American Economic Journal: Economic Policy 2023, 15(1): 510–542
https://doi.org/10.1257/pol.20200563

Student Performance, Peer Effects, and Friend Networks: Evidence from a Randomized Peer Intervention†

By Jia Wu, Junsen Zhang, and Chunchao Wang*

We estimate the effects of an educational peer intervention in which previously high- and low-achieving students are randomly paired as deskmates in elementary schools in China. Our treatment boosts the mathematics scores of the low-achieving students. Moreover, the treatment enhances the extraversion and agreeableness of high- and low-achieving students, thereby providing evidence that peers can influence personality traits. The positive treatment effects can be attributed to the deskmate-level peer effects. We document friend network structure changes and present additional evidence on how friendship ties help us better understand the mechanisms behind peer effects. (*JEL* I21, I26, J24, O15, P36, Z13)

Peer effects in education are widely heralded, as they play an important role in the improvement of students' performance in a favorable and cost-effective way. It is not surprising that economists show great interest in the economics of education, with burgeoning research revealing the link between peers' quality and child academic and long-run market consequences.[1] However, there is virtually no research examining whether children's personality traits can be affected by their peers. To fill this gap in the literature, we leverage a randomized peer intervention in elementary schools and analyze to what extent an individual's personality traits are affected by the personality traits of his/her nearby peers.

Moreover, the literature rarely presents evidence on the mechanisms behind the observed peer effects. This study peers into this black box and identifies possible mechanisms. Jackson (2008) conjectures that networks of relationships play a central

*Wu: School of Business, Shandong University; Weihai Institute for Interdisciplinary Research (email: jwu@sdu.edu.cn); Zhang: School of Economics, Zhejiang University; Department of Economics, The Chinese University of Hong Kong (email: jszhang@cuhk.edu.hk); Wang: Department of Economics, Jinan University; Jinan University–University of Birmingham Joint Institute (email: twangcc@jnu.edu.cn). The experiment in this paper is registered in the AEA RCT Registry (RCT ID: AEARCTR-0006127). C. Kirabo Jackson was coeditor for this article. We are indebted to the three anonymous referees for their valuable comments and suggestions. The authors also thank many colleagues and friends for their helpful comments. We acknowledge the funding support provided by the National Natural Science Foundation of China (Grant Nos. 72073051 and 72034006), the National Social Science Fund of China (Grant Nos. 20&ZD076 and 18ZDA081), the Natural Science Foundation of Guangdong Province (Grant Nos. 2020A1515010421 and 2021A1515012304), and the Fundamental Research Funds for the Central Universities (Grant No. 19JNKY06).

† Go to https://doi.org/10.1257/pol.20200563 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] See Lu and Anderson (2015); Feld and Zölitz (2017); Duflo, Dupas, and Kremer (2011); Burke and Sass (2013); Sacerdote (2011); Garlick (2018); Booij, Leuven, and Oosterbeek (2017); Golsteyn, Non, and Zölitz (2021); and Carrell, Hoekstra, and Kuka (2018).

role in human interaction, which helps develop peer effects. Carrell, Sacerdote, and West (2013) show that students with similar abilities are likely to sort into study (friendship) groups, thereby explaining the effects of homogenous peers. Apart from these studies, empirical studies that link networks and peer effects are lacking. An important reason is that there are formidable challenges in demonstrating this issue using observational data owing to endogenous peer composition and network formation.[2] However, this issue is important not only by itself but also because alternative mechanisms have different implications for economic efficiency and policy design in numerous contexts, in both developed and developing countries. To complement the literature, we use a randomized peer intervention and collect detailed data on students' friend networks. Hence, we are well positioned to explore an important mechanism through which peer effects work by exploring students' friend networks.

In China, the fixed-seat system in elementary schools, which requires students to sit beside the same deskmates throughout the semester, provides a good opportunity to explore the aforementioned issue. We build on an experiment originally proposed by Li, Han, Zhang, and Rozelle (2014) to design a randomized peer intervention in the mixed-seating environment with two distinct schemes to study peer intervention on student performance.[3] In the first scheme, low- and high-achieving 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 MS classes are randomly selected, and a tournament-type incentive is provided for upper-track students. To motivate high-achieving students to help low-achieving students, upper-track students receive a monetary award and a certificate of commendation if their deskmates' improved scores rank in the top 10 percent among lower-track students in class in the midterm or final examinations. We refer to these classes as "mixed-seating with reward" (MSR).[4] In the control classes, students' seats are randomly assigned. In their study, Li et al. (2014) find that pairing students and providing incentives can help low-achieving students improve their scores. The current paper attempts to reassess the effects of this successful peer intervention. Our study offers complementary evidence of the impact of the mixed-seating environment on student academic performance, but it breaks new ground by analyzing personality traits and testing the mechanisms or channels through which peer effects may be effective in improving students' performance when taking into account the role of incentives in forming their friend networks.

Furthermore, given the important role of personality traits in human capital, it is crucial to gain a better understanding of the formation of these skills. Although considerable progress has been made in documenting the determinants of personality traits or noncognitive skills (Kautz et al. 2014; Akee, Copeland, Costello, and

[2] Moreover, collecting network information is costly and time consuming.

[3] Section IA and especially footnote 7 will introduce the differences between the original design in Li et al. (2014) and ours.

[4] In this study, "lower track" and "low achieving" are used interchangeably. Similarly, "upper track" and "high achieving" are used interchangeably.

Simeonova 2018), to the best of our knowledge, no study estimates the causal impacts of peer intervention and peers on personality traits in a randomized environment. By randomly assigning a neighboring student with an individual, this study provides fresh evidence on these issues.

Our empirical analyses and findings can be summarized in four major aspects. First, after the treatment, students in the MS class without external incentives do not perform differently from those in the control class. However, when rewards are offered to high-achieving students as incentives to help their low-achieving deskmates in the MSR classes, the mathematics scores of low-achieving students improve by 0.24 standard deviations in the final examination. High-achieving students are nearly unaffected in the MS and MSR classes.

Second, we estimate the effects of MS and MSR interventions on students' "big five" personality traits. Being assigned to the MSR classes improves lower- and upper-track students' extraversion by 6 percent and 5 percent, respectively, and their agreeableness by 8 percent and 7 percent, respectively. Given that extraversion and agreeableness are related to social skills, these findings indicate that the MSR intervention leads students to be more sociable. Conscientiousness level, which relates to "organized and hardworking," is also promoted for lower-track students in the MSR classes.

Third, we show evidence of peer effects at the deskmate level in the MSR classes, thus complementing the scarce literature on micro-level peer effects (Lu and Anderson 2015).[5] Being assigned with a high-performing deskmate increases the lower-track students' test scores by 0.09 standard deviations. Moreover, we also show peer effects on personality traits. In particular, sitting beside an extraverted and agreeable deskmate increases a student's extraversion and agreeableness by 0.23 and 0.11 points, respectively. However, no peer effects are observed in the MS classes. These results imply that the development of students' personality traits can be molded by the personality traits of their peers.

Finally, we observe a higher probability of students regarding their deskmates as friends in the MSR classes than in the control classes. By matching the names of the nominated friends for each individual and the names 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 nominating each other as friends by 8.4–9.6 percentage points. Moreover, the MSR treatment induces lower-track students to select fewer low-achieving friends, and they nominate more high-achieving friends instead. In parallel, upper-track students in the MSR classes select more low-achieving friends. Overall, these findings present a plausible mechanism of the deskmate-level peer effects: the incentivized mixed-seating arrangement promotes between-track friendship formation. As friends, low-achieving students can interact easily with their high-achieving deskmates.

[5] The study of Lu and Anderson (2015) appears to be the only work on the relationship between the gender composition of neighboring-seating students and students' test scores. They randomly assign students' seats for the entire class and find that being surrounded by five female students rather than five male students increases female students' test scores.

We present several tests to exclude the alternative explanations of the treatment effects as well as peer effects. (i) To show 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. (ii) 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-achieving students in the MSR classes. (iii) We also exclude the possibility that our results are driven by deskmates' collusion to cheat in the exam.

We make two major contributions to the literature. First, we contribute to the burgeoning literature that focuses attention on the personality traits of children (Kautz et al. 2014; Akee et al. 2018). Early childhood appears to be the most malleable period and is formative for an individual's personality traits in adulthood. We present the first study to examine the causal impact of the (incentivized) mixed-seating environment on students' "big five" personality traits. We make use of peer intervention and compare the personality traits of individuals between the treatment and control classes. Moreover, leveraging on the random seating assignment, we also investigate an important but barely discussed issue: is the personality trait of a student affected by the personality traits of peers who sit next to him/her?

The second novelty of this study is that we provide empirical evidence on the shrouded mechanism behind the peer effects by exploring students' networks. The role of networks in peer effects has been documented in several theoretical frameworks. 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 the identification of the endogenous and exogenous peer effects under a network interaction. Owing to the challenges of endogenous peer composition and network formation, only few studies have linked peer effects and social relationships in empirical research. To our knowledge, one work of Carrell et al. (2013) is the only study that focuses on this issue. They show that students are likely to segregate themselves into separate groups according to abilities (homophily) and are reluctant to interact with peers of different abilities, thereby resulting in decreased benefits in squadrons that mix high- and low-ability students. Homophily that students with similar academic performance like to regard each other as friends is also detected in our data. However, students tend to choose their deskmates who are of different abilities as friends under the treatment. That is, homophily can be remolded in the mixed-seating environment if a proper incentive is provided.

The remainder of this paper is structured as follows. Section I describes our experiment design. Section II presents the model specifications and the main results. Section III examines the possible mechanisms of the treatment effect and peer effects. Section IV discusses several related issues. Section V compares our findings with the related literature, and the last section concludes the paper.

I. Experiment

A. *Study Design*

Our experiment closely followed Li et al. (2014) and was implemented in four elementary schools in Longhui County, Hunan Province, China.[6,7] 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, 20 weeks). The four schools were not selected randomly.[8] We required each selected school to have at least three classes in each grade from grades 3–5. For those with more than three classes, we randomly selected three classes from each grade in each school as our target classes. A total of 36 classes were selected.[9]

At the beginning of the semester, we selected 2 classes randomly in each school-by-grade cell as the treated classes (24 treated classes). The seats of students in each of the selected 24 treated classes were reassigned to form the 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 in Chinese and mathematics on 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 would be assigned to the upper track. Otherwise, he/she would 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 percent, middle 40 percent, and bottom 30 percent 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.[10] 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

[6] More information on Longhui can be found in the online Appendix.

[7] Li et al. (2014) design two treatment arms. The first treatment arm is "individual incentive classes," where the top three progressive low-performing students (i.e., low-performing students who achieve the greatest increase in scores between the baseline and endline tests) are rewarded. The second treatment arm is "peer incentive classes," where the high- and low-achieving students are paired as deskmates. In addition, the top three progressive low-achieving students and their paired high-achieving deskmates are rewarded with an equivalent cash prize to study together in the second treatment arm. As aforementioned, our experiment design is built on Li et al. (2014) but differs from their design in two aspects. First, we randomly pair all high- and low-achieving students as deskmates in the classes, which is a full pairing process. By comparison, Li et al. (2014) use a partial pairing process where very high- and low-performing students are assigned as deskmates (i.e., those in the middle are not paired). Second, given that awards provided for low-performing students alone have no impact, as shown by Li et al. (2014), we remove the incentives of low-achieving students. This condition is beneficial to discern the effects of peer-to-peer interaction, because such an approach can avoid the money-seeking motivation of low-achieving students.

[8] All selected schools are commuter public schools.

[9] Class sizes range from 20 to 60 students.

[10] Figure A1 in the online Appendix shows a graph of a typical desk.

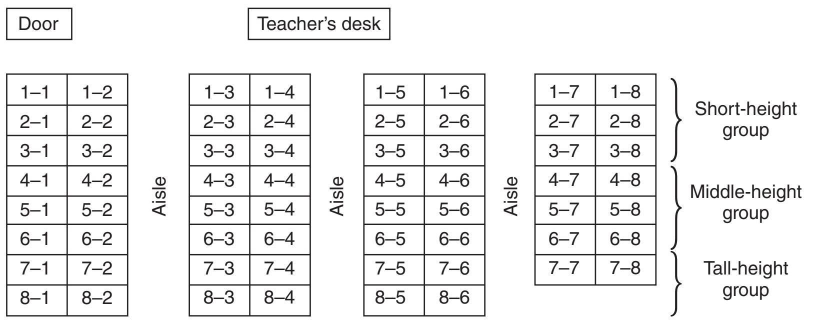

Figure 1. Desk Arrangement in a Typical Classroom

Note: Each rectangle indicates a desk; "*x*–*y*" means a student is seated in row *x* and column *y*.

classroom for each pair is from left to right and from front to back. After assigning the seats of students in the short-height group, we assigned seats for students in the middle-height group and those 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.[11] In our design, taller students mostly sat behind shorter students.[12]

Next, we selected one of the two treated classes in each school-by-grade cell to be the MSR treated class. Thus, we had 12 MSR and 12 MS classes. In the 12 MSR classes, we offered monetary awards and certificates of commendation as incentives for the 30 upper-track students in the middle and at 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-achieving deskmates achieved. Specifically, an upper-track student was offered 100 CNY if his/her deskmate's improved score gain ranked in the top 10 percent in the midterm or final exams among lower-track

[11] Several students were dissatisfied with the seating assignment and requested to change seats after the seating arrangement. In our sample, 38 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 percent of the treated sample. We allowed the class head teacher to make minor adjustments to the seats for students who requested to change seats. A dissatisfied student would be repaired with a new deskmate who was in the same track and same height group as his/her originally assigned deskmate. Table A1 in the online Appendix describes noncompliers and their original deskmates. We also report in detail noncompliers' new deskmates and students who eventually sit with noncompliers' original deskmates (we call them the "extended noncompliers"). The 38 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 38 noncompliers as normal students because they account for a very small fraction of the treated sample. Furthermore, a dissatisfied student is only allowed to exchange a new deskmate in the same track and in the height group with his/her originally assigned deskmate, and thus, this scenario does not affect the complete mixed-seating arrangement of the class. Even if we consider the noncompliers, their original deskmates, and the extended noncompliers together, the number of nonrandomly assigned students only accounts for 8 percent of the overall sample. We use two methods to deal with these nonrandomly 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.

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

students in the class. Given that the average class size is 50, a 10 percent award rate implies that 2.5 students in an MSR class can receive rewards for each exam (in total, $12 \times 2.5 = 30$ upper-track 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 US$30). This amount can be considered a financial incentive for upper-track students to help their low-achieving deskmates to learn.[13] 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 US$900), which is a low cost to motivate high-achieving students to offer their help to low-achieving 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. Students must stay in their seats during the lessons. 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.

We have 12 control classes. The protocol for the control group is simple, and no intervention is carried out. The class head teacher determines seats in the control classes; the teacher mainly considers student height—that is, short students sit in front of tall students. We will show student height does not relate to student performance in Section IB. Students can neither decide where they want to sit nor select their deskmates, and they sit with the same deskmate for the entire semester. As in MS and MSR classes, two columns of desks were rotated together in the control classes every week.

B. *Data*

Two rounds of surveys and three rounds of examinations were conducted. Figure 2 illustrates the timeline of our data collection and experiment. The baseline tests on Chinese and mathematics were conducted in June 2015. 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 county.[14] To prevent deskmates from possibly colluding to earn the rewards,[15] we required each student to sit individually and to be separated by a vacant seat on either side (i.e., each student took the exam in a separate table). Moreover, students' seats in exam venues were arranged according to their student numbers, and deskmates were

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

[14] The midterm 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.

[15] For example, upper-track students may help their low-achieving deskmates cheat on the test, which would contaminate our estimated treatment effects.

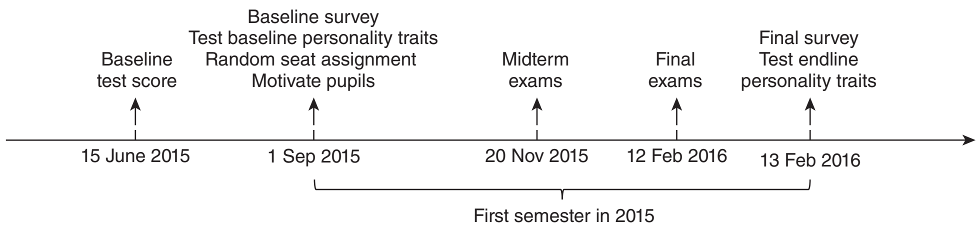

Figure 2. Timeline of Data Collection and Experiment

less likely to sit together during the exam. Lastly, we recruited additional teachers to proctor the exam rigorously for the 36 classes to prevent students from cheating.

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. A follow-up survey with similar questions was conducted in February 2016 after the final exam. In two rounds of the survey, we designed a questionnaire asking for students' personal information, such as gender, age, height, ethnicity, 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.

Chinese and mathematics scores are used to measure students' academic performance. We convert the original Chinese and mathematics scores into standardized scores (z-score). We used a well-accepted model, namely, the "big five" personality test, to measure personality traits. 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 and Li, Mak, and Wang, 2020). The "big five" personality test employed in this study is composed of 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. Students' test scores and "big five" personality traits were collected before and after the treatment.

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.[16] 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. Each track has 901 students.

[16] Among the 109 students, 38 students belong to the control classes, 37 students belong to the MS classes, and 34 students belong to the MSR classes. We also conduct a balance test for the 109 students to check whether the baseline characteristics are different across treatment status. We find no evidence that these variables are significantly different between treatment and control classes.

Table 1—Summary Statistics

	Control classes		MS classes		MSR classes	
	Mean (1)	Standard deviation (2)	Mean (3)	Standard deviation (4)	Mean (5)	Standard deviation (6)
Panel A. Baseline outcomes						
Chinese test raw score	66.87	19.66	67.04	17.83	66.33	19.56
Mathematics test raw score	57.66	24.05	56.13	23.73	57.38	24.68
Extraversion	38.13	5.50	39.39	5.56	38.18	5.57
Agreeableness	34.45	7.48	35.59	6.37	34.52	7.94
Openness	40.20	3.75	41.18	4.63	40.55	4.45
Neuroticism	38.75	5.08	39.34	4.84	38.64	4.97
Conscientiousness	38.30	5.90	39.23	5.42	38.43	5.60
Panel B. Endline outcomes						
Chinese test raw score	71.25	18.46	70.33	16.80	70.02	19.10
Mathematics test raw score	57.28	23.59	55.57	23.85	60.53	22.99
Extraversion	37.44	6.61	37.24	5.61	39.62	6.80
Agreeableness	32.30	9.34	32.84	5.90	35.16	9.71
Openness	39.74	4.35	39.59	4.29	40.53	4.39
Neuroticism	37.85	5.81	38.57	4.69	38.16	5.69
Conscientiousness	37.74	7.01	38.05	5.40	38.86	6.64
Panel C. Student characteristics and family background						
Male	0.57	0.50	0.55	0.50	0.58	0.49
Age	8.55	0.89	8.72	0.94	8.64	0.84
Height (inches)	52.29	3.99	52.02	4.54	51.85	3.90
Health status (good = 1)	0.89	0.32	0.85	0.35	0.76	0.43
Nonminority (yes = 1)	0.99	0.10	0.99	0.12	1.00	0.04
Hukou registration status (urban hukou = 1)	0.05	0.22	0.03	0.18	0.06	0.23
Father's education (# of years)	9.78	2.85	9.59	2.44	9.83	2.32
Mother's education (# of years)	9.27	2.73	8.89	2.46	9.25	2.29
Student's household possesses a computer (yes = 1)	0.28	0.45	0.29	0.46	0.30	0.46
Student's household possesses a car (yes = 1)	0.18	0.39	0.17	0.38	0.16	0.37
Number of individuals	574		634		594	

Notes: Columns 1, 3, and 5 present the means of the variables in the control and treatment classes, respectively. Columns 2, 4, and 6 present the standard deviations. Student characteristics and family background are measured in the baseline survey.

Table 1 reports the mean values by treatment status for variables measuring student test scores and personality traits, personal characteristics, and family background. Panel A shows the key baseline outcome variables—namely, average Chinese and mathematics raw test scores and "big five" personality traits. The baseline test scores and personality traits are similar across treatment status. Panel B shows the outcome variables measured after the treatment. Students' mathematics scores show a substantial 3.2 points increase for those in the MSR classes. Moreover, the extraversion and agreeableness scales slightly increase for students in the MSR classes. However, we do not find any changes for students in the MS classes. Panel C presents students' characteristics and their family backgrounds. Among the 1,802 students in the sample, slightly more than half of the students are male. Most of the students are nonminority and without urban hukou, as suggested by the fact that 1 percent is the minority, and 95 percent 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 29 percent of households have computers, and 18 percent of households have a car.

Table 2—Balance Test of Pretreatment Variables

	Lower-track students				Upper-track students			
	Diff. between MS and control	p-value	Diff. between MSR and control	p-value	Diff. between MS and control	p-value	Diff. between MSR and control	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Chinese test raw score	−1.92	0.52	−4.12	0.19	2.27	0.36	3.05	0.23
Mathematics test raw score	−5.99	0.20	−6.00	0.16	2.92	0.28	5.44	0.04
Extraversion	0.64	0.54	−0.11	0.94	1.88	0.17	0.21	0.90
Agreeableness	0.59	0.62	−0.19	0.92	1.69	0.30	0.33	0.89
Openness	0.64	0.43	0.24	0.81	1.33	0.09	0.46	0.64
Neuroticism	0.15	0.87	−0.50	0.67	1.03	0.32	0.28	0.85
Conscientiousness	0.56	0.61	−0.02	0.99	1.29	0.29	0.28	0.86
Male	0.01	0.79	0.06	0.15	−0.06	0.15	−0.05	0.32
Age	0.19	0.58	0.10	0.77	0.15	0.65	0.10	0.73
Height (inches)	−0.50	0.68	−0.53	0.61	−0.03	0.98	−0.35	0.70
Health status (good = 1)	−0.07	0.29	−0.13	0.02	−0.00	0.98	−0.13	0.04
Nonminority (yes = 1)	0.00	0.91	0.01	0.16	−0.01	0.36	0.01	0.25
Hukou registration status (urban hukou = 1)	−0.01	0.81	0.01	0.83	−0.02	0.49	0.00	0.87
Father's education (# of years)	−0.27	0.53	0.15	0.68	−0.13	0.78	−0.05	0.92
Mother's education (# of years)	−0.43	0.33	0.13	0.74	−0.36	0.38	−0.19	0.68
Student's household possesses a computer (yes = 1)	−0.01	0.86	0.00	0.99	0.05	0.58	0.04	0.66
Student's household possesses a car (yes = 1)	−0.02	0.67	−0.01	0.84	0.00	0.97	−0.03	0.60
Number of individuals	901				901			

Notes: This table presents the estimated differences in means between MS/MSR and control classes for lower- and upper-track students, which are obtained from regressing the variable of interest on the treatment dummies. Standard errors are clustered at the class level.

Table 2 provides the balance tests. We use baseline measures to assess the sample's balance across treatment status. Columns 1, 3, 5, and 7 report the estimated differences in means, which are obtained from regressing the variable of interest on the treatment dummies. Standard errors are clustered at the class level. Columns 2, 4, 6, and 8 present the estimated p-values that test whether the variable of interest is significantly different across treatment status. We perform the balance test separately for the subsamples of upper- and lower-track students. The key outcome variables do not differ significantly across treatment and control groups (rows 1–7). In addition, no significant differences by treatment status can be observed in terms of students' characteristics and family background (rows 8–17). Therefore, the MS/MSR and control groups are statistically indistinguishable in terms of the differences between the baseline characteristics and for the subsamples of upper- and lower-track students.

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 (Table 1, last row). 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

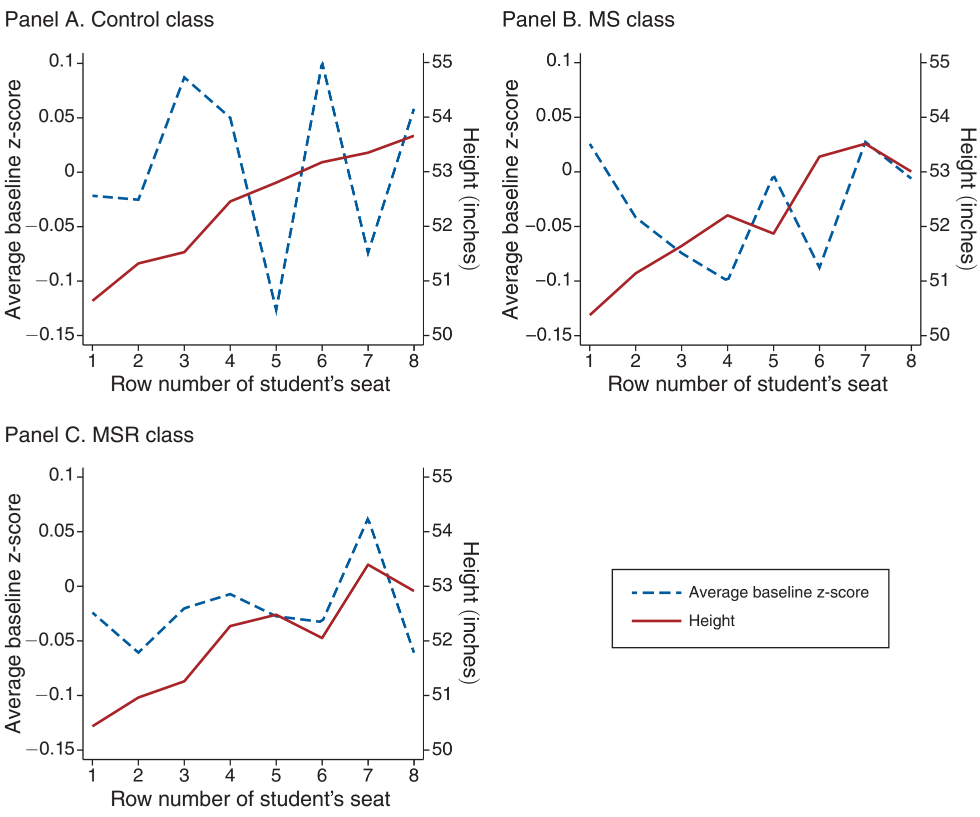

Figure 3. Average Baseline Score and Height for Students in Different Rows

may transfer to another school. In view of this concern, Table A2 in the online Appendix tests whether treatments can predict the probability of sample attrition due to the following reasons: (i) students transferred to other schools or had grade retention, (ii) missing values of key variables, and (iii) missing observation due to either case. There is no evidence of the significant difference in the probability of missing students between the treatment and control groups. The results also hold for the subsamples of upper- and lower-track students. Hence, it seems that the sample attrition is not related to the treatment.

As aforementioned, class head teachers generally assign seats on the basis of student height. To visualize this arrangement, we plot the average height of students for each row, separated according to 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. Tall students basically sit behind short 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). Students' 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 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 A3 in the online 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. By comparison, the absolute difference in percentile rank is positively associated with the probability of students becoming deskmates in the MS and MSR classes, implying that the treated classes negatively pair students with different performance as 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.

Figure 4 visualizes the results from the seating arrangement of students in the treatment and control groups. The figure shows histograms of deskmates' baseline z-scores for low-achieving 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-achieving students as deskmates.[17]

II. Empirical Model Specification and Main Estimation Results

A. *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:

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

where $y_{ijg}^{endline}$ and $y_{ijg}^{baseline}$ denote the endline and baseline standardized test scores/personality traits 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

[17] 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. These results imply that the treatment effectively creates a negative assortative pairing class. Positive relations are observed in terms of deskmates' ages and heights 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.

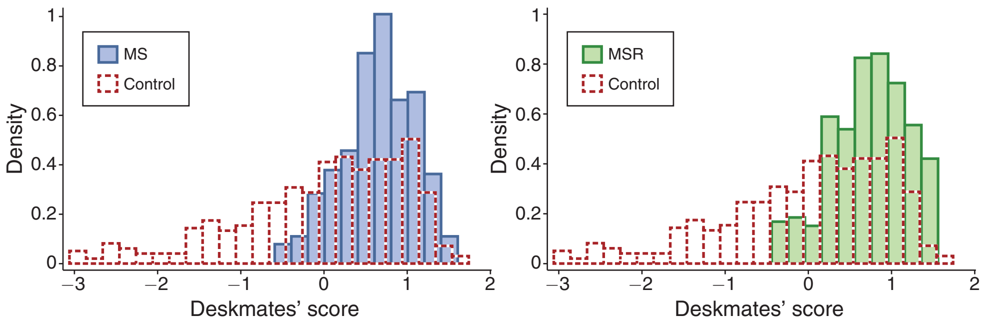

FIGURE 4. DISTRIBUTION OF DESKMATE BASELINE Z-SCORE FOR LOW-ACHIEVING STUDENTS

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 et al. (2011) and Alan and Ertac (2018), we include the lagged dependent variable in the right hand of the regression equation.

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, height, 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

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

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

As low- or high-achieving 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 a student's endline achievement. The peer's baseline average achievement is used as a proxy for student ability:

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

where $y_{-i,jg}^{baseline}$ refers to the deskmate's baseline z-score or personality traits. 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 a student's endline achievement.

B. *Main Estimation Results*

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 to learn.

Panel A of Table 3 reports the estimated results for lower-track students. Column 1 shows that the effect of MS is weak and statistically insignificant (0.01 with SE = 0.122). The MSR effect is positive but not significant (0.14 with SE = 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 (SE = 0.065), suggesting that low-achieving students benefit from their incentivized high-achieving 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-achieving students' Chinese scores. Column 6 shows that when all controls are added, the MSR intervention promotes lower-track students' mathematics scores by 0.24σ (SE = 0.097). Such an effect is statistically different from 0 at the 5 percent significance level. By comparison, the MS intervention has a weak effect on 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.

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

Effects of the MS and MSR Programs on Student Personality Traits.—In this subsection, we examine whether the mixed-seating class changes the personality traits of students. This is an important issue. For example, pairing students may provide additional opportunities for students to communicate with peers in different abilities and may induce high level of extraversion and sociability. Offering rewards in the 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 subsection answers these questions.

Table 4 reports the estimated results using equation (2). Several features can be drawn from the table. Students in the MSR classes have high scores on the

Table 3—Effects of the Interventions on Students' 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 own baseline scores, gender, age, height, health status, hukou registration status, minority status, 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 reported in brackets.

extraversion scale (column 1). Specifically, being assigned to the MSR classes promotes lower- and higher-track students' extraversion by 2.33 (SE = 0.759) and 2.07 (SE = 0.653) points, respectively, which are equivalent to 6 percent and 5 percent increases at the mean value of extraversion of students in the control classes. Extraversion measures one's interests and energies toward 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.

Column 2 shows that the MSR intervention significantly increases low- and high-achieving students' agreeableness by 2.68 (SE = 1.256) and 2.39 (SE = 1.175) points, respectively, which are equivalent to 8 percent and 7 percent increases 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. 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 IIIB.

Openness, which measures one's curious thinking and openness to new aesthetic, cultural, and intellectual experiences, does not change among students in the

Table 4—Effects of Interventions on the "Big Five" Personality Traits

	Extraversion (1)	Agreeableness (2)	Openness (3)	Neuroticism (4)	Conscientiousness (5)
Panel A. Lower-track students					
MS	0.553	1.525	0.019	0.985	0.847
	[0.832]	[1.392]	[0.626]	[0.756]	[0.914]
MSR	2.325	2.677	1.140	0.535	1.317
	[0.759]	[1.256]	[0.761]	[0.652]	[0.781]
Controls	Yes	Yes	Yes	Yes	Yes
p-value of intervention in the lower track (MS = MSR)	0.085	0.451	0.161	0.613	0.640
Mean of the dependent variable for students in control classes	37.96	33.42	40.00	38.02	37.86
Observations	901	901	901	901	901
Panel B. Upper-track students					
MS	−0.918	−0.676	−0.203	0.507	−0.067
	[0.773]	[1.168]	[0.367]	[0.650]	[0.819]
MSR	2.066	2.386	0.589	0.072	0.953
	[0.653]	[1.175]	[0.447]	[0.606]	[0.732]
Controls	Yes	Yes	Yes	Yes	Yes
p-value of intervention in the upper track (MS = MSR)	0.001	0.009	0.161	0.492	0.337
Mean of the dependent variable for students in control classes	38.21	33.45	39.89	38.40	38.59
Observations	901	901	901	901	901
p-value of MS effect (lower track = upper track)	0.010	0.007	0.694	0.247	0.197
p-value of MSR effect (lower track = upper track)	0.507	0.494	0.399	0.272	0.409

Notes: This table reports the regression estimates of treatment effects of MS and MSR interventions on students' "big five" personality traits. The dependent variable is "big five" personality traits 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 own baseline personality trait, gender, age, height, health status, hukou registration status, minority status, 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 reported in brackets.

mixed-seating classes (column 3). Although the level of competition may increase among high-achieving students in the MSR classes, the level of student anxiety does not increase. The MS and MSR treatments have no effect on the neuroticism (column 4). Column 5 shows that being assigned to the MSR classes promotes the level of conscientiousness of lower-track students by 1.32 points (SE = 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. Overall, our analyses support that personality traits can be affected by peer intervention and remain malleable in primary school.

A natural question is whether the treatment is particularly successful for a type of student. For example, the treatment effects may be more effective in promoting test scores of students at the bottom, with top-ten high-performing deskmates, or among deskmates who have large score gaps. Nonlinear treatment effects in terms of baseline personality traits may exist. Moreover, treatment effects may vary by family background.

First, we employ a nonlinear specification to estimate heterogeneous MS and MSR effects.[18] The effects of MSR intervention on test scores are more effective for the bottom three quartiles of students (Table A4 in the online Appendix). Then, we generate a dummy "Top ten students."[19] We include this dummy and its interaction with each of the two treatment variables in equation (2) to test heterogeneity in treatment effects in terms of deskmate's initial ranking. In a similar vein, we generate a dummy "Large difference between deskmates" to analyze whether treatment effects vary by initial score gaps between deskmates.[20] We find no evidence that low-achieving students who sit with the top ten students or with deskmates who have higher scores benefit more (Table A5 in the online Appendix). Second, to test whether treatment effects on personality traits vary by students' baseline personality traits, for each "big five" personality trait, we generate a dummy indicating whether a student's personality trait measured in the baseline is below the sample median. This dummy and its interaction with MS and MSR are further included in equation (2). Table A6 in the online Appendix shows that the treatment effects of MSR on extraversion and agreeableness are not sensitive to students' baseline extraversion and agreeableness. However, the MSR intervention is more effective in increasing the openness of high-openness students. Third, to test whether treatment effects on test scores vary by students' personality traits, we include the "big five" personality traits measured in the baseline and their interactions with each of the two treatment variables in equation (2). The MSR intervention increases test scores of students who are more open and less conscientious in the baseline (Table A7 in the online Appendix). This may be due to the fact that open students are more likely to seek new experiences, and therefore they feel more comfortable in the MSR classes. Students who score low on conscientiousness are less organized and less hardworking and thus perform worse. That less conscientious students benefit more from the treatment is consistent with our finding that MSR intervention promotes lower-track students' test scores. Lastly, to analyze whether treatment effects vary by students' family background, we generate two dummies, "Low-educated parents" and "High income."[21] Each dummy and its interaction with each of the two treatment variables are included in equation (2). Our analyses do not reveal any systematic heterogeneity in treatment effects on test scores and personality traits with respect to parental educational level (Table A8 in the online Appendix). However, being assigned to the MSR classes increases low-income students' conscientiousness (Table A9 in the online Appendix). Thus, the positive treatment effects shown in the last column in Table 4 are likely driven by low-income students.

[18] The specification is as follows: $zscore_{ijg}^{endline} = \beta_0 + \beta_1 zscore_{ijg}^{baseline} + \gamma_1 MS_{ijg} \cdot Q_{ijg} + \gamma_2 MSR_{ijg} \cdot Q_{ijg} + X_{ijg}\Gamma + \eta_g + \varepsilon_{ijg}$. Q_{ijg} is a column vector; each element is a dummy indicating the quartile of the student's baseline score. γ_1 and γ_2 are two row vectors to be estimated.

[19] "Top ten students" takes a value of 1 if students are assigned with deskmates who rank in the top ten in the baseline examination in their classes and a value of 0 otherwise.

[20] "Large difference between deskmates" takes a value of 1 if the difference in baseline average scores between deskmates is above the deskmate-level median difference score in the class and value of 0 otherwise.

[21] "Low-educated parents" takes a value of 1 if both parents did not attain a high school education and a value of 0 otherwise. "High income" takes a value of 1 if a student's household possesses a car and a computer and takes a value of 0 otherwise.

III. Mechanism Analysis

Peer effects are the possible channel through which lower-track students benefit from the mixed-seating 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.

A. *Peer Effects*

In our two treatment schemes, as students are randomly assigned with new deskmates, we can estimate the peer effects of deskmates. Table 5 presents the estimate of the influence of peer baseline test scores and personality traits on high- and low-achieving students' corresponding performance measured in the endline and in the MS and MSR classes separately by equation (3). We additionally control for peer's gender because Lavy, Silva, and Weinhardt (2012) indicate that it should be a concern.

Panel A presents estimated results for lower-track students in the MS classes. Being assigned to peers with high baseline z-scores or high levels of personality traits causes no change in performance for lower-track students in the MS classes. All estimated coefficients are small and indifferent from 0. Panel B shows the estimated results for upper-track students in the MS classes. Results show that peers' test scores and personality traits nearly have no impact. All estimated coefficients are small and insignificant, except for neuroticism, which is significant at the 10 percent level.

Panel C employs the same specification to estimate peer effects for lower-track students in the MSR classes. Low-achieving students' academic performance benefits from incentivized peers with high baseline test scores: the coefficient on peer's baseline z-score is 0.09σ with a standard error of 0.05 (column 1). Being assigned to deskmates with high extraversion and agreeableness scales increases lower-track students' corresponding traits by 0.23 (SE $=$ 0.124) and 0.11 (SE $=$ 0.053) points (columns 2 and 3). This finding indicates that students' personality traits can be affected by the personality traits of their peers and supports the recent research in social skill formation (Zárate 2019).[22] Such deskmate-level peer effects only exist when rewards are provided for high-achieving students.

When we move to upper-track students in the MSR class (panel D), column 3 indicates that being assigned to more agreeable lower-track deskmates improves upper-track students' agreeableness by 0.17 (SE $=$ 0.099) points. It implies that upper-track students are more likely to get along with others if they are assigned agreeable deskmates. A related question is whether peer personality traits affect students' test scores or vice versa. To test cross-skill peer effects, we regress students' endline z-scores/personality traits on peers' baseline personality traits/z-scores separately in MS and MSR classes, individual controls, and class-by-height-group fixed effect. Students' test scores are not sensitive to peers' baseline personality traits.

[22] Zárate (2019) conducts a large-scale field experiment in boarding schools in Peru and shows that having more sociable peers enhances the formation of social skills.

Table 5—Peer Effects in the Mixed-Seating Classes

	Average z-scores (1)	Extra-version (2)	Agree-ableness (3)	Openness (4)	Neuroticism (5)	Conscientiousness (6)
Panel A. Lower-track students in MS classes						
Deskmate's baseline performance	−0.029	0.054	−0.059	−0.048	0.026	0.044
	[0.106]	[0.044]	[0.038]	[0.057]	[0.051]	[0.049]
Class-by-height-group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	317	317	317	317	317	317
Panel B. Upper-track students in MS classes						
Deskmate's baseline performance	−0.032	0.054	0.084	0.055	0.095	0.074
	[0.034]	[0.044]	[0.054]	[0.078]	[0.052]	[0.092]
Class-by-height-group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	317	317	317	317	317	317
Panel C. Lower-track students in MSR classes						
Deskmate's baseline performance	0.094	0.225	0.110	0.108	−0.075	−0.022
	[0.050]	[0.124]	[0.053]	[0.061]	[0.101]	[0.086]
Class-by-height-group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	297	297	297	297	297	297
Panel D. Upper-track students in MSR classes						
Deskmate's baseline performance	0.033	0.152	0.168	−0.089	0.058	−0.021
	[0.038]	[0.112]	[0.099]	[0.067]	[0.059]	[0.077]
Class-by-height-group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	297	297	297	297	297	297

Notes: This table reports the regression estimates of the effect of peer quality by examining the effect of peers' baseline test scores and personality traits on students' corresponding performance. A peer is defined as one's deskmate. The dependent variables are the student's endline average z-scores (column 1) or "big five" personality traits (columns 2–6). Regressions are run for upper- and lower-track students in the MS and MSR classes separately. The estimated equation is equation (3). "Deskmate's baseline performance" is the corresponding outcome variable of the deskmate measured in the baseline. Controls include the own baseline performance, gender, age, height, health status, hukou registration status, minority status, 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 reported in brackets.

However, the anxiety level of upper-track students rises with the baseline scores of lower-track students in the MS classes (see Table A10 in the online Appendix).[23]

Overall, the analysis of academic performance indicates that being with high achievers promotes lower-track students' test scores in the MSR classes. In addition, the analysis of personality traits indicates that being with more sociable deskmates increases lower- and upper-track students' social skills in the MSR classes. Coupled with no detected peer effects in the MS classes, our findings imply that a certificate of commendation and a small amount of money reward provided to high-achieving students are important for them to interact with and offer help to their deskmates. These findings underlie low-achieving students' test score gain in

[23] We find no significant results of neuroticism for upper-track students in the MSR classes. One possible explanation is that deskmates tend to form friendships in the MSR classes, thereby leading them to feel less anxious with their friends.

the MSR classes. In the next subsection, we discuss the potential mechanism of peer effects by exploring student friend networks.

B. *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 et al. 2013). Such a case may appear in the control and pretreatment 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, original friendship connections can be reduced due to the physical separation of students with similar abilities. Furthermore, our previous findings show that the MSR intervention promotes students' sociability, thus implying that students may form new friendships with their assigned deskmates. Hence, students were surveyed regarding 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 friends (three boys and three girls) in class.[24] For each student, his/her friends are identified.[25] By matching the name of the nominated friends of each individual with the name of his/her classmates, we construct the network for each class,[26] where a link is formed between two students if either student nominates the other as a friend. Students communicate reasonably with their friends daily; hence, this network represents 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-achieving students make more connections with upper-track students by exploring the fraction of high-achieving 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

[24] In the data, 22 and 18 students nominated 6 friends in the first and second round of surveys, respectively. Therefore, the truncated number of friends at six is not a serious concern.

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

[26] We match the nominated friends' names to students' names using Chinese characters 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, then we match them together because a student may incorrectly spell 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 percent of reported friends can be matched with students in the same class.

underlying mechanism to the deskmate-level peer effects and attempt to understand the existence of the MSR treatment effect through these analyses.

Are Students More Likely to Nominate Their Deskmates as Friends?—As Table 5 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 6 presents the difference in treatment versus controls in terms of whether the status of a deskmate nominated as a friend 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-achieving students increases the probability by 8.3 percentage points (SE $=$ 0.048) for low-achieving students to nominate their deskmates as friends. By comparison, the MS intervention has a small and statistically insignificant effect.

Column 1 in panel B shows that MSR intervention also encourages high-achieving students to form friendships with their low-achieving deskmates. The estimated effect is 0.1 (SE $=$ 0.05), indicating that incentivized high-achieving students are 10 percentage points more likely to nominate low-achieving 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 more than those in the control classes.

Do Low-Achieving Students Form More Network Ties with High-Achieving Students?—We first measure the fraction of high-achieving 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 6 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 (SE $=$ 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-achieving friends of students in the upper track. The result shows that the estimated coefficient on the MSR treatment is 0.04 (SE $=$ 0.022), indicating that the MSR intervention increases the fraction of high-achieving friends for upper-track students. However, the magnitude of these estimates is smaller than that in column 1. By comparison, the coefficient on the MS treatment is 0.02 (SE $=$ 0.023) and is statistically insignificant. Overall, panel B of Table 6 shows that lower-track students

Table 6—Effects of the Interventions on Student Friend Networks

	Whether a deskmate is nominated as a friend	Fraction of upper-track students in one's network	Number of lower-track friends	Number of upper-track friends
	(1)	(2)	(3)	(4)
Panel A. Lower-track students				
MS	0.042	0.036	−0.051	0.141
	[0.039]	[0.019]	[0.027]	[0.086]
MSR	0.084	0.066	−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	0.039	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, health status, hukou registration status, minority status, 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 reported in brackets.

make more friends with high-achieving students in the MSR classes. Although the MS intervention prompts social ties with high-achieving students, its effect is small.

We provide a robustness check of the results in column 2 in panel B of Table 6 (see Table A11 in the online Appendix). We separate the two students sharing the same desk into relatively low- and high-achieving students according to their average baseline scores.[27] 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 6.

Do Connections among Low-Achieving Students Decline?—A related question is whether our interventions crowd out within-track friendships or simply generate new between-track friendships. We examine the number of friends to verify this issue.

[27] 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.

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 6 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-achieving students decreases by 0.05 in the MS classes (SE = 0.027). The magnitude of the estimated effect is tripled in the MSR classes: MSR intervention reduces the selected number of lower-track friends by 0.19 for low-achieving students, which amounts to a 35 panel decrease at the mean value in the control classes (-0.53).[28] Column 3 in panel B shows that for high-achieving students, the MSR intervention increases the number of their lower-track friends by 0.09 (SE = 0.042), which offsets nearly 30 percent of the negative mean of control classes (-0.29).

Column 4 in panel A shows that low-achieving 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 (SE = 0.086) and 0.26 (SE = 0.097), respectively. 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, the mixed-seating environment 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 A2 in the online Appendix presents a picture of the friend networks between low- and high-achieving students in a selected MSR class to illustrate the results in Table 6. Evidence clearly shows that students with similar abilities are likely to group together before the treatment. By comparison, low-achieving students are more likely to select high-achieving students as friends after the treatment. The online Appendix presents more details on Figure A2. As a formal test of whether ties among lower-track students declined in the 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.[29] To perform statistical inference, we permute the networks 100 times to obtain the p-value to deal with the correlation of error terms across observations. Table A12 in the online Appendix reports the results. Students in the lower track are more likely to connect with lower-track students (0.004 with p-value < 0.01). Our

[28] We compare the number of friends reported by lower-track students before and after the treatment. The average number of reported friends is 3.7 in the baseline survey, compared with 3.4 in the endline survey. Thus, the negative effect is not caused by lower-track students having more friends after the treatments.

[29] Using the friendship information, we construct the adjacency matrix, which is a $1{,}802 \times 1{,}802$ 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 $Friend_{pq} = \beta_0 + +\beta_1 lower\ track_{pq} + \beta_2 lower\ track_{pq} \times MS_{pq} + \beta_3 lower\ track_{pq} \times MSR_{pq} + X'_{pq}\Gamma + \nu_{pq}$. The dependent variable $Friend_{pq}$ is a dummy that takes a value of 1 if a pair of students (p, q) are friends (including the scenario when p nominates q as a friend 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, and parental educational level; both have a computer or a car; and both are interested in Chinese or mathematics, and it takes a value of 0 otherwise. ν_{pq} is the error term. The estimates are obtained by using OLS.

variable of interest, interaction term "MSR × Lower track," is significantly negative, implying that MSR helps break the connections among lower-track students (-0.008 with p-value $= 0.01$). Moreover, friendship patterns are characterized by homophily: two students are more likely to be friends when they have the same age, number of siblings, and mother's education and are both interested in Chinese.

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 their newly assigned high-achieving deskmates. In addition, incentivized high-achieving students are more likely to select low-achieving deskmates as their friends. These findings reconcile the evidence that students are more sociable in the MSR classes. 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.

IV. Further Analyses

A. *Treatment Effect by Gender and Gender Composition*

In this subsection, we examine the heterogeneous treatment effect through two aspects—namely, by gender and gender composition of deskmates. First, boys and girls may perform differently in the treatment classes. Second, the positive treatment effect of MSR arising from our random assignment may increase the possibility of mixed gender within a desk. Boys were more likely to match with girls if girls generally performed better in the baseline score than boys or vice versa.[30]

Table 7 presents the heterogeneity in treatment effects by student gender.[31] We generate a dummy "Boy" that takes a value of 1 if the student is a boy and a value of 0 otherwise. We interact this dummy with each of the two treatment variables. For brevity, we only report the estimated coefficients on the MSR intervention and its interaction with gender. We still find significant treatment effects of MSR on student extraversion, agreeableness, and lower-track students' mathematics scores. However, none of the estimated coefficients on "MSR × Boy" are statistically different from 0. Thus, our estimated results show no evidence that the MSR treatment effect varies between student gender.

[30] In our sample, the baseline average raw score of girls is 6.4 percent higher than that of boys.
[31] The lower track consists of 546 boys and 355 girls, and the upper track has 473 boys and 428 girls.

Table 7—Effects of MSR Intervention by Student Gender

	Average score (1)	Chinese score (2)	Mathematics score (3)	Extraversion (4)	Agree–ableness (5)	Openness (6)	Neuroticism (7)	Conscientiousness (8)
Panel A. Lower-track students								
MSR	0.089	−0.071	0.214	2.024	2.182	0.789	0.045	1.058
	[0.115]	[0.098]	[0.131]	[1.099]	[1.594]	[0.982]	[0.836]	[1.004]
MSR × Boy	0.076	0.102	0.037	0.492	0.774	0.575	0.808	0.422
	[0.104]	[0.103]	[0.113]	[0.910]	[1.283]	[0.725]	[0.812]	[0.780]
Boy	−0.101	−0.255	0.015	−0.596	−0.246	−0.598	−0.810	−0.778
	[0.076]	[0.089]	[0.075]	[0.670]	[0.862]	[0.590]	[0.491]	[0.574]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	901	901	901	901	901	901	901	901
Panel B. Upper-track students								
MSR	0.024	−0.043	0.088	1.669	2.502	0.420	0.416	1.032
	[0.060]	[0.067]	[0.079]	[0.694]	[1.350]	[0.489]	[0.732]	[0.767]
MSR × Boy	0.040	0.089	0.041	0.748	−0.202	0.324	−0.636	−0.141
	[0.062]	[0.078]	[0.086]	[0.717]	[0.894]	[0.537]	[0.756]	[0.689]
Boy	−0.007	−0.116	0.037	−0.526	0.347	−0.053	0.365	−0.007
	[0.036]	[0.051]	[0.067]	[0.514]	[0.674]	[0.381]	[0.601]	[0.478]
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 MSR intervention on students' test scores and "big five" personality traits. "Boy" is a dummy that takes a value of 1 if the student is male and 0 otherwise. Panel A reports estimated results for lower-track students, and panel B reports estimated results for upper-track students. MS and its interaction with "Boy" are also included in all regressions. The controls include the corresponding own baseline scores, age, height, health status, hukou registration status, minority status, 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 reported in brackets.

Table 8 presents an analysis of the gender composition of deskmates.[32,33] We generate a dummy "Different gender" that takes a value of 1 if the gender of two students sharing a desk is different and a value of 0 otherwise. Then, we interact this dummy with each of the two treatment variables. Columns 1–3 in panel A show that for lower-track students, the estimated effects of the MSR intervention on test scores do not vary by the gender composition of deskmates. Column 3 in panel B shows that for upper-track students, the estimated coefficient on "MSR × Different gender" is 0.24 (SE = 0.096), implying that upper-track students gain in mathematics scores if they are assigned with deskmates of different gender in the MSR classes. Columns 4–8 in Table 8 report the heterogeneous analyses of MSR treatment effects on students' "big five" personality traits. For lower-track students, the estimated coefficients on "MSR × Different gender" are −1.45 (SE = 0.836) and −1.68 (SE = 0.835) for extraversion and conscientiousness, respectively (columns

[32] One concern is that the gender composition of deskmates may be endogenous. We calculate the means of students' characteristics for each class and perform a class-level regression. In particular, we regress the proportion of deskmates of different gender on the proportion of boys, proportion of nonminorities, average age, average height, average years of schooling of parents, proportion of students who have urban hukou, proportion of students who have computers and/or cars at home, and school-by-grade fixed effects. Standard errors are clustered at the school-by-grade level. We find that our treatments cannot significantly affect the proportion of mixed-gender deskmates in the classes

[33] The total number of pairs of same-gender and different-gender deskmates is 434 (434) and 467 (467), respectively, in the lower track (upper track).

Table 8—Effects of MSR Intervention by Gender Composition of Deskmates

	Average score (1)	Chinese score (2)	Mathematics score (3)	Extraversion (4)	Agree-ableness (5)	Openness (6)	Neuroticism (7)	Conscientiousness (8)
Panel A. Lower-track students								
MSR	0.177	0.055	0.238	3.042	3.618	1.317	1.014	2.148
	[0.074]	[0.078]	[0.089]	[0.691]	[1.091]	[0.709]	[0.570]	[0.736]
MSR × Different gender	−0.092	−0.148	0.005	−1.446	−1.746	−0.464	−1.007	−1.684
	[0.075]	[0.093]	[0.091]	[0.836]	[1.112]	[0.676]	[0.737]	[0.835]
Different gender	0.052	0.081	−0.016	0.243	−0.218	0.512	0.400	0.275
	[0.057]	[0.056]	[0.076]	[0.734]	[0.774]	[0.587]	[0.607]	[0.763]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	901	901	901	901	901	901	901	901
Panel B. Upper-track students								
MSR	−0.022	−0.011	−0.007	1.668	1.725	0.951	−0.617	0.327
	[0.087]	[0.072]	[0.103]	[0.866]	[1.407]	[0.584]	[0.793]	[0.918]
MSR × Different gender	0.145	0.040	0.240	0.839	1.365	−0.753	1.295	1.221
	[0.086]	[0.099]	[0.096]	[0.820]	[1.256]	[0.583]	[0.798]	[0.854]
Different gender	−0.113	−0.047	−0.177	−0.268	−0.704	0.650	−0.016	−0.149
	[0.049]	[0.079]	[0.045]	[0.517]	[0.762]	[0.430]	[0.631]	[0.671]
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 MSR intervention on students' test scores and "big five" personality traits. "Different gender" is a dummy that takes a value of 1 if the gender of two students sharing a desk is different and 0 otherwise. Panel A reports estimated results for lower-track students, and panel B reports estimated results for upper-track students. MS and its interaction with "Different gender" are also included in all regressions. Controls are added in all regressions. The controls include the corresponding own baseline scores, age, height, health status, hukou registration status, minority status, 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 reported in brackets.

4 and 8 in panel A). It implies that MSR intervention is more helpful in promoting the extraversion and conscientiousness of lower-track students with deskmates of the same gender. We find no evidence of difference in the treatment effects on lower-track students' openness, agreeableness, and neuroticism. When moving to columns 4–8 in panel B, which present heterogeneous analyses of treatment effects on "big five" personality traits of upper-track students, our results do not reveal any systematic heterogeneity in treatment effects with respect to gender composition.

Lu and Anderson (2015) show that peers' gender in neighboring students is an important factor affecting students' test scores. To rule out that our treatment effects are affected by peers' gender, we test the existence of heterogeneity in treatment effects with respect to the share of female students in the neighboring five students. Our results reveal that students' test scores are positively associated with the share of female neighboring students, which is consistent with the finding of Lu and Anderson (2015). However, no heterogeneity in MSR treatment effects has been detected for either test scores or personality traits (see Table A13 in the online Appendix).

B. *Role of Teachers*

In the 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 in accordance to student needs (Burke and Sass 2013). This subsection reports on tests of explanations related to the role of teachers. First, we will show that our treatment does not alter teachers' effort. Second, we present 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 detected peer effects.

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 (i) learning instruction, (ii) daily care for students, and (iii) fairness to students. Additional details of this questionnaire are presented in Table A14 in the online Appendix. 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 A15 in the online Appendix presents the estimated results. Students in treatment classes provide similar evaluations of their teachers compared with those in the control classes, indicating that the teacher's function may not be the driving force of the treatment effect of the MSR intervention.[34]

Peer Effects and Teacher Quality.—High-quality teachers may tailor their teaching methods to meet students' needs in mixed-seating classes. In this sense, the estimated peer effects may include a teacher effect. Table A16 in the online Appendix 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 a value of 0 otherwise. We employ a similar specification to equation (3) except that we add an interaction term between deskmates' baseline z-score and teacher quality.[35] The coefficient on this interaction term can tell whether peer effects vary by teacher quality. None of the coefficients of the "Deskmate's baseline z-score $\times$ Teacher with a bachelor degree" is statistically different from 0. Moreover, the estimated effects of peer quality are close to those in Table 5. 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.

[34] We do not conclude that teachers have 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' scores may be associated with teacher effects. However, even if this possibility 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' personality traits.

[35] Teacher quality is absorbed in the class-by-height-group fixed effect.

C. *Corrected Standard Errors*

One concern is that 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, Gelbach, and Miller (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 percent confidence interval is reported. Table A17 in the online Appendix reports the estimated treatment effects using the corrected standard errors. We find that significant estimates all survive, confirming the findings in Tables 3 and 4.

Data obtained from randomized control trial do not meet the requirements to rely on asymptotic properties, and thus results may be obtained by chance (Young 2019). To reassure that our results are robust, we follow Young (2019) to use randomization inference to calculate p-values. Our analysis shows that none of the estimated coefficients switch from being statistically significant to insignificant (Table 9).

D. *Multiple Hypotheses*

Another concern is that we estimate the effects of the treatment on multiple outcomes, which may raise the issue of multiple hypothesis testing. Table 9 provides Romano and Wolf's (2005, 2016) and Westfall and Young's (1993) p-values along with the original ones. Given that only MSR intervention promotes students' test scores, personality traits, and friendship formation, we suppress the results for MS intervention. Romano–Wolf's (2005, 2016) p-values show that the precision of most of our estimated MSR treatment effects survive this adjustment. An exception is that the treatment effects on the probability of students reporting their deskmates as friends and the number of upper-track friends reported by lower-track students switch from being statistically significant to insignificant. However, Westfall–Young's (1993) p-values show that the effect on the number of upper-track friends remains significant after the adjustment.

V. Comparison with the Literature

In this section, we compare our estimated effects with related literature. In a partial pairing process, Li et al. (2014) find that pairing high- and low-performing classmates as benchmates and offering them rewards improves low-performing 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 the fact that Li et al. (2014) use a partial pairing of very high- and very low-performing students, which may generate larger effects. Moreover, they incentivize both high and low achievers. By comparison, only

Table 9—Multiple Hypothesis and Randomization Tests

	Lower-track students				Upper-track students			
	Original p-value	Romano–Wolf p-value	Westfall–Young p-value	Randomization test p-value	Original p-value	Romano–Wolf p-value	Westfall–Young p-value	Randomization test p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Chinese test raw score	0.885	0.955	0.945	0.888	0.917	0.687	0.401	0.922
Mathematics test raw score	0.018	0.005	0.035	0.018	0.181	0.393	0.313	0.190
Extraversion	0.000	0.005	0.007	0.004	0.005	0.005	0.008	0.004
Agreeableness	0.001	0.010	0.096	0.053	0.054	0.005	0.107	0.058
Openness	0.021	0.099	0.252	0.146	0.245	0.433	0.351	0.212
Neuroticism	0.391	0.970	0.330	0.424	0.979	0.886	0.684	0.907
Conscientiousness	0.091	0.309	0.192	0.108	0.300	0.493	0.366	0.216
Whether a deskmate is nominated as a friend	0.087	0.204	0.157	0.088	0.062	0.279	0.123	0.069
Fraction of upper-track students in one's network	0.011	0.099	0.025	0.013	0.098	0.393	0.175	0.100
Number of lower-track friends	0.000	0.005	0.000	0.000	0.053	0.040	0.111	0.060
Number of upper-track friends	0.012	0.204	0.024	0.013	0.151	0.438	0.284	0.168

Notes: Columns 1 and 5 report original p-values. Columns 2, 3, 6, 7 report p-values for multiple hypothesis test, which controls the probability of rejecting at least one true null hypothesis in a family of hypotheses. Columns 2 and 6 report Romano–Wolf's (2005, 2016) p-values, and columns 3 and 7 report Westfall–Young's (1993) p-values. Columns 4 and 8 report p-values of randomization tests.

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 3 with the within-school tracking effects in the literature.[36] Hoffer (1992) finds that students placed in the high-achieving group in schools outperform otherwise similar students in schools without tracking. The size of the effect for high-achieving students ranges from 0.18–0.26 standard deviations in tracking schools. Argys, Rees, and Brewer (1996) report that tracking boosts scores of students in high-achieving classes by 8.4 percent (or nearly 0.25 standard deviations). In recent literature, Duflo et al. (2011) show that after 18 months in tracking schools, students in the top half of the preassignment 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.[37]

In this study, 20 weeks of studying in the mixed-seating environment with incentivized high-achieving students increased the endline mathematics scores of low-achieving students in the classes by 0.24 standard deviations. These effects are smaller than the effect in the study of Card and Giuliano (2016), but they are close

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

[37] Table A18 in the online Appendix details the comparison between the existing literature and our study, including the estimated magnitudes and standard deviations. Our results are comparable to those in the literature even if we take into account the sample dispersion.

to the magnitude estimated by Hoffer (1992) and Duflo et al. (2011). Our estimates imply that the mixed-seating classroom with incentives benefits students, especially those who belong to the bottom of the class. Combined with the findings in tracking literature, we may draw the conclusion that tracking is economically efficient for 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-achieving students. Table 5 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. Booij et al. (2017) find that a 1 standard deviation increase of the mean performance of peers raises the number of credit points a student collects by 0.1–0.15 standard deviations. Feld and Zölitz (2017) show that the increased fraction of middle-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 personality traits, Zárate (2019) shows that being assigned with more sociable peers increases students' extraversion and agreeableness scales by 0.067 and 0.066 standard deviations, respectively. Our estimated results show that an 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.

VI. Conclusion

This paper presents novel evidence on two important issues. First, does a peer effect on personality traits exist? Second, what is the possible underlying mechanism of peer effects on student performance? We conduct a randomized peer intervention in elementary schools, where low- and high-achieving students were randomly paired as deskmates in mixed-seating classrooms. To motivate high-achieving students to help their low-achieving deskmates learn, we offer a monetary award and a certificate of commendation to high-achieving students if their deskmates' improved scores rank in the top 10 percent among the low-achieving students in the class. We analyze the outcomes for three groups: normal-seating classes (the control group), mixed-seating classes without incentives (the MS classes), and mixed-seating classes with incentives (the MSR classes).

We estimate the treatment effects of the mixed-seating classes. On the one hand, we find that lower-track students achieve higher scores in mathematics in the MSR classes than those in the control classes, whereas upper-track students are nearly unaffected. In terms of personality traits, upper- and lower-track students become more extraverted and agreeable in the MSR classes. On the other hand, we observe no treatment effects in the MS classes. These results demonstrate the importance of mixed-seating classes with incentives for high-achieving students.

We detect deskmate-level peer effects in the MSR classes: the deskmate's higher test scores promote student's academic performance. Moreover, having a deskmate with high levels of extraversion and agreeableness fosters the development of a student's extraversion and agreeableness. This finding highlights that personality traits can be affected by peers, which is rarely studied in the literature.

Our mechanism analysis shows that students' friendship connections contribute to the peer effects detected at the deskmate level. The granular information of nominated friends for each student allows us to identify the network structure of students in a class and, thus, to pin down the mechanism through which peer effects occur. We observe a higher probability of friendship formation between deskmates in MSR classes than those in control classes. The MSR treatment also helps separate part of social connections among lower-track students who instead would like to form more social ties with high-achieving students. This finding suggests that homophily, which is considered one source of the low performance of low-achieving students, can be remolded. Overall, our analysis of the friend networks is helpful in gaining a better understanding of the peer effects in the educational context.

Our study answers important questions but still has many open questions, thus calling for further research. First, whether similar results can be obtained in different contexts is unknown. For example, ethnicity may be a concern if students of different ethnicities are reluctant to interact. Given that our data contain few minority students, we cannot say if any difference would occur when the number of minority students increases in a class. Second, the analysis raises the question of how sustainable the newly formed friendship is, which relates to an important question of whether the benefits from peers with high academic performance and good personality traits can exist in the long run. Finally, a key factor that undergirds the detected effects of the mixed-seating classes is the reward provided for high-achieving students. It is crucial to know how to ensure continued incentives after the intervention has ended. Thus, our results provide important insights for possible policy intervention, but more research is warranted for long-run policy feasibility and effectiveness.

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. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126 (5): 1865–1911.

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): 623–45.

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 (1): 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): 621–34.

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." *Review of Economic Studies* 76 (4): 1239–67.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90 (3): 414–27.

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., Mark Hoekstra, and Elira Kuka. 2018. "The Long-Run Effects of Disruptive Peers." *American Economic Review* 108 (11): 3377–3415

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): 855–82.

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) 1739–74.

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.

Garlick, Robert. 2018. "Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment." *American Economic Journal: Applied Economics* 10 (3): 345–69.

Golsteyn, Bart H.H., Arjan Non, and Ulf Zölitz. 2021. "The Impact of Peer Personality on Academic Achievement." *Journal of Political Economy* 129 (4): 1052–99.

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 (4): 388–406.

Hoffer, Thomas B. 1992. "Middle School Ability Grouping and Student Achievement in Science and Mathematics." *Educational Evaluation and Policy Analysis* 14 (3): 205–27.

Hsieh, Chih-Sheng, and Lung Fei Lee. 2016. "A Social Interactions Model with Endogenous Friendship Formation and Selectivity." *Journal of Applied Econometrics* 31 (2): 301–19.

Jackson, Matthew O. 2008. *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

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 Paper 110.

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.

Li, Li, Eric Mak, and Chunchao Wang. 2020. "Deskmates: Reconsidering Optimal Peer Assignments within Classroom." Unpublished.

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.

Romano, Joseph P., and Michael Wolf. 2005. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73 (4): 1237–82.

Romano, Joseph P., and Michael Wolf. 2016. "Efficient Computation of Adjusted p-Values for Resampling-Based Stepdown Multiple Testing." *Statistics and Probability Letters* 113: 38–40.

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

Westfall, Peter H., and S. Stanley Young. 1993. *Resampling-Based Multiple-Testing—Examples and Methods for p-Value Adjustment*. New York: John Wiley and Sons.

Wu, Jia, Junsen Zhang, and Chunchao Wang. 2023. "Replication data for: Student Performance, Peer Effects, and Friend Networks: Evidence from a Randomized Peer Intervention." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.38886/E149262V1.

Young, Alwyn. 2019. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *Quarterly Journal of Economics* 134 (2): 557–98.

Zárate, Román Andrés. 2019. "Social and Cognitive Peer Effects: Experimental Evidence from Selective High Schools in Peru." Unpublished.

Zhang, Li-fang. 2011. "Hardiness and the Big Five Personality Traits among Chinese University Students." *Learning and Individual Differences* 21 (1): 109–13.

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