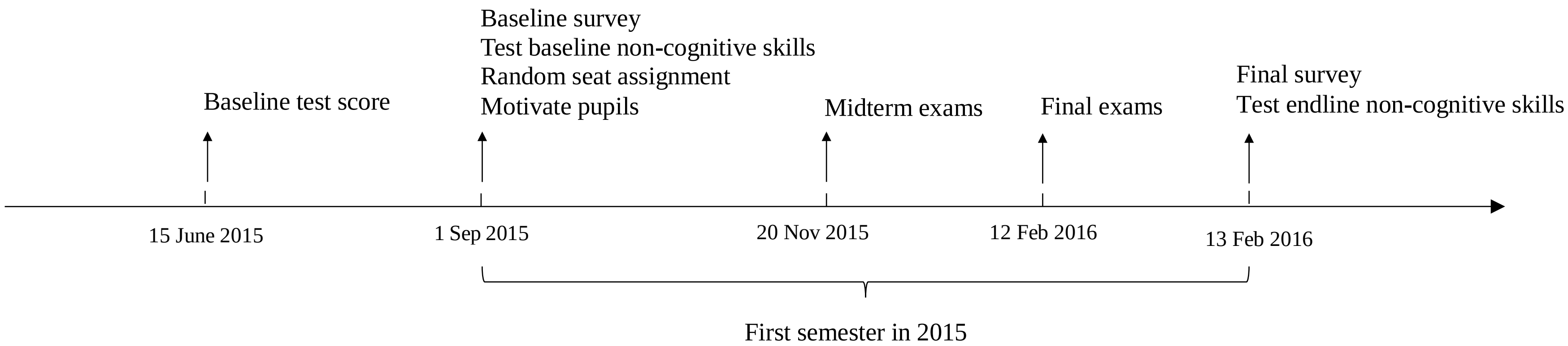


Figure 2. Timeline of Data Collection and Experiment

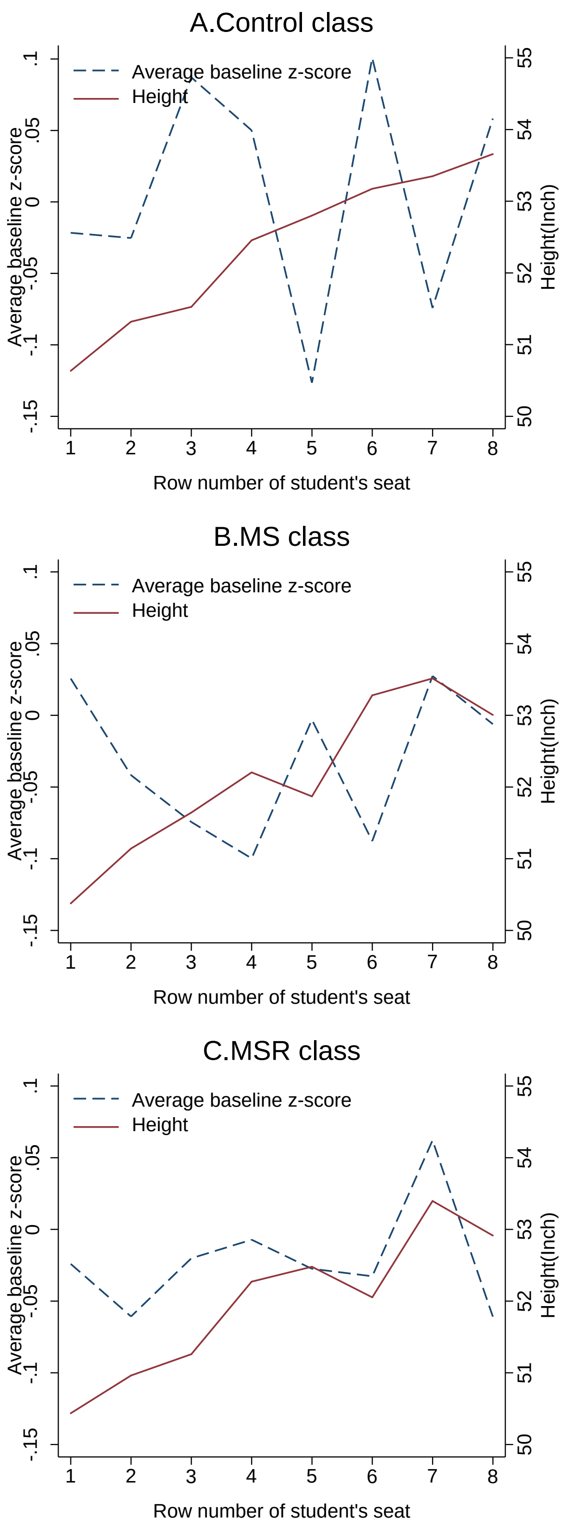


Figure 3. Average Baseline Score and Height for Students in Different Rows

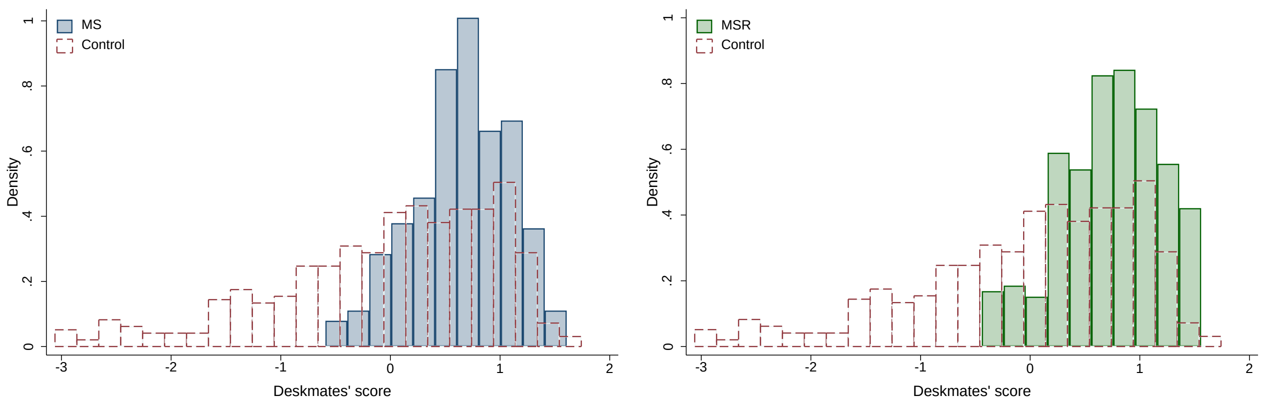


Figure 4. Distribution of Deskmate Baseline Z-Score for Low-Ability Students

Appendix

Table A1. Comparison of Student Characteristics and the "Big Five" Personality Traits between Non-randomly Assigned Students and Compliers

	Non-randomly assigned students				Compliers
	All	Non-compliers	Non-compliers' original assigned deskmates	Extended non-compliers	
	(1)	(2)	(3)	(4)	(5)
Panel A. Achievement and characteristics					
Chinese test score	67.35	69.28	65.76	66.09	66.70
	(17.99)	(16.75)	(19.74)	(19.27)	(19.09)
Mathematic test score	57.37	60.92	59.48	55.50	57.00
	(24.32)	(25.09)	(20.55)	(26.38)	(24.13)
Male	0.59	0.39	0.74	0.62	0.56
	(0.49)	(0.50)	(0.45)	(0.49)	(0.50)
Age	8.67	8.79	8.66	8.66	8.72
	(1.05)	(1.04)	(1.05)	(1.01)	(0.97)
Height (inch)	51.27	51.09	51.48	51.79	52.11
	(4.72)	(4.71)	(4.79)	(4.75)	(4.10)
Weight (kg)	25.35	25.86	26.05	25.16	25.73
	(5.62)	(5.53)	(5.75)	(5.30)	(6.13)
Non-minority (yes=1)	0.05	0.11	0.00	0.04	0.05
	(0.21)	(0.31)	(0.00)	(0.20)	(0.21)
Hukou registration status (urban *hukou*=1)	0.99	1.00	1.00	0.99	0.99
	(0.08)	(0.00)	(0.00)	(0.11)	(0.09)
Interest in Chinese (yes=1)	0.57	0.55	0.61	0.53	0.60
	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)
Interest in mathematics (yes=1)	0.52	0.55	0.61	0.46	0.56
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Father's education (# of years)	9.38	9.24	9.47	9.78	9.76
	(2.18)	(2.45)	(2.37)	(2.27)	(2.57)
Mother's education (# of years)	8.88	8.37	9.00	9.25	9.16
	(2.05)	(2.43)	(1.97)	(2.29)	(2.54)
Student's household has a computer (yes=1)	0.29	0.16	0.24	0.30	0.29
	(0.46)	(0.37)	(0.43)	(0.46)	(0.45)
Student's household has a car (yes=1)	0.13	0.13	0.08	0.16	0.17
	(0.34)	(0.34)	(0.27)	(0.37)	(0.38)
Panel B. Big Five personality traits					
Neuroticism	39.47	40.34	39.37	39.08	38.90
	(5.24)	(5.38)	(4.53)	(5.42)	(4.95)
Extraversion	39.03	38.87	39.13	39.07	38.84
	(6.05)	(6.19)	(5.23)	(6.31)	(5.62)
Openness	40.85	40.68	39.79	41.28	40.89
	(4.94)	(5.26)	(4.27)	(4.97)	(4.39)
Agreeableness	34.54	34.37	33.79	34.79	34.95
	(7.79)	(7.79)	(6.31)	(8.19)	(7.23)
Conscientiousness	39.43	39.34	38.47	40.41	39.19
	(5.95)	(6.16)	(5.39)	(6.03)	(5.84)
Number of individuals	148	37	37	74	1,650

Notes: This table reports the mean and standard deviation of the variables. The sample in column 1 includes students who were not randomly allocated. Samples in columns 2 and 3 include students who rejected the original seating arrangement and their original deskmates. The sample in column 4 includes non-compliers' new deskmates and students who eventually sat with non-compliers' original deskmates (i.e., 74 students whose seats were reassigned due to the 37 non-compliers). The sample in column 5 includes students who accepted the original seating arrangement.

Table A2. Prediction of Student Characteristics on Deskmate Pairing

	Control group		MS group		MSR group	
	Estimated coefficient	*p*-value	Estimated coefficient	*p*-value	Estimated coefficient	*p*-value
	(1)	(2)	(3)	(4)	(5)	(6)
Percentile rank	0.000	(0.93)	0.055***	(0.00)	0.061***	(0.00)
Male	−0.003	(0.12)	0.005	(0.10)	0.004	(0.19)
Age	−0.001	(0.53)	0.000	(0.91)	0.001	(0.10)
Height (inch)	−0.025	(0.19)	−0.052	(0.01)	−0.025	(0.07)
Weight (kg)	0.000	(0.22)	0.000	(0.33)	0.000	(0.49)
Non−minority (yes=1)	0.003	(0.05)	−0.001	(0.46)	−0.002	(0.31)
Hukou register status (urban *hukou*=1)	0.000	(0.04)	0.005***	(0.00)	0.000	(0.94)
Interest in Chinese (yes=1)	0.001	(0.76)	0.001	(0.44)	−0.005	(0.16)
Interest in Mathematics (yes=1)	−0.001	(0.68)	−0.002	(0.35)	0.001	(0.55)
Father's education (# of years)	0.000	(0.03)	0.000	(0.65)	0.000	(0.19)
Mother's education (# of years)	0.000	(0.76)	0.000	(0.28)	0.000	(0.34)
Student's household has a computer (yes=1)	−0.001	(0.59)	−0.001	(0.10)	−0.001	(0.71)
Student's household has a car (yes=1)	0.000	(0.97)	0.000	(0.90)	0.001	(0.55)
Class fixed effect	Yes		Yes		Yes	
Number of observations	14,403		17,093		14,541	

Notes: This table reports the estimates of the effects of the similarity of characteristics of two students on the probability of students who can be deskmates. We run OLS regression on the dyadic data, which are obtained by pairing each individual with his/her classmates to finalize a dataset. Each observation of the data represents a pair of students. The dependent variable is a dummy that has a value of 1 if the paired students are deskmates. The independent variables are the absolute difference of characteristics between a student and his/her potential deskmate. *P*-values are obtained by permuting student pairs in the control, MS, and MSR classes 500 times and reported in the parentheses.

* Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

We employ a nonlinear model to estimate a complex set of the MS and MSR effects.[33] We separate 1,802 students or the entire sample into quartiles according to the baseline score in each class (bottom, second-to-bottom, second-to-top, and top quarters). Four dummy variables, which indicate whether students belong to each quarter are generated. We then interact four dummies with MS and MSR. Therefore, eight interaction terms are created. Individual controls and school-by-grade fixed effects are added. Standard errors are clustered at the class level.

Table A3 reports the estimated results. Column 1 shows that students in the lower and upper middle reap significant score gain in their average scores from the MSR intervention. The MSR effects are 0.17σ (s.e.=0.104) and 0.1σ (s.e.=0.059), respectively. However, students in the upper middle and top quarters gain little from the MSR intervention. Column 2 shows that for Chinese language scores, the MSR intervention cannot alter scores of students in each of the quartiles. Column 3 shows that for mathematics, the magnitudes of the MSR effects among low- and middle-ability students are positive and statistically significant. For example, for the bottom three quarter students, the estimated MSR effects are 0.22σ (s.e.=0.129), 0.15σ (s.e.=0.08), and 0.21 (s.e.=0.092), respectively. For the top-quarter students, their score gain in mathematics is the smallest and insignificant. To sum up, our results indicate that the complete mixed-seating strategy accompanied by incentivized high-ability deskmates is effective for mathematics and the low- and middle-ability students. Despite the heterogeneity observed in the MSR effects between students in the bottom and top quartiles, the *p*-values of the *F*-test of whether the MSR effect in the bottom-quarter students is the same as the effect in each of the other three quarters show that they are not significantly different. Besides, the effects of interaction between the MS and student ability on all subjects are small and insignificant.

[33] The specification is as follows: $zscore_{ijg}^{endline} = \beta_0 + \beta_1 zscore_{ijg}^{baseline} + \gamma_1 MS_{ijg} \cdot Q_{ijg} + \gamma_2 MSR_{ijg} \cdot Q_{ijg} + X_{ijg}\Gamma + \eta_g + \varepsilon_{ijg}$, where Q_{ijg} is a column vector that indicate the initial quartile of the student's baseline score distribution. γ_1 and γ_2 are two row vectors to be estimated. We consider this equation in which the two treatment variables are interacted with the initial quartile of the baseline score to determine whether student test scores can vary with student's initial cognitive ability.

Table A3. Non-linear Effects of the Interventions on Test Scores in the Entire Sample

	Average score (1)	Chinese score (2)	Mathematics score (3)
(1) Bottom quarter×MS	0.047	−0.152	−0.067
	[0.106]	[0.101]	[0.128]
(2) Second-to-bottom quarter×MS	−0.017	−0.065	−0.028
	[0.081]	[0.066]	[0.102]
(3) Second-to-top quarter×MS	−0.056	0.007	−0.043
	[0.062]	[0.050]	[0.083]
(4) Top quarter×MS	−0.052	0.006	0.091
	[0.069]	[0.052]	[0.102]
(5) Bottom quarter×MSR	0.173*	−0.132	0.216*
	[0.104]	[0.085]	[0.129]
(6) Second-to-bottom quarter×MSR	0.102*	−0.051	0.152*
	[0.059]	[0.065]	[0.080]
(7) Second-to-top quarter×MSR	0.103	0.038	0.210**
	[0.068]	[0.052]	[0.092]
(8) Top quarter×MSR	−0.020	0.059	0.098
	[0.077]	[0.071]	[0.084]
Controls	Yes	Yes	Yes
p-value of total MS effect=0	0.824	0.585	0.671
p-value of total MSR effect=0	0.255	0.588	0.172
p-value of MSR effect (bottom quarter=upper middle quarter)	0.393	0.358	0.592
p-value of MSR effect (bottom quarter=lower middle quarter)	0.540	0.104	0.964
p-value of MSR effect (bottom quarter=top quarter)	0.210	0.129	0.503
Observations	1,802	1,802	1,802

Notes: This table reports the nonlinear treatment effects of the MS and MSR interventions on students' endline average scores and Chinese and mathematics test scores. A group of dummies indicating the quartile of the student's baseline score distribution is interacted with two treatment variables. We estimate the nonlinear treatment effect using the entire sample (1,802 students). Controls include the corresponding baseline scores, gender, age, height, *hukou* registration status, minority, father's education, mother's education, and whether the student's household has a computer or a car. Robust standard errors clustered at the class level are presented in brackets.

* Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

Table A4. Effects of Interventions on Friendship Ties

	Estimated coefficient	*p*-value
	(1)	(2)
Lower track	0.012**	(0.02)
MS×Lower track	–0.002***	(0.01)
MSR×Lower track	–0.005***	(0.01)
MS	0.006***	(0.01)
MSR	0.005**	(0.03)
Male	–0.001**	(0.05)
Age	0.006***	(0.01)
Height (inch)	–0.013	(0.24)
Weight (kg)	0.006	(0.12)
Non-minority (yes=1)	0.003	(0.73)
Hukou registration status (urban *hukou*=1)	0.000	(0.95)
Interest in Chinese (yes=1)	0.007***	(0.01)
Interest in mathematics (yes=1)	0.003	(0.13)
Father's education (# of years)	0.001	(0.68)
Mother's education (# of years)	0.006**	(0.04)
Student's household has a computer (yes=1)	0.007***	(0.00)
Student's household has a car (yes=1)	–0.001	(0.53)
Dyad	1,622,701	

Notes: This table reports the estimates of the effects of the similarity of characteristics of two students on the probability of students who can be friends. We run OLS regression on the dyadic data, which are obtained by pairing each individual with other students to finalize a dataset. Each observation of the data represents a pair of students. The dependent variable is a dummy, which takes a value of 1 if a pair of students are friends (including A nominates B as friends or vice versa). The independent variables are a set of dummies indicating whether pairs of students have the same characteristics (i.e. both belong to the lower track, its interactions with MS and MSR, both receive MS or MSR treatment, have the same gender, height, weight, race, *hukou* registration status, parental educational level, both have a computer or a car, and both are interested in Chinese or mathematics). *P*-values are obtained by permuting network 500 times and reported in the parentheses.

* Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

Table A5. Effects of Interventions on the Fraction of Upper-Track Students in one's Friend Network

	Relatively low-achieving students	Relatively high-achieving students
	(1)	(3)
MS	0.034*	0.023
	[0.020]	[0.024]
MSR	0.062***	0.039*
	[0.024]	[0.022]
Controls	Yes	Yes
p-value of the effect (MS=MSR)	0.212	0.430
Mean of the dependent variable for students in control classes	−0.058	−0.035
Observations	901	901

Notes: This table reports the regression estimates of treatment effects of MS and MSR interventions on the fraction of upper-track students in one's friend network. The dependent variable is the first-difference of the fraction of upper-track students in one's friend network. The sample is divided into subsamples according to the relative baseline score of two students sharing the same desk [The relatively high-performing (low-performing) students of the deskmates are the upper-track (lower-track) students in the treatment classes, but this may not be the case in the control classes]. All specifications include school-by-grade fixed effects. Controls include gender, age, height, *hukou* registration status, minority, father's education, mother's education, and whether the student's household has a computer or a car. Robust standard errors clustered at the class level are presented in brackets.

* Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

Table A6. Effects of Interventions on the Students' Endline Cognitive and Non-cognitive Performance (Considering Student Centrality)

	Average score	Chinese	Mathematics	Extraversion	Agreeableness	Openness	Neuroticism	Conscientiousness
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Lower-track students								
MS×In-degree centrality	0.021	0.018	0.011	0.088	0.110	0.043	–0.000	0.093
	[0.013]	[0.015]	[0.013]	[0.146]	[0.217]	[0.103]	[0.117]	[0.154]
MSR×In-degree centrality	–0.014	–0.022	–0.015	0.032	0.030	0.161	0.049	–0.061
	[0.016]	[0.017]	[0.022]	[0.143]	[0.216]	[0.114]	[0.144]	[0.154]
MS	–0.062	–0.089	–0.043	0.250	1.110	–0.061	0.947	0.505
	[0.107]	[0.082]	[0.129]	[1.064]	[1.426]	[0.776]	[0.953]	[1.062]
MSR	0.182	0.060	0.287**	2.251**	2.518*	0.758	0.406	1.563
	[0.109]	[0.082]	[0.134]	[0.938]	[1.343]	[0.962]	[0.895]	[1.061]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	901	901	901	901	901	901	901	901
Panel B. Upper-track students								
MS×In-degree centrality	–0.000	0.006	–0.009	0.077	0.212*	0.067	0.043	0.035
	[0.007]	[0.007]	[0.009]	[0.083]	[0.117]	[0.074]	[0.087]	[0.112]
MSR×In-degree centrality	–0.005	0.002	–0.013	–0.135	0.095	–0.053	–0.055	–0.065
	[0.009]	[0.008]	[0.012]	[0.087]	[0.091]	[0.079]	[0.094]	[0.102]
MS	–0.069	–0.084	–0.003	–1.353	–1.584	–0.542	0.285	–0.328
	[0.069]	[0.053]	[0.097]	[0.916]	[1.262]	[0.556]	[0.729]	[1.020]
MSR	0.073	–0.003	0.171*	2.491***	1.856	0.746	0.195	1.060
	[0.075]	[0.060]	[0.098]	[0.829]	[1.154]	[0.630]	[0.741]	[0.931]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	901	901	901	901	901	901	901	901

Notes: This table reports the regression estimates of treatment effects of the MS and MSR interventions on student's endline cognitive and non-cognitive performance. In-degree centrality is the number of other students who nominate an individual as a friend. Controls include the corresponding baseline performance, in-degree centrality, gender, height, age, *hukou* registration status, minority, father's education, mother's education, and whether the student's household has a computer or a car. Robust standard errors clustered at the class level are presented in brackets.

* Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

Table A7. Items from Student Evaluation of the Class Head Teacher

Item domain	Item statement	Answer scale
Learning instruction	He looks impatient when I ask him questions. (R)	1(rarely)–4(often)
	He criticizes me when I cannot solve math problems well. (R)	1(rarely)–4(often)
	He praises me when I make progress in my study.	1(rarely)–4(often)
	He patiently gives answers when I ask him questions.	1(rarely)–4(often)
	I feel he does not concern about my learning. (R)	1(rarely)–4(often)
Daily care	I feel he does not care too much about my existence. (R)	1(rarely)–4(often)
	He praises me when I do well in other things except in studying.	1(rarely)–4(often)
	He criticizes me when I do not do well. (R)	1(rarely)–4(often)
	He is very concerned about my feelings.	1(rarely)–4(often)
	He communicates with me.	1(rarely)–4(often)
	I feel that he does not pay attention to me. (R)	1(rarely)–4(often)
	He is very friendly to me.	1(rarely)–4(often)
Fairness	He shows partiality to some students. (R)	1(rarely)–4(often)
	He despises me because of my dress. (R)	1(rarely)–4(often)
	He despises me because of my appearance. (R)	1(rarely)–4(often)
	He despises me because of my family background. (R)	1(rarely)–4(often)
	He scores fairly.	1(rarely)–4(often)
	He does not distinguish between boys and girls.	1(rarely)–4(often)

Notes: The class head teacher of the respondent is referred to as “he” in the questionnaire. “(R)” indicates the reverse-scored item, which negatively assesses teachers.

Table A8. Effects of Interventions on Student Evaluations on Teacher Behaviors

	Learning instruction		Daily care		Fairness	
	Lower-track	Upper-track	Lower-track	Upper-track	Lower-track	Upper-track
	(1)	(2)	(3)	(4)	(5)	(6)
MS	−0.069	−0.078	−0.127	−0.023	0.053	0.180
	[0.159]	[0.137]	[0.138]	[0.146]	[0.174]	[0.164]
MSR	−0.013	0.012	−0.060	0.112	−0.019	−0.119
	[0.188]	[0.161]	[0.150]	[0.160]	[0.140]	[0.152]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
p-value of intervention (MS=MSR)	0.754	0.573	0.643	0.407	0.668	0.091
Mean of the dependent variable for students in control classes	0.076	0.013	0.080	−0.031	0.046	−0.077
Observations	901	901	901	901	901	901
p-value of MS intervention (lower-track=upper-track)	0.824		0.479		0.321	
p-value of MSR intervention (lower-track=upper-track)	0.896		0.164		0.371	

Notes: This table reports the regression estimates of treatment effects of MS and MSR interventions on student evaluation on teacher behavior regarding the three domains, including learning instruction, daily care, and fairness. The dependent variable is the first-difference of each of the evaluation on teacher behavior before and after the experiment. All specifications include school-by-grade fixed effects. Controls include the corresponding outcome variable measured in the baseline survey, gender, age, height, *hukou* registration status, minority, father's education, mother's education, and whether the student's household has a computer or a car. Robust standard errors clustered at the class level are presented in brackets.

* Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

Table A9. Peer Effects and Teacher Quality

	Estimation for MS classes						Estimation for MSR classes					
	Average z-scores	Extra-version	Agree-ableness	Openness	Neuro-ticism	Conscientio-usness	Average z-scores	Extra-version	Agree-ableness	Openness	Neuro-ticism	Conscientio-usness
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. Lower-track students												
Deskmate's baseline performance	–0.011	0.012	–0.033	–0.114	0.007	0.031	0.093	0.275**	0.136**	0.126**	–0.018	–0.033
	[0.162]	[0.035]	[0.045]	[0.072]	[0.058]	[0.064]	[0.057]	[0.121]	[0.049]	[0.060]	[0.081]	[0.044]
Deskmate's baseline performance×Teacher with a bachelor degree	–0.022	0.153	–0.079	0.189	0.055	0.038	–0.003	–0.243	–0.087	–0.109	–0.298	0.077
	[0.204]	[0.185]	[0.088]	[0.111]	[0.155]	[0.136]	[0.135]	[0.189]	[0.076]	[0.101]	[0.185]	[0.064]
Class-by-height group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	317	317	317	317	317	317	297	297	297	297	297	297
Panel B. Upper-track students												
Deskmate's baseline performance	–0.035	0.049	0.094	0.077	0.075	0.104	0.047	0.083	0.177*	–0.073	0.045	–0.030
	[0.037]	[0.047]	[0.064]	[0.040]	[0.049]	[0.069]	[0.042]	[0.085]	[0.101]	[0.065]	[0.064]	[0.081]
Deskmate's baseline performance×Teacher with a bachelor degree	0.007	0.018	–0.045	–0.091	0.067	–0.053	–0.101	0.196	–0.063	–0.073	0.017	0.060
	[0.078]	[0.162]	[0.134]	[0.093]	[0.105]	[0.123]	[0.109]	[0.172]	[0.129]	[0.164]	[0.053]	[0.115]
Class-by-height group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	317	317	317	317	317	317	297	297	297	297	297	297

Notes: This table reports estimated results to test whether the deskmate-level peer effects vary by teacher quality. A peer is defined as one's deskmate. The outcome variables are the student's endline average scores (columns 1 and 7) or Big Five personality traits (columns 2–6 and 8–12). The estimated equation is similar to Equation (3) except that we include the interaction between deskmate's cognitive/non-cognitive abilities and teacher quality. Teacher quality is measured by a dummy indicating whether the class head teacher has a bachelor's degree. Controls include teacher quality, gender, age, height, *hukou* registration status, minority, father's education, mother's education, and whether the student's household has a computer or a car. Robust standard errors clustered at the class level are presented in brackets.

* Significant at the 10% level. **Significant at the 5% level. ***Significant at the 1% level.

Table A10. Effects of Interventions on the Students' Endline Performance (Corrected Standard Error)

	Average score	Chinese score	Mathematics score	Extraversion	Agreeableness	Openness	Neuroticism	Conscientiousness
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Lower-track student								
MS	0.013	−0.020	−0.004	0.575	1.095	−0.379	0.963	0.883
	(−0.162, 0.188)	(−0.162, 0.126)	(−0.221, 0.407)	(−1.129, 2.234)	(−1.216, 4.266)	(−1.449, 1.487)	(−0.475, 2.445)	(−0.942, 2.635)
MSR	0.139	−0.009	0.242	2.341	2.610	0.805	0.550	1.358
	(−0.036, 0.294)	(−0.183, 0.065)	(0.012, 0.494)	(0.601, 4.049)	(−0.141, 5.494)	(−0.365, 2.646)	(−0.967, 2.037)	(−0.518, 3.152)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	901	901	901	901	901	901	901	901
Panel B. Upper-track student								
MS	−0.063	−0.051	−0.031	−0.086	0.260	−0.176	0.483	−0.129
	(−0.178, 0.056)	(−0.137, 0.037)	(−0.208, 0.236)	(−2.412, 0.575)	(−2.943, 1.591)	(−1.089, 0.683)	(−0.803, 1.871)	(−1.732, 1.598)
MSR	0.045	0.004	0.109	1.911	2.327	−0.659	−0.016	0.766
	(−0.066, 0.168)	(−0.081, 0.097)	(−0.036, 0.333)	(0.555, 3.578)	(0.074, 4.698)	(−0.312, 1.489)	(−1.266, 1.410)	(−0.746, 2.653)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	901	901	901	901	901	901	901	901

Notes: This reports the treatment effects of the MS and MSR interventions on students' endline average test scores and non-cognitive abilities using the corrected standard errors. We obtain Moulton standard error, which is then magnified by $\sqrt{K(K-1)}$, where K is the number of classes. Panel A reports the estimated results for lower-track students and Panel B reports the estimated results for upper-track students. The estimated equation is specified by Equation (2). The 95% confidence intervals are reported in parentheses. Controls include the baseline score of the corresponding outcome variable, gender, age, height, *hukou* registration status, minority, father's education, mother's education, and whether the student's household has a computer or a car.

Figure A1. Typical Desk in Chinese Elementary Schools

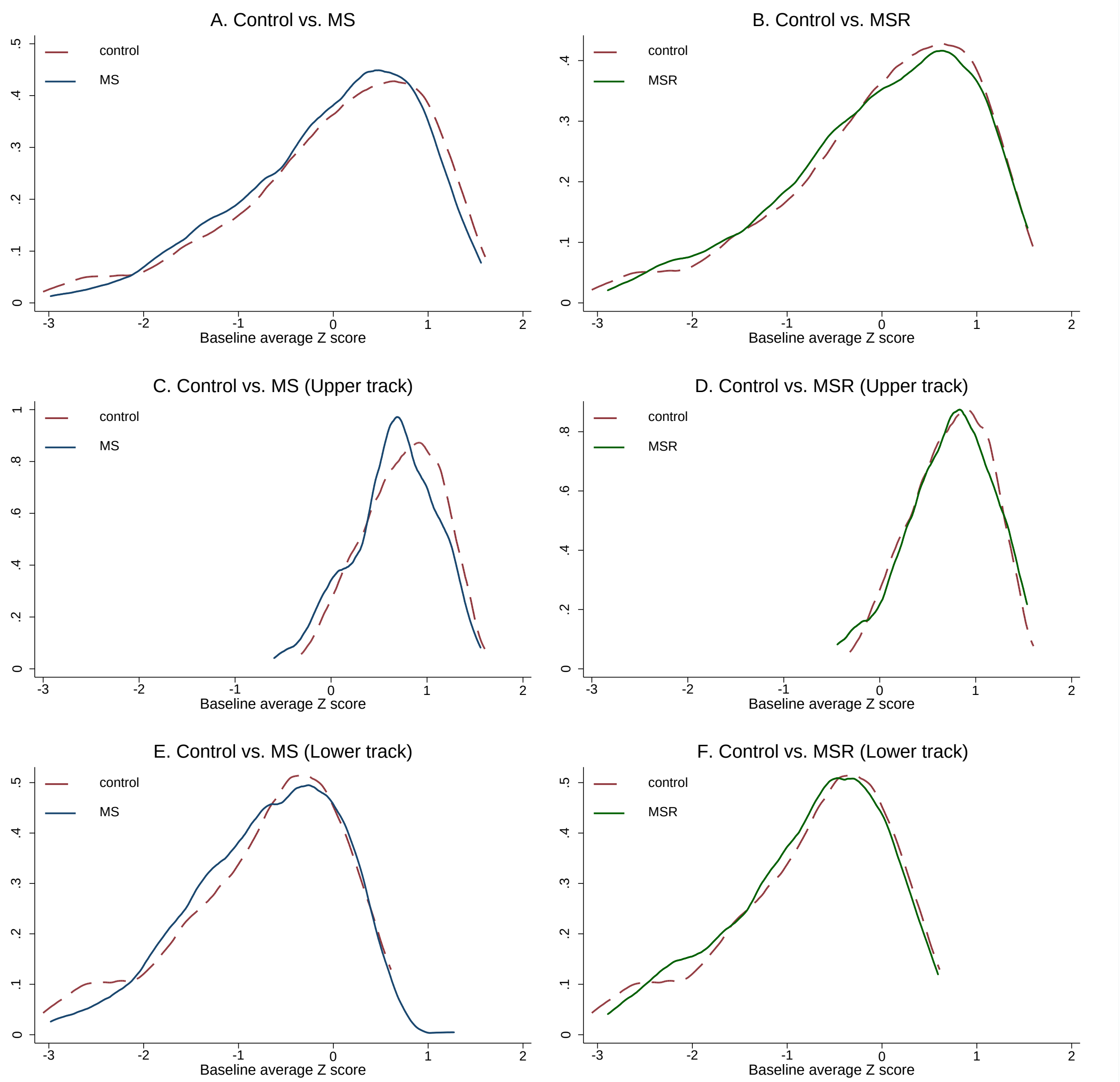


Figure A2. Density Distributions of the Baseline Test Scores

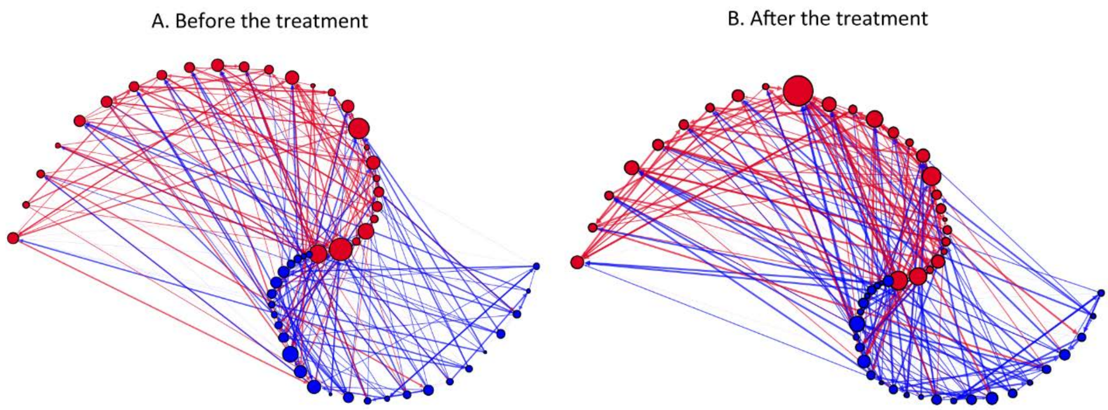


Figure A3. Structure of Student Networks in a Selected Class with the MSR Intervention

Notes: This figure presents a picture of the friend networks between low- and high-ability students in a selected MSR class. The figure on the left shows the pattern before the treatment, while the figure on the right shows the pattern after the treatment. A red dot indicates an upper-track student, whereas a blue dot represents a lower-track student. The size of the dot means the number of other students who nominate this student as a friend. Dots closer to the center indicate that the student's baseline achievement is closer to the class median baseline score. Clear evidence that can be found in the left figure is that students with similar abilities are likely to group together: the red lines dominate the top semi-cycle, indicating that upper-track students mostly select upper-track students as friends; similarly, the blue lines dominate the bottom semi-cycle, indicating that lower-track students mostly select lower-track students. In comparison, the right figure shows that blue lines in the top semi-cycle (red area) increase considerably, implying that more low-ability students select high-ability students as friends after the treatment.