Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups [*]

Fangwen Lu
Renmin University of China
lufangwen@ruc.edu.cn

Michael L. Anderson
UC Berkeley
mlanderson@berkeley.edu

October 31, 2012

Abstract

Many believe that classroom interactions play an important role in students' academic achievement, but there is little evidence on peer effects within sub-classroom groups. We exploit random seat assignment in a Chinese middle school to estimate how the gender of neighboring students affects a student's academic achievement. We find that being surrounded by five females rather than five males increases a female's test scores by 0.2 to 0.3 standard deviations but has no significant effects on a male's test scores. These results suggest a low-cost way to potentially improve performance within the world's largest school system.

JEL Codes: I21, J16

Keywords: Network effects, spillover effects, classroom environment

[*] We thank Jeff Perloff for insightful comments. Any mistakes are the authors' own.

1. INTRODUCTION

Social interactions are believed to play an important role in students' academic achievement. Most peer effects studies in primary and secondary education define peers at the classroom or school level and test whether students are influenced by classroom or school-level averages. However, recent work by Carrell, Sacerdote and West (2011) finds that students may form subgroups within larger peer groups, implying that peer effect analyses at the classroom or school level may miss important interactions within sub-classroom groups.

This paper examines peer effects among subgroups of Grade 7 students by exploiting an experiment with random seat assignment in a Chinese middle school. As is common in most Chinese schools, students in this school stay at a fixed seat in a fixed classroom for most classes, while teachers rotate through the classrooms. In this experiment, students were assigned to blocks of rows based on height and then randomly assigned to seats within blocks. This within-block randomization controls for non-random sorting of students into groups and allows an exploration of peer effects in a microenvironment (i.e., a sub-classroom group).

We find that the gender of nearby students influences a student's performance, but the effects vary according to the student's gender. Having a female deskmate (a student who shares the same desk) increases a student's test scores by 0.05 to 0.08 standard deviations regardless of the student's gender. Being surrounded by five females rather than five males increases a female student's test scores by 0.2 to 0.3 standard deviations but has no effect on a male student's test scores. These effects suggest welfare gains from rearranging students within classrooms. In comparison, however, there is little evidence that baseline test scores of nearby students affect academic performance.

In addition to estimating the effects of nearby students on academic performance, we also explore how students evaluate the influence of deskmates on their academic performance. Students report that deskmates with higher baseline test scores positively influence their own academic performance. Likewise, students report that they positively influence the academic performance of their deskmates if they themselves have higher baseline scores. However, the same measures are not significantly affected by a student's own gender or her deskmate's gender. The discrepancy between the self-reported impacts of peer test scores and gender and the actual impacts of peer test scores and gender highlights the risks of using subjective evaluations as proxies for actual outcomes.

2. LITERATURE REVIEW

A large set of peer effects studies leverage variation in peer groups at the classroom or school level (Hanushek et al., 2003; Angrist and Lang, 2004; Arcidiacono and Nicholson, 2005; Hoxby and Weingarth, 2005; Lyle 2007; Ammermueller and Pischke, 2009; Gould, Lavy and Paserman, 2009), while others explore living arrangements among college students (Sacerdote, 2001; Zimmerman, 2003). These studies generally find evidence of positive spillovers in academic performance. However, to the best of our knowledge no studies leverage experimental or quasi-experimental variation to estimate peer effects within sub-classroom groups.

A related literature explores the effects of student gender on peer outcomes. Morse (1998) and Mael et al. (2005) review observational studies comparing students in single-sex and coeducational classes; some studies suggest single-sex schooling may be beneficial while others indicate no difference. Hoxby (2002) and Lavy and Schlosser (2011) explore plausibly exogenous variation in the gender composition of coeducational schools and find that the proportion of female students has positive effects on students' cognitive achievements. However, gender composition does not have differential effects on boys and girls. Whitmore (2005) finds that students assigned to classrooms with higher proportions female in the Tennessee STAR experiment do better in kindergarten and second grade, with some evidence of differential effects on boys and girls.

Our study extends the rich academic peer effects literature to sub-classroom groups. The results reveal that even within micro-level environments, there can be strong peer effects. This finding has policy relevance because teachers have significant discretion in organizing groups within classrooms. Implementing single-sex groups within classrooms, for example, is less controversial than implementing single-sex classrooms or single-sex schools. Changes to classroom arrangements thus represent a low-cost way to potentially improve academic performance.

3. SCHOOL ENVIRONMENT

This experiment was implemented in a coeducational public middle school in Jiangsu, China. At the beginning of the school year, students in Grade 7 – the starting grade – were assigned to a fixed classroom. They stayed in the same classroom for most classes over the semester, while teachers rotated from classroom to classroom. This arrangement is standard for this middle school and most other schools in China. The middle school is not considered an elite school and does not have special entrance requirements.

Desks and benches are provided in classrooms, typically arranged in sets of rows and columns (see Figure 1). Each desk seats two students, and there are four desks per row with aisles between desks. There are six, seven or eight rows in each classroom depending on the number of students. All students are assigned to fixed seats, and they must stay in their assigned seats during class time. The practice of assigning students to fixed seats helps teachers detect absentees and identify students that misbehave during class.

An administrative teacher assigns seats for the classroom.[1] When assigning seats, student height is a major consideration. Classrooms are typically crowded, and taller students sitting in the front may block the view of shorter students behind them. In a non-experimental setting, the administrative teacher may have personal preferences for assigning seats. For example, some administrative teachers like to put students of the same gender together while others tend to mix genders. Seats may also be dynamically adjusted during the school year as administrative teachers learn more about students. In addition, some parents may request to have their children moved to the front of the classroom or near high performing students.

A typical day consists of a 30-minute reading session in the early morning, four 45-minute lecture sessions in the morning, three 45-minute lecture or study sessions in the afternoon, and one 40-minute study or physical exercise session in the late afternoon. During most sessions, students must stay in their own seats. In lectures, chatting is generally prohibited. During study sessions, students choose what to study for themselves. Students are typically allowed to talk in low voice with neighboring students during study sessions. However, seating arrangements remained fixed during study sessions, so in most cases students can only communicate with neighboring students.

[1] An administrative teacher is a regular teacher with additional managerial responsibilities, which include arranging class events, disciplining misbehavior, communicating with parents, and assigning student seats.

Neighboring students have many opportunities to interact with each other, and different peer groups may influence students in different ways. For example, students can talk with their deskmates without moving at all, but they generally have to turn around to talk with students at adjacent desks. Deskmates can always observe each other with ease, but it is difficult to observe details across rows. Though students may interact with other students across aisles, the columns of desks in these classes rotated every several weeks. Students thus had fewer opportunities to form lasting peer groups across aisles.

Since each column of desks stayed together during this experiment, we define peer groups within columns of desks. The first peer group is the desk itself – each student has a single deskmate. The second peer group consists of "neighbor-4 students." A student's neighbor-4 peers are the two students sitting at the desk directly in front of her and the two students sitting at the desk directly behind her. Students sitting across aisles are not a relevant peer group since columns are rotated every few weeks. For students sitting in the first and last row, their neighbor-4 peers consist of fewer than four students. The last peer group, "neighbor-5 students," consolidates the first two groups. A student's neighbor-5 peers are her neighbor-4 peers plus her deskmate.

To see a concrete example of these peer groups, consider Student 1 in the second row and second column of Figure 1. Student 2 is his deskmate, Students 3 through 6 are his neighbor-4 peers, and Students 2 through 6 are his neighbor-5 peers. For Student 3, as no students sit in front of her, her neighbor-4 peers and neighbor-5 peers include only two (1 and 2) and three (1, 2, and 4) students respectively.

4. EXPERIMENTAL DESIGN

In this experiment, a research group in the local Department of Education randomly assigned students' seats with input from the authors. During the first week of the Fall 2009 semester, the Department of Education requested information on students' names, gender, and heights in each classroom. The basic mechanism for assigning seats is as follows. First, students were sorted from shortest to tallest by gender within each classroom. Then, the first eight students were placed in Block 1 (corresponding to Row 1), the next 16 students were placed in Block 2 (Rows 2 and 3), and the 16-student blocks continued until all students were assigned to blocks. Students taller than 5 feet, 6.5 inches (169 centimeters) were put in a separate block. Finally, a random sequence was generated, and students were randomly permuted and assigned to seats within each block. The size of the last two blocks varies depending on the number of students and distribution students' height within classrooms. Students in shorter groups always sit in front of students in taller groups, but within a block taller students may sit in front of shorter students as a result of the randomization. This did not present challenges in the classroom as all students within the same block are of roughly similar height. Due to the one-child family planning policy and a frequent preference for sons, the ratio of boys to girls was 1.27 in the sample school. As boys and girls were of similar height in Grade 7, we placed four boys and four girls in the first block, and then nine boys and seven girls (1.28 boys per girl) in subsequent blocks until it became infeasible.[2]

Some students required special seat assignments due to near-sightedness, and in some cases parents lobbied for favorable seat assignments. To increase compliance rates, the researchers allowed administrative teachers to list several student names for favorable seat treatments; students in the favored list account for 9% of all students. Students on this list received a seat assignment in either a front row or a middle column.[3] The remaining students in each block were randomly assigned seats. Normal students are thus randomly assigned with respect to their peers, but "favored" students are not randomly assigned with respect to their peers. In particular, favored students are more likely to sit adjacent to other favored students. We thus drop the outcomes for all favored students from our analysis, as these students' seats are not randomly assigned (though our main results are not sensitive to including them). Favored students are still used to construct surrounding peer measures (i.e., the right hand side variables), however; excluding the favored students would introduce measurement error in those measures. We can summarize the assignment procedure as first non-randomly assigning a small number of students, and then randomly assigning the remaining students to the remaining seats.[4]

Administrative teachers were asked to cooperate by adopting the random seat assignments and avoiding seat adjustments over the semester. There were no financial

[2] Boys were 0.4 inches taller than girls on average.

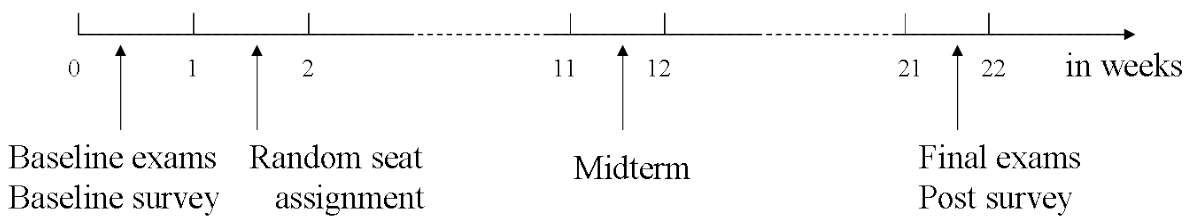
[3] For students on this list, if they were originally assigned to the first four rows, they were moved to the middle columns in the same row. If they were originally in Row 5, they were moved to the middle columns in Row 4. If they were originally behind Row 5, they were moved to Row 5.

[4] The only confounding factor that the non-random assignment of favored students could introduce would be a correlation between sitting near the center of the room and peer characteristics that favored students tend to have. However, we can control for sitting near the center of the room by including column fixed effects. Including these fixed effects does not change our results.

incentives provided to administrative teachers, however, and students were not informed of the research project. It is likely that administrative teachers adjusted seat assignments during the semester so that some students were moved away from their original assignments, but there was no systematic check for compliance in seat arrangement. Strictly speaking, our estimates represent "intent to treat" effects.

5. DATA

The data for this study consist of three rounds of test scores and two rounds of surveys for students of Grade 7 in Fall 2009. We illustrate the data collection timeline by week of semester as follows.



The baseline test and baseline survey were administrated during the first week of the semester before random seat assignment. The random seat assignment was announced during the second week. Students sat according to the random assignment unless the administrative teachers made adjustments. For the midterm and final exams, due to the school's efforts to prevent cheating, students were seated such that students in the same classroom were generally spread over more than ten rooms and no student sat immediately adjacent to another student from the same class. Two teachers monitored the exams in each classroom. The post survey was administrated right after the final exam when students were still seated according the seat arrangement for the final exam. As students took exams and surveys in different seats than their experimental assignments, any correlations in outcomes among randomly assigned peers are not likely generated by communication among students when taking the exams or surveys.

Grading was rigorously conducted. Teachers in the same subjects allocated exam questions among themselves so that the same question was always graded by the same teacher. In addition, students' names were hidden during the grading process. In the baseline test, the school tested students on three major subjects – Chinese, English and math. In the midterm and final, the school tested seven subjects – Chinese, English, math, politics, history, geography, and biology. Each of the three major subjects accounted for 150 points in the raw scores, and the other four subjects accounted for approximately 50 points each. The midterm (or final) score represents the sum of all seven scores. We also create a combined measure of the midterm and final scores – the "combined score" – which is the mean of the midterm and final scores. All test scores are standardized to have mean 0 and standard deviation 1. Figure 2 presents the kernel density of students' baseline scores by gender. The test scores are left skewed for both boys and girls, and girls had higher scores overall. Figure 3 illustrates the kernel densities of total scores for the midterm and final exams.

In addition to the administrative data on students' gender, height and test scores, the surveys provide information on students' family backgrounds and subjective interests. The surveys also report students' evaluations of peer influences. Panel A of Table 1 presents baseline summary statistics, while Panel B presents post-experiment summary statistics.

We use three types of peer groups in this study – deskmates, neighbor-4 peers, and neighbor-5 peers. For each type of peer group, we construct two measures of the peer characteristics: gender composition (whether the deskmate is female or the proportion of females among neighbor-4 and neighbor-5 peers) and baseline total test score. Panel C of Table 1 presents summary statistics of peer characteristics.

Random seat assignments allow us to identify the casual effects of neighboring students. The traditional method for verifying random assignment is to regress baseline characteristics of student *i* on student *i*'s peers' characteristics. However, since sampling is performed without replacement, a simple bivariate regression may generate mechanical negative correlations. If student *i* has a high share of female peers, for example, then it is more likely that student *i* is male. We address this issue in our main regressions by controlling for student *i*'s own value of the corresponding peer characteristic in the regression (Guryan, Kroft, and Notowidigdo 2009). For example, when testing whether deskmate gender is correlated with a student's academic background, we control for the student's own gender to eliminate the mechanical negative correlation. We thus estimate regressions of the form:

$$X_{cbi} = \alpha_1 Peer_{cbi} + \alpha_2 Female_{cbi} + \alpha_3 Baseline\ Score_{cbi} + \lambda^{cb} + u_{cbi} \quad (1)$$

The variable X_{cbi} represents a baseline characteristic for student *i* in block *b* of class *c*. The regressor of interest, $Peer_{cbi}$, represents the gender or baseline test score of student *i*'s deskmate, neighbor-4, or neighbor-5 students. The regressors $Female_{cbi}$ and $Baseline\ Score_{cbi}$ control for student *i*'s gender and baseline test score to eliminate the mechanical negative correlation discussed above. The term λ^{cb} contains block fixed effects; these fixed effects are important for identification since randomization occurs within blocks.

Statistical inference in equation (1) (and all other regressions in our paper) is complicated by the clustered nature of the data. Outcomes are likely correlated within classrooms, and neighbor-4 peer measures are correlated within classrooms by construction – each student belongs to more than one neighbor-4 peer group. One solution is to cluster by classroom, but with only twelve classrooms there are very few clusters, potentially biasing the clustered standard errors. Instead, we use our knowledge of the randomization procedure to perform exact permutation tests. These tests are derived solely from the actual randomization and thus have the appropriate size regardless of the dependence structure of the data (Rosenbaum 2007). In essence, we rerun the experiment 10,000 times and compute the resulting distribution of *t*-statistics. Under the sharp null hypothesis of no treatment effect, this distribution can be used for statistical inference. Unlike cluster bootstrap-based techniques (e.g., Cameron, Gelbach, and Miller 2008), these tests remain valid even for small numbers of clusters, since they are derived from the randomization procedure itself.

To implement the exact permutation tests, we randomly permute the seat assignments according to the original assignment procedure. For each permutation, we calculate $Peer_{cbi}$ based on the permuted seat assignments and estimate equation (1). We then collect the *t*-statistics from 10,000 permutations and compute the distribution of these *t*-statistics. We compare the actual *t*-statistic for a given regression to the distribution of *t*-statistics from the 10,000 random permutations. The *p*-value, reported in italics in all tables, represents the fraction of random permutation *t*-statistics that are larger than the actual *t*-statistic.

Table 2 tests for non-random sorting into peer groups. The table presents results from regressions of student *i*'s predetermined characteristics on peer gender composition and test scores, controlling for block fixed effects and student *i*'s baseline characteristics (when possible). Each cell represents a separate regression. For gender and baseline test score – the first two rows – we cannot control for student *i*'s gender or baseline test score (otherwise the regression would have an R^2 of 1). We thus omit student *i*'s gender as a regressor in the first row and omit student *i*'s baseline test scores as a regressor in the second row. The resulting regressions display weak evidence of negative mechanical correlation. However, the permutation *p*-values, reported in italics, account for the mechanical negative correlation, since it is inherent in the sampling procedure.[5] Their insignificance demonstrates that the negative coefficients are consistent with the null hypothesis of random assignment.

The subsequent rows of Table 2 test for correlations between student *i*'s characteristics and peer gender or peer baseline test scores. Characteristics include height, age, birth order, mother's and father's education, and interest in Chinese, math, and English. The results are consistent with null effects for all tests. Of the 66 tests in Panel C, only one test, the relationship between age and deskmate gender, is statistically significant at the 5% level (p = 0.04 for this test).[6] Nevertheless, to be conservative we control for baseline characteristics in all subsequent regressions. Our conclusions are unaffected by the inclusion of these controls. The tests are also all insignificant if we estimate the regressions separately for males and females (results available upon request).

The last row of Table 2 tests for correlations between attrition and peer gender or peer baseline test scores. There is a 14% attrition rate in our main regressions, but almost all of this attrition is due to missing values of baseline covariates rather than missing outcomes (the attrition rate for the midterm exam is 1% and the attrition rate for the final exam is 0.6%). We thus expect attrition to occur randomly, since there is no way for seating assignment to affect the attrition of baseline covariates. Indeed, the last row confirms that there is no significant relationship between attrition and peer gender or baseline test scores.

6. RESULTS

[5] The random permutation *t*-statistics in the first three columns of row 1 and the last three columns of row 2 are centered around negative values due to mechanical negative correlation. We thus test whether the observed *t*-statistics in these rows are extreme relative to the permutation *t*-statistics, which by construction are affected by mechanical correlation.

[6] Unlike many randomized trials, we could not mechanically enforce covariate balance by repeating the randomization procedure until all covariate balance tests are statistically insignificant. This is because the randomization procedure was performed prior to the processing of the survey data.

6.1 Main Effects of Peers on Academic Performance

Given the within-block randomization of seats, we estimate the main effects of peers using the following equation:

$$Y_{cbi} = \beta_1 Peer_{cbi} + \gamma X_{cbi} + \lambda^{cb} + e_{cbi} \tag{2}$$

The outcome Y_{cbi} represents standardized test scores for student *i* in block *b* in class *c*. The regressor of interest, $Peer_{cbi}$, represents the gender or baseline test score of student *i*'s deskmate or neighbor-4 students. The term X_{cbi} now includes all of student *i*'s baseline characteristics – gender, baseline test score, height, age, birth order, mother's and father's education, and interest in Chinese, math, and English.[7] The term λ^{cb} contains block fixed effects, which are necessary since randomization occurs within blocks. Statistical inference is performed using exact permutation tests, as described in Section 5. For additional power, each table contains one or two columns in which we estimate the effects of peer characteristics on the "combined" midterm and final score. This consolidates the number of tests by estimating the effect of peer characteristics on overall exam performance.

Table 3 presents results from estimating equation (2). Each column represents a separate regression. Overall, a female deskmate increases a student's test scores by an average of 0.07 standard deviations (last column; s.e. = 0.03). The effect on the combined midterm and final score is statistically significant, though the effects on the final score itself are not significant. A deskmate's baseline test score, however, does not affect student *i*'s midterm or final score (β_1 = 0.012, s.e. = 0.018). The inclusion of neighbor-4 gender or baseline test scores does not affect the coefficients on deskmate gender or baseline test scores, which is not surprising since deskmates and neighbor-4 students should be uncorrelated under the random seat assignment. The coefficients on the proportion of females among neighbor-4 students and the baseline test scores of neighbor-4 students are generally positive (β_1 = 0.018 and β_1 = 0.030 respectively) but statistically insignificant.

6.2 Heterogeneous Treatment Effects

[7] Excluding all baseline characteristics except gender and baseline test score (which are necessary to control for mechanical negative correlation) does not change our conclusions.

If peer gender has similar effects on test scores for all students, then it is difficult to achieve net improvements on test scores by rearranging students. However, heterogeneous treatment effects are of policy interest because they may provide opportunities for improving welfare by rearranging students. The first three columns in Table 4 present regressions in which we estimate the effects of peer gender composition separately for males and females by interacting the peer measures with gender. The coefficients on the interaction between student *i*'s gender and having a female deskmate, presented in the first row, are small (0.02 standard deviations or less) and statistically insignificant; female deskmates do not appear to affect male and female students differently. The coefficients on the interaction between student *i*'s gender and the proportion females among neighbor-4 students, presented in the second row, imply that female students benefit significantly more than males from sitting near other females. The interaction coefficients are significant for both midterm and final scores, and the effect on the combined midterm and final score is highly significant (0.30 standard deviation effect, s.e. = 0.09). The main effect of neighbor-4 gender composition, however, is negative and statistically insignificant (–0.13 standard deviations, s.e. = 0.11), suggesting that males do not benefit from sitting near females.

The last three columns in Table 4 examine whether peer gender composition affects students with different academic backgrounds differently. The interactions between student *i*'s baseline score and the gender of student *i*'s deskmate and neighbors range from –0.04 to 0.12 standard deviations and are all statistically insignificant, suggesting that the effects of peer gender does not vary strongly by baseline test score.

To estimate the combined effects of deskmates and neighbor-4 students, Table 5 estimates the effects of neighbor-5 students for each gender (recall that student *i*'s neighbor-5 peers are her deskmate plus her neighbor-4 peers). For every outcome, female students have positive and statistically significant effects on neighboring female students. However, females have no effects on neighboring male students. Moving a female student from an all-boy microenvironment to an all-girl microenvironment, for example, increases her test scores by approximately 0.2 standard deviations. Male students, in contrast, display no similar response to gender peer composition, and we can rule out an effect for males above 0.13 standard deviations with 95% confidence. If anything, males may benefit from moving to a all-boy microenvironment, as the coefficient on the female share is –0.095 for males (s.e.

= 0.130). However, this result is statistically insignificant. There is also no evidence that students with better baseline test scores have significant impacts on students seated nearby.

Table 6 tests the robustness of our results to dropping the first and last rows from the analysis. Recall that students in the first (last) row have smaller neighbor-4 and neighbor-5 groups because there are no students sitting directly behind (ahead) of them. This fact may attenuate our estimates, particularly if the effect increases in the number of surrounding students. Dropping the first and last rows from our analysis (but not from the constructed peer measures) increases the effect sizes for both females and males. For females, the effect of moving from an all-boy microenvironment to an all-girl microenvironment is now 0.27 standard deviations (s.e. = 0.08). For males, the effect of moving from an all-boy microenvironment to an all-girl microenvironment is now –0.31 standard deviations (s.e. = 0.17). The male coefficient is now marginally significant, suggesting that males may also benefit from gender homogeneous microenvironments.

Table 7 explores whether the effect of neighboring female students is linear in female share. We modify equation (1) to include dummy variables for four categories: female share of neighbor-5 students is 20–40%, female share of neighbor-5 students is 40–60%, female share of neighbor-5 students is 60–80%, and female share of neighbor-5 students is 80–100%. The omitted category is a female share of neighbor-5 students from 0–20%. The dummy variable coefficients are estimated with limited precision, but for female students there appears to be a general upward trend in the female share coefficients. In general, the 80–100% share coefficient is larger than the 60–80% share coefficient, the 60–80% share coefficient is larger than the 20–40% share coefficient, and the 20–40% share coefficient is positive. The only anomaly occurs with the 40–60% share coefficient, which is always smaller than the 20–40% coefficient. Overall there is no evidence of strong nonlinearities in the relationship between a female student's performance and the female share of her neighbor-5 peers. There is also no evidence of a relationship between a male student's performance and the female share of his neighbor-5 peers.

6.3 Effects of Peers on Self-Reported Peer Relationships

The endpoint survey includes four specific questions on the relationship between the surveyed student and his deskmate: (1) what is the effect of the deskmate on the surveyed student's academic performance; (2) what is the effect of the surveyed student on the deskmate's academic performance; (3) how well do the surveyed student and the deskmate get along with each other; (4) how strongly does the surveyed student wish to remain in her current seat.

Table 8 presents the effects of peer gender and academic background on survey responses.[8] The table also presents coefficients on student *i*'s gender and baseline test scores for the purpose of comparing columns 1 and 2. Column 1 presents the reported effects of deskmate gender and academic background on student *i*'s academic performance. A deskmate that scores one standard deviation better is reported to have a positive and statistically significant effect on student *i*'s academic performance (0.13 standard deviations, s.e. = 0.06). A female deskmate has a positive perceived effect (0.16 standard deviations, s.e. = 0.11), but the effect falls short of statistical significance. Column 2 demonstrates that when student *i* evaluates his own effect on his deskmate's academic performance, student *i* believes that he is more helpful if he has a better test score (0.21 standard deviations, s.e. = 0.06). This result is consistent with the results in column 1, which show that students believe their deskmates' positive influence is increasing in deskmate test scores. However, neither result is consistent with the actual effects of higher scoring deskmates, which are small and statistically insignificant (see Table 3). In comparison, student *i* does not feel that she has a significantly larger effect on her deskmate if student *i* is female instead of male (0.07 standard deviations, s.e. = 0.08).

It is commonly presumed in Chinese culture that good students have positive academic impacts on others, as expressed in old sayings such as "being close to red and then turn red; being close to black and then turn black". In contrast, there is no common belief that female students have positive academic impacts relative to male students. The discrepancy between the perceived effects of high scoring peers and the actual effects of high scoring peers thus suggests that the survey responses may be biased towards common beliefs. Our results therefore cast doubt on the use of subjective evaluations as proxies for actual outcomes.

---

[8] Although the survey responses are categorical variables, we choose to present results from linear regression models rather than ordered logit models. This is because the within block randomization requires us to control for block fixed effects, and the ordered logit model may suffer from the incidental parameters problem with a large number of fixed effects. Nevertheless, the coefficient signs and significance levels are similar if we instead estimate ordered logit models.

Columns 3 and 4 examine whether deskmate gender or academic background affect deskmate relationships. Peer characteristics have no significant effects on either the surveyed student's relationship with her deskmate or her desire to change seats. Students with better academic backgrounds, however, report better relationships with their deskmates.[9]

7. CONCLUSION

We identify peer effects within subgroups inside classrooms by exploiting the random assignment of seats in a Chinese middle school. The results suggest that having a female deskmate is beneficial for both boys and girls, while having more female neighbors has significant positive effects on girls but no positive impact on boys. The differing patterns between the deskmate results and the neighbor-4 results may be due to differences in interactions – interactions between deskmates are easier and to some degree unavoidable, while interactions with neighboring students are optional. However, we lack sufficient power to conclude that boys benefit from female deskmates. In other words, we cannot reject the hypothesis that a female deskmate and a female neighbor-4 student have the same effect on females, but neither has an effect on males.

It is interesting to compare the results of this study to the results of classroom-level and school-level studies on gender peer effects. Whitmore (2005) finds that increasing the classroom female share by 20 percentage points increases kindergarten test scores by approximately 0.1 standard deviations. In comparison, we find that increasing the female share of neighboring students by 20 percentage points increases female test scores by approximately 0.04 standard deviations. Whitmore finds no difference in effects for males and females in kindergarten, but a large difference in effects for males and females in third grade – increasing the female share by 20 percentage points increases female test scores by 0.13 standard deviations but decreases male test scores by 0.16 standard deviations.

Lavy and Schlosser (2011) use Israeli data to measure the effect of the fraction female within a school grade level on peer test scores. In eighth grade (the grade closest to our study), they find that a 20 percentage point increase in the female share increases female test scores by 0.06 to 0.08 standard deviations on average. These effects are 50% to 100% larger than our equivalent effects for neighboring female students. They find no significant

[9] We do not view this relationship as having a causal interpretation; it is simply a descriptive fact.

effect of female share on eighth grade male test scores. However, a 20 percentage point increase in female share raises high school test scores by 0.04 to 0.05 standard deviations for both males and females. These effects are similar in magnitude to our effects (for females).

Externally validity is a key issue in our study as we focus on one middle school in China. Though the classroom layout and rotation of teachers through classrooms in our study is typical of Chinese schools, the characteristics of children in our school may not be representative of the average Chinese child. Table 9 presents summary statistics comparing households in our study's area to the average Chinese household.[10] The first two columns compare households living in all Chinese urban areas to households living in our study's urban area. Households in our study's urban area are slightly more educated than households in the average Chinese urban area. They are also less likely to have running water or toilets. However, these differences are modest in magnitude even when statistically significant (e.g., less than 0.25 standard deviations). The last two columns of Table 9 compare all Chinese areas to our study's overall area. The differences between the last two columns are even smaller than the differences between the first two columns, perhaps because the sample sizes are larger in the last two columns. In the case of parental education, the difference between all urban areas and all areas is larger than the difference between all urban areas and our study's urban area. This suggests that the main issue for external validity may be the urban-rural divide rather than the specific area in which we conducted our study. We would caution against drawing specific conclusions regarding effect sizes in rural Chinese areas.

We are likewise hesitant to extrapolate our results to other countries. The Chinese education system is heavily structured around the classroom and includes in-class study sessions. Furthermore, students stay in the same seat throughout the day and do not switch rooms. This stands in sharp contrast to systems in the United States and the United Kingdom, where middle school students change classrooms many times throughout the day. These frequent room changes ensure that a student's classroom peers are constantly changing, likely affecting how peer relationships impact performance.

These caveats notwithstanding, our results demonstrate the potential for net test score gains from improving classroom arrangements within the Chinese context. This finding suggests a low cost way to improve test scores within a significant segment of the

[10] Data come from a 0.1% sample of the 2000 Chinese Census.

world's largest education system, and it underscores the potential return to further research on peer effects within sub-classroom microenvironments.

REFERENCES

Ammermueller, Andreas, and Jörn‐Steffen Pischke. 2009. "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study." *Journal of Labor Economics* 27(3): 315-348.

Angrist, Joshua D., and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *The American Economic Review* 94(5): 1613-1634.

Arcidiacono, Peter, and Sean Nicholson. 2005. "Peer effects in medical school." *Journal of Public Economics* 89(2-3): 327-350.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90(3): 414-427.

Carrell, S.E., B.I. Sacerdote, and J.E. West. 2011. "From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects." *NBER Working Paper*.

Gould, Eric D, Victor Lavy, and M. Daniele Paserman. 2009. "Does Immigration Affect the Long‐Term Educational Outcomes of Natives? Quasi‐Experimental Evidence." *The Economic Journal* 119(540): 1243-1269.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo. 2009. "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments." *American Economic Journal: Applied Economics* 1(4): 34-68.

Hanushek, Eric A et al. 2003. "Does peer ability affect student achievement?" *Journal of Applied Econometrics* 18(5): 527-544.

Hoxby, C.M., and G. Weingarth. 2005. "Taking race out of the equation: School reassignment and the structure of peer effects."

Hoxby, Caroline M. 2002. "How Does the Makeup of a Classroom Influence Achievement?" *Education Next* 2(2): 56-63.

Lavy, Victor, and Analía Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics* 3(2): 1-33.

Lyle, David. 2007. "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point." *Review of Economics and Statistics* 89(2): 289-299.

Mael, Fred et al. 2005. *Single-Sex Versus Coeducational Schooling: A Systematic Review*. Jessup, MD: US Department of Education.

Morse, S. 1998. *Separated by Sex: A Critical Look at Single-Sex Education for Girls*. Washington, DC: Amer Assn of Univ Women.

Rosenbaum, Paul R. 2007. "Interference Between Units in Randomized Experiments." *Journal of the American Statistical Association* 102: 191-200.

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

Whitmore, Diane. 2005. "Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment." *The American Economic Review* 95(2): 199-203.

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

Figure 1: Arrangement of a classroom

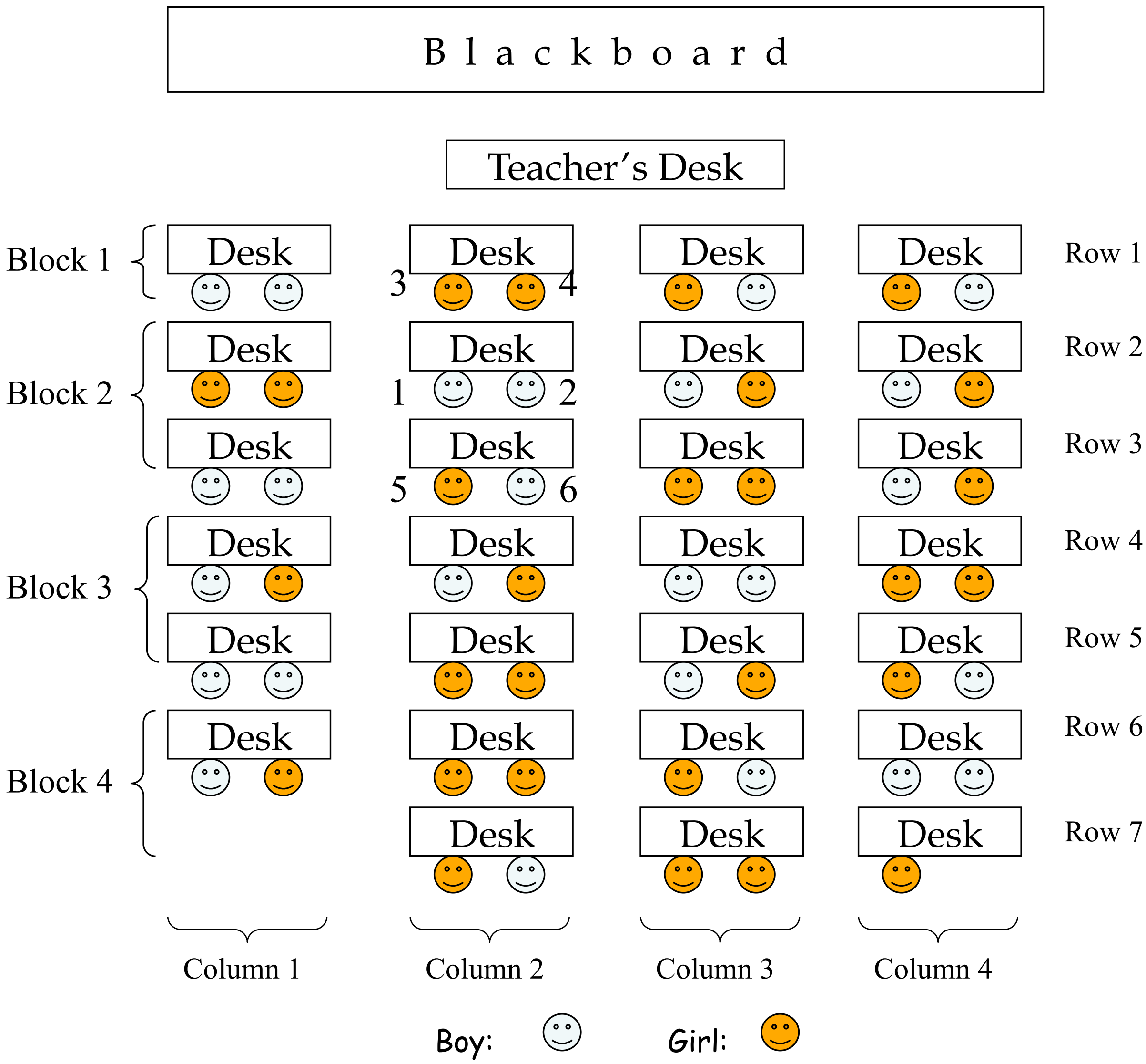


Figure 2: Distribution of baseline scores by genders

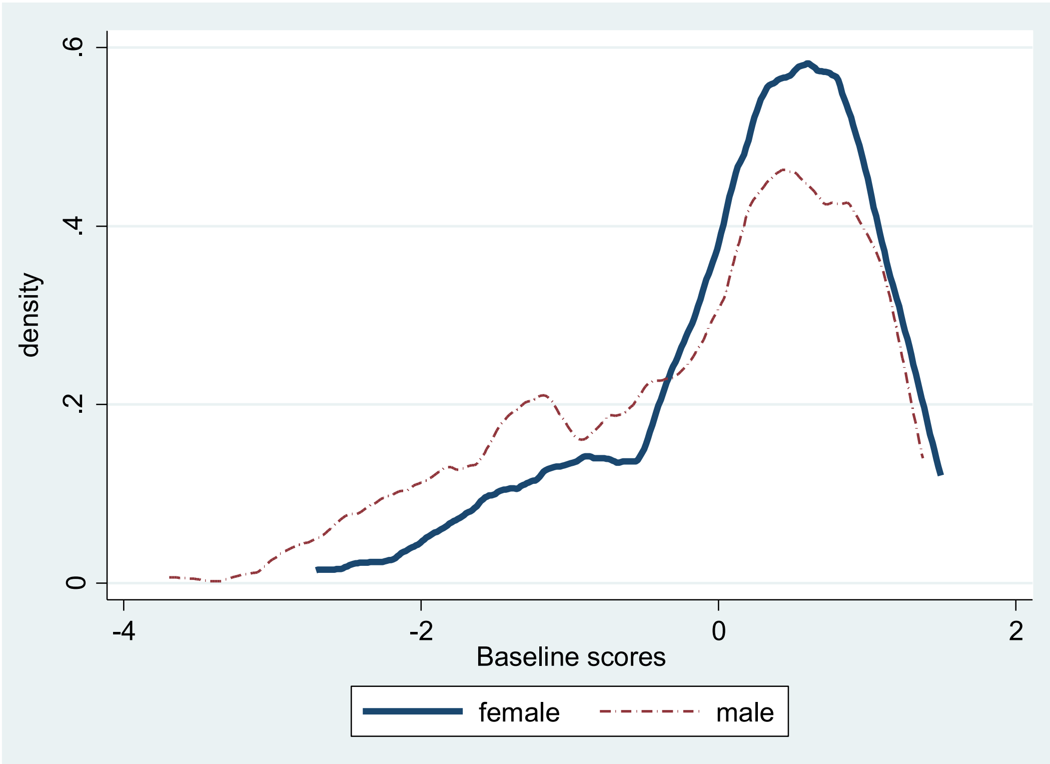


Figure 3: Distributions of midterm and final scores by genders

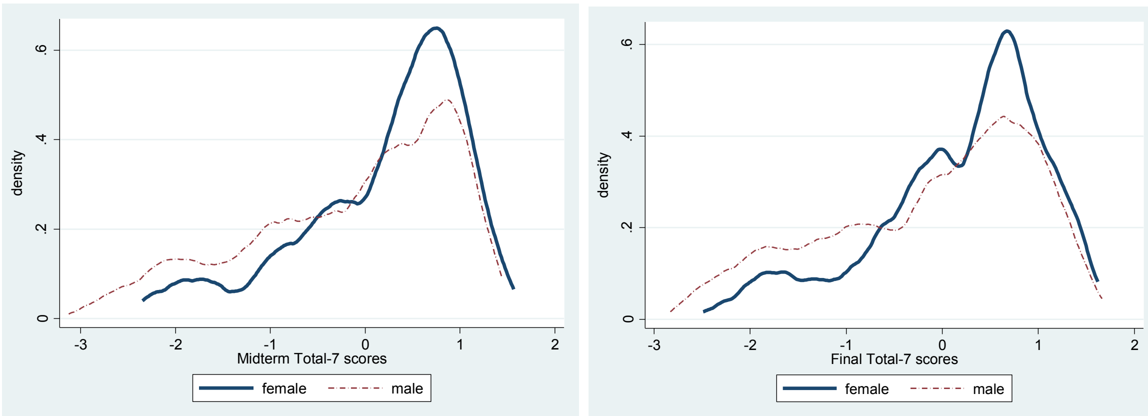


Table 1: Summary Statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
Panel A: Baseline characteristics					
Female	682	0.43	0.5	0	1
Baseline test score	680	0.00	1.00	-3.69	1.50
"Favored" student	682	0.09	0.29	0	1
Body height (cm)	682	156.43	6.71	135	180
Age (years)	655	12.47	0.55	10.17	14.75
Birth Order	680	1.42	0.75	1	7
Father's education	650	11.65	3.04	3	19
Mother's education	647	10.84	3.07	3	19
Interest in Chinese (pre)	643	4.02	0.81	1	5
Interest in English (pre)	638	3.93	1.05	1	5
Interest in Math (pre)	643	4.05	0.85	1	5
Panel B: Post experiment characteristics					
Midterm score	675	0.00	1.01	-3.12	1.57
Final score	677	0.01	1.01	-2.83	1.66
Effects of deskmate on student *i*'s study	663	3.33	1.04	1	5
Effects of student *i* on deskmate's study	661	3.48	0.86	1	5
Relations with deskmate	661	3.74	1.13	1	5
Desire to change seats	669	3.34	1.34	1	5
Panel C: Characteristics of peers					
Female deskmate	672	0.44	0.5	0	1
% females in surrounding 4 students	682	0.44	0.27	0	1
% females in surrounding 5 students	682	0.44	0.24	0	1
Baseline score of deskmate	670	0.001	1	-3.69	1.5
Average baseline score of surrounding 4 students	682	-0.001	0.57	-2.33	1.31
Average baseline score of surrounding 5 students	682	0	0.5	-1.97	1.31

Table 2: Relationships Between Peer Measures and Baseline Characteristics

Peer measure (independent variable):	Deskmate female	Neighbor-4 female share	Neighbor-5 female share	Deskmate score	Neighbor-4 avg score	Neighbor-5 avg score
Baseline characteristic (dependent variable):						
Female	-0.039	-0.155	-0.235	-0.012	0.024	0.004
	(0.064)	(0.095)	(0.135)	(0.023)	(0.038)	(0.048)
	0.450	*0.646*	*0.916*	*0.580*	*0.534*	*0.946*
Baseline test score	-0.040	0.084	0.018	-0.059	-0.098	-0.159
	(0.094)	(0.148)	(0.179)	(0.058)	(0.130)	(0.175)
	0.653	*0.580*	*0.943*	*0.685*	*0.704*	*0.474*
Height (cm)	0.136	0.442	0.669	-0.090	0.196	0.157
	(0.193)	(0.687)	(0.764)	(0.102)	(0.273)	(0.320)
	0.499	*0.542*	*0.415*	*0.399*	*0.494*	*0.640*
Age (years)	-0.066	0.029	-0.053	-0.015	0.017	-0.009
	(0.029)	(0.091)	(0.107)	(0.026)	(0.044)	(0.048)
	0.042	*0.766*	*0.639*	*0.580*	*0.707*	*0.855*
Birth Order	-0.118	-0.117	-0.272	0.003	-0.125	-0.136
	(0.059)	(0.142)	(0.140)	(0.032)	(0.065)	(0.076)
	0.073	*0.442*	*0.081*	*0.927*	*0.083*	*0.104*
Father's education	0.139	0.339	0.623	-0.149	0.242	0.104
	(0.323)	(0.584)	(0.607)	(0.142)	(0.216)	(0.338)
	0.675	*0.575*	*0.330*	*0.319*	*0.294*	*0.775*
Mother's education	-0.075	0.629	0.590	-0.048	-0.100	-0.125
	(0.209)	(0.315)	(0.286)	(0.083)	(0.293)	(0.346)
	0.728	*0.070*	*0.066*	*0.583*	*0.739*	*0.735*
Interest in Chinese (pre)	0.049	0.051	0.092	-0.061	0.062	-0.014
	(0.074)	(0.156)	(0.125)	(0.033)	(0.051)	(0.051)
	0.519	*0.749*	*0.474*	*0.092*	*0.253*	*0.794*
Interest in English (pre)	0.110	-0.120	-0.011	-0.051	0.000	-0.051
	(0.075)	(0.157)	(0.146)	(0.032)	(0.050)	(0.043)
	0.165	*0.467*	*0.943*	*0.147*	*1.000*	*0.264*
Interest in Math (pre)	-0.064	-0.145	-0.239	-0.012	0.047	0.040
	(0.040)	(0.116)	(0.137)	(0.023)	(0.080)	(0.085)
	0.136	*0.237*	*0.109*	*0.604*	*0.579*	*0.653*
Missing Any Covariate Values	0.013	-0.047	-0.033	0.014	-0.009	0.015
	(0.035)	(0.047)	(0.070)	(0.016)	(0.042)	(0.037)
	0.725	*0.339*	*0.648*	*0.409*	*0.839*	*0.702*

Notes: Each cell represents a separate regression. The cell in row *i* and column *j* reports the results from a regression of the dependent variable in row *i* on the peer measure in column *j*. "Favored" students with non-random seat assignments are excluded. All regressions control for gender and baseline test score (if possible). Parentheses contain standard errors clustered by classroom. Permutation-based *p*-values are reported in italics.

Table 3: Effects of Peer Gender and Baseline Score on Test Scores

Dependent variable:	Midterm	Midterm	Final	Final	Combined Midterm + Final	Combined Midterm + Final
Peer measures (independent variables):						
Female deskmate	0.084	0.083	0.047	0.048	0.065	0.065
	(0.039)	(0.035)	(0.029)	(0.030)	(0.031)	(0.029)
	0.055	*0.035*	*0.134*	*0.135*	*0.060*	*0.048*
Baseline score of deskmate	0.002	0.004	0.020	0.020	0.011	0.012
	(0.018)	(0.018)	(0.019)	(0.019)	(0.018)	(0.018)
	0.915	*0.829*	*0.316*	*0.318*	*0.553*	*0.516*
Share female in Neighbor-4		0.026		0.010		0.018
		(0.081)		(0.090)		(0.083)
		0.760		*0.916*		*0.838*
Avg baseline score of Neighbor-4		0.061		-0.003		0.030
		(0.039)		(0.030)		(0.032)
		0.153		*0.919*		*0.374*
Observations	532	532	535	535	532	532

Notes: Each column represents a separate regression. "Favored" students with non-random seat assignments are excluded. All regressions control for gender, baseline test score, and baseline covariates from Table 2. Parentheses contain standard errors clustered by classroom. Permutation-based *p*-values are reported in italics.

Table 4: Heterogeneous Effects of Peer Gender and Baseline Score on Test Scores

Dependent variable:	Midterm	Final	Combined Midterm + Final	Midterm	Final	Combined Midterm + Final
Peer measures (independent variables):						
Female*Female deskmate	-0.022	0.002	-0.013			
	(0.069)	(0.061)	(0.062)			
	0.763	*0.975*	*0.840*			
Female*Share female in Neighbor-4	0.323	0.272	0.299			
	(0.108)	(0.082)	(0.089)			
	0.020	*0.008*	*0.010*			
Baseline score*Female deskmate				-0.051	-0.041	-0.046
				(0.043)	(0.031)	(0.034)
				0.274	*0.227*	*0.219*
Baseline score*Share female in Neighbor-4				0.082	0.120	0.102
				(0.063)	(0.099)	(0.078)
				0.239	*0.273*	*0.242*
Female deskmate	0.087	0.042	0.065	0.086	0.050	0.067
	(0.047)	(0.049)	(0.045)	(0.037)	(0.029)	(0.029)
	0.099	*0.419*	*0.181*	*0.041*	*0.115*	*0.046*
Baseline score of deskmate	0.006	0.022	0.014	0.005	0.021	0.013
	(0.017)	(0.018)	(0.017)	(0.018)	(0.019)	(0.018)
	0.731	*0.249*	*0.431*	*0.786*	*0.297*	*0.487*
Share female in Neighbor-4	-0.129	-0.122	-0.125	0.028	0.013	0.021
	(0.109)	(0.119)	(0.110)	(0.080)	(0.086)	(0.080)
	0.279	*0.348*	*0.299*	*0.738*	*0.891*	*0.805*
Avg baseline score of Neighbor-4	0.060	-0.003	0.029	0.060	-0.005	0.029
	(0.037)	(0.028)	(0.030)	(0.039)	(0.030)	(0.032)
	0.141	*0.917*	*0.361*	*0.160*	*0.868*	*0.394*
Observations	532	535	532	532	535	532

Notes: Each column represents a separate regression. "Favored" students with non-random seat assignments are excluded. All regressions control for gender, baseline test score, and baseline covariates from Table 2. Parentheses contain standard errors clustered by classroom. Permutation-based *p*-values are reported in italics.

Table 5: Effects of Peer Gender and Baseline Score on Test Scores for Females and Males

Dependent variable:	Midterm	Final	Combined Midterm + Final	Midterm	Final	Combined Midterm + Final
Gender:	Female	Female	Female	Male	Male	Male
Peer measures (independent variables):						
Share female in Neighbor-5	0.239	0.179	0.209	-0.099	-0.092	-0.095
	(0.089)	(0.079)	(0.073)	(0.115)	(0.157)	(0.130)
	0.026	*0.046*	*0.018*	*0.420*	*0.575*	*0.488*
Avg baseline score of Neighbor-5	0.017	-0.005	0.008	0.114	0.052	0.084
	(0.067)	(0.058)	(0.060)	(0.089)	(0.077)	(0.081)
	0.809	*0.933*	*0.900*	*0.242*	*0.520*	*0.339*
Observations	246	248	246	296	297	296

Notes: Each column represents a separate regression. "Favored" students with non-random seat assignments are excluded. All regressions control for baseline test score and baseline covariates from Table 2. Parentheses contain standard errors clustered by classroom. Permutation-based *p*-values are reported in italics.

Table 6: Effects of Peer Gender and Baseline Score on Test Scores (Front and Rear Rows Dropped)

Dependent variable:	Midterm	Final	Combined Midterm + Final	Midterm	Final	Combined Midterm + Final
Gender:	Female	Female	Female	Male	Male	Male
Peer measures (independent variables):						
Share female in Neighbor-5	0.271	0.265	0.270	-0.354	-0.270	-0.311
	(0.088)	(0.097)	(0.076)	(0.169)	(0.183)	(0.166)
	0.013	*0.018*	*0.005*	*0.075*	*0.188*	*0.101*
	3.08	2.73	3.55	2.09	1.48	1.87
Avg baseline score of Neighbor-5	-0.019	-0.050	-0.030	0.100	0.055	0.077
	(0.063)	(0.082)	(0.069)	(0.098)	(0.085)	(0.087)
	0.776	*0.560*	*0.687*	*0.343*	*0.546*	*0.406*
	0.30	0.61	0.43	1.02	0.65	0.89
Observations	171	173	171	205	206	205

Notes: Each column represents a separate regression. Students in the front and rear rows, as well as "favored" students with non-random seat assignments, are excluded. All regressions control for baseline test score and baseline covariates from Table 2. Parentheses contain standard errors clustered by classroom. Permutation-based *p*-values are reported in italics.

Table 7: Nonlinear Effects of Peer Gender and Baseline Score on Test Scores for Females and Males

Dependent variable:	Midterm	Final	Combined Midterm + Final	Midterm	Final	Combined Midterm + Final
Gender:	Female	Female	Female	Male	Male	Male
Peer measures (independent variables):						
20-40% females in Neighbor-5	0.056	0.048	0.056	0.038	-0.009	0.015
	(0.131)	(0.111)	(0.118)	(0.129)	(0.171)	(0.147)
	0.683	*0.684*	*0.652*	*0.792*	*0.963*	*0.928*
40-60% females in Neighbor-5	0.028	-0.001	0.017	-0.045	-0.018	-0.031
	(0.141)	(0.080)	(0.107)	(0.118)	(0.135)	(0.124)
	0.855	*0.990*	*0.878*	*0.730*	*0.901*	*0.825*
60-80% females in Neighbor-5	0.151	0.151	0.152	-0.005	-0.035	-0.019
	(0.106)	(0.068)	(0.083)	(0.130)	(0.165)	(0.145)
	0.201	*0.055*	*0.106*	*0.972*	*0.842*	*0.905*
80-100% females in Neighbor-5	0.226	0.144	0.189	-0.100	-0.087	-0.094
	(0.137)	(0.102)	(0.118)	(0.187)	(0.231)	(0.204)
	0.161	*0.215*	*0.169*	*0.629*	*0.732*	*0.665*
Avg baseline score of Neighbor-5	0.021	-0.001	0.012	0.110	0.055	0.083
	(0.069)	(0.059)	(0.061)	(0.091)	(0.075)	(0.082)
	0.769	*0.985*	*0.853*	*0.262*	*0.489*	*0.336*
Observations	246	248	246	296	297	296

Notes: Each column represents a separate regression. "Favored" students with non-random seat assignments are excluded. All regressions control for baseline test score and baseline covariates from Table 2. Parentheses contain standard errors clustered by classroom. Permutation-based *p*-values are reported in italics.

Table 8: Perceived Effects of Peer Gender and Baseline Score

Dependent variable:	Perceived effect of deskmate on student *i*'s study	Student *i*'s perceived effect on deskmate's study	Relations with deskmate	Desire to change seats
Independent variables:				
Deskmate female	0.162	0.032	-0.056	0.238
	(0.111)	(0.101)	(0.143)	(0.142)
	0.177	*0.754*	*0.699*	*0.126*
Deskmate's baseline score	0.133	0.026	0.083	0.032
	(0.056)	(0.049)	(0.057)	(0.090)
	0.041	*0.604*	*0.182*	*0.738*
Student *i* female	0.193	0.065	0.039	0.195
	(0.123)	(0.075)	(0.107)	(0.119)
	0.147	*0.406*	*0.720*	*0.133*
Student *i*'s baseline score	0.135	0.212	0.195	-0.068
	(0.087)	(0.062)	(0.059)	(0.085)
	0.158	*0.008*	*0.007*	*0.459*
Share female in Neighbor-4	0.190	0.008	0.139	-0.025
	(0.170)	(0.114)	(0.141)	(0.226)
	0.301	*0.947*	*0.355*	*0.920*
Avg baseline score of Neighbor-4	0.108	-0.023	-0.000	0.024
	(0.082)	(0.066)	(0.098)	(0.106)
	0.227	*0.730*	*1.000*	*0.833*
Observations	526	525	523	530

Notes: Each column represents a separate regression. "Favored" students with non-random seat assignments are excluded. All regressions control for gender, baseline test score, and baseline covariates from Table 2. Parentheses contain standard errors clustered by classroom. Permutation-based *p*-values are reported in italics.

Table 9: Comparisons Between Study Areas and All Areas

Variable	All Urban	Study Urban	All Areas	Study Area
Years of Education	10.2	10.3	8.8	8.8
	(2.8)	(2.6)	(2.8)	(2.4)
Education ≥ 9 years	0.87	0.91*	0.72	0.75*
	(0.33)	(0.28)	(0.45)	(0.44)
HH size	4.0	4.0	4.2	4.3
	(1.5)	(1.4)	(1.5)	(1.4)
Running water available	0.77	0.75	0.40	0.42
	(0.42)	(0.44)	(0.49)	(0.49)
Toilet available	0.74	0.63	0.70	0.76*
	(0.44)	(0.49)	(0.46)	(0.43)
Households	16,864	51	1,382	186

Notes: Standard deviations in parentheses. * denotes statistically different than All Urban/All Areas average at 5% level.