Gender Peer Effects on Students' Academic and Noncognitive Outcomes: Evidence and Mechanisms*

Jie Gong[†] Yi Lu[‡] Hong Song[§]

This version: July 2019

Abstract

This paper examines gender peer effects on students' academic and noncognitive outcomes. We use a nationally representative survey of middle school students in China and focus on schools that randomly assign students to classrooms. Our findings show that having a higher proportion of female peers in class improves students' test scores and noncognitive outcomes, which include their social acclimation and general satisfaction in school. A further decomposition of channels suggests that teacher behavior, greater student effort, and the improved classroom environment are the primary channels through which peers' gender influences student outcomes.

Keywords: peer effects; gender; education; noncognitive outcomes

JEL Classification: I21, J16, Z13

*We thank Nicola Bianchi, David Figlio, Jonathan Guryan, Kirabo Jackson, Ofer Malamud, Jessica Pan, Ivan Png, Songfa Zhong, and other scholars and seminar participants at Northwestern University, National University of Singapore, Asian Meeting of the Econometric Society and International Symposium on Contemporary Labor Economics for helpful comments. We acknowledge financial support from the Singapore MOE AcRF Tier 1 (R313000129115), Natural Science Foundation of China (71803027), Ministry of Education of China (18YJC790139), and Shanghai Pujiang Talent Program (18PJC012). Any errors are our own.

[†]NUS Business School, National University of Singapore, Singapore. Email: gong@nus.edu.sg

[‡]School of Economics and Management, Tsinghua University, China; Email: luyi@sem.tsinghua.edu.cn

[§]School of Economics, Fudan University, China; Shanghai Institute of International Finance and Economics, China; Email: songhong@fudan.edu.cn

I. Introduction

This paper investigates how classroom gender composition affects students' academic and noncognitve outcomes. Researchers and policymakers have long believed that peer effects—e.g., gender, race, ability, and social background—play important roles in determining student outcomes (e.g., Sacerdote 2001, Zimmerman 2003, Angrist and Lang 2004, Arcidiacono and Nicholson 2005, Ammermueller and Pischke 2009, Carrell et al. 2009, Gould et al. 2009).[1] Understanding the interaction of gender in the educational production function is particularly relevant for the optimal grouping of students in schools and classrooms, and may shed light on the debate concerning single-sex and coeducational schools.

Along this line, previous studies have emphasized the influence of the presence of girls on peers' academic outcomes (e.g., Hoxby 2000; Whitmore 2005; Lavy and Schlosser 2011; Black et al. 2013; Hu 2015). However, little is known about how peers influence other students' noncognitive outcomes. We attempt to fill this gap in the literature by using unique data on individual students' mental stress, social acclimation and general satisfaction in school. These outcomes are valuable not only because they provide a more comprehensive view of students' development through schooling, but also as good predictors of their long-run well-being. Since Jencks et al. (1979), studies have extensively documented the importance of noncognitive skills in explaining long-term significant life outcomes and labor market success (Heckman and Rubinstein 2001; Heckman et al. 2013; Bertrand and Pan 2013). We therefore aim to expand the boundaries of student outcomes and explicitly consider students' noncognitive skills as an output of the educational production process.

Another contribution of our work to the literature lies in our decomposition of the mechanism. Lavy and Schlosser (2011) examine several channels through which peers influence student learning that go beyond the focus on the gender dimension, and suggest that a further important step is to quantify the relative weight of each channel. We exploit rich questionnaires from a nationally representative survey of Chinese middle schools that includes student and teacher behaviors and classroom environments, and use a method following Heckman et al. (2013) and Gelbach (2016) to quantify the importance of each channel.

[1] See Epple and Romano (2011) for an extensive review of the literature.

A common challenge in uncovering peer effects at school is the nonrandom grouping of students. Students with similar backgrounds or characteristics tend to associate with one another, and peer groups tend to be self-selected.[2] For our research question, if there are unobserved characteristics of students that are associated with both gender composition in the classroom and students' outcomes, the estimation of gender peer effects would be biased. To address this identification problem, researchers often exploit cross-cohort variation (Hoxby 2000; Gould et al. 2009; Carrell et al. 2009; Lavy and Schlosser 2011; Black et al. 2013) or use random assignment (Sacerdote 2001; Zimmerman 2003; Carrell et al. 2009; Kremer et al. 2011; Chetty et al. 2011; Shue 2013; Hu 2015). Here, we rely on unique information on classroom assignments, which were obtained from the survey questionnaire, and focus on middle schools in which students are randomly assigned to classrooms.

We use the China Education Panel Survey 2014 (CEPS 2014), which is a nationally representative survey of middle school students and teachers, to estimate how peers' gender composition, as measured by the proportion of female classmates, affects students' academic and noncognitive outcomes. We restrict the sample to schools that randomly place students in classroom. Students in our refined sample cannot self-select into classrooms, and those assigned to the same classroom stay together for learning and extracurricular activities throughout the three years of middle school. A balancing test and robustness checks further comfirm randomized assignment. The main outcome variables include students' test scores, obtained from school administrators, and noncognitive outcomes obtained from their survey responses regarding their mental stress, social acclimation, and general satisfaction in school.

We find that having a higher proportion of female peers in the classroom positively affects students' academic and noncognitive outcomes. Specifically, a 10 percentage point increase in the proportion of female classmates raises students' test scores by 10.2% of a standard deviation and improves their social acclimation and satisfaction in school by 7.7% of a standard deviation. These results are robust after controlling for student and teacher

[2] Manski (1993) documents three types of effects that can generate similar peer outcomes: (1) correlated effects arise when individuals with similar backgrounds self-select into the same group; (2) exogenous effects arise when individuals' predetermined characteristics affect their peers' outcomes; and (3) endogenous effects arise when individuals' outcomes directly affect their peers' outcomes. Since we are interested in the effects of gender, which is fixed, the endogenous effects are not applicable here. The focus of our identification strategy is to separate exogenous effects from correlated effects.

characteristics. We also find some heterogeneity of gender peer effects. For instance, the positive effect on test score is stronger among male students, or when the teacher is male.

By further exploring the mechanisms behind the benefits of having female peers, we find support for four channels: a more interactive teaching style, more time allocated to teaching-related tasks, improved classroom environment, and greater effort exerted by students in learning. In particular, when there are more female students in class, teachers behave differently; they tend to introduce more discussions with and among students, allocate more time to teaching and grading, and be more patient with and responsible for their students. Students also report that the environment is friendlier and more satisfying, and that they devote more hours to homework and tutorials. These changes associated with gender composition may be attributed to the observed benefits for students' learning and noncognitive outcomes. We do not find strong support for ability-based spillover from female students.

Lavy and Schlosser (2011), the closest study to ours, find positive gender peer effects on cognitive outcomes and examine students' behavioral outcomes as mechanisms. Our study also demonstrates a significant benefit for test scores, but we treat noncognitive outcomes as an output of the educational production function and as important measures of student development. Accordingly, we include a richer set of noncognitive measures, such as mental stress and satisfaction in school, which are excluded from Lavy and Schlosser (2011). Another difference is that in examining the mechanisms, Lavy and Schlosser (2011) find that gender composition in the classroom changes students' perceived classroom environment and inter-student and teacher-student relations, but not students' or teachers' own behaviors. Instead, our analysis shows that teacher and student behavior varies with gender composition in the classroom. In addition, Lavy and Schlosser (2011) are unable to identify the relative weight of each mechanism, and instead emphasize the importance of further studies to "distinguish between peer effects that result from changes in individual behavior and peer effects that result from externalities on the classroom environment" (p. 32). By exploiting rich questions from the survey, we find evidence for both individual behaviors and classroom environment, and we further decompose the weight of each channel.

More broadly, our results contribute to the understanding of peer effects in the educational production function. In this literature, the definition of "peers" varies by context, including peer cohorts within the same school (Angrist and Lang 2004; Arcidiacono and Nicholson 2005; Ammermueller and Pischke 2009; Gould et al. 2009), roommates in college dorms (Sacerdote 2001; Zimmerman 2003), and peer groups in military academies (Carrell et al. 2009; Lyle 2007). Peer effects have been found in several dimensions. Along the ability dimension, several studies have found that peers' abilities have positive effects on student achievement (see, e.g., Sacerdote 2001; Zimmerman 2003; Ammermueller and Pischke 2009). Along the racial and social background dimensions, Angrist and Lang (2004) evaluate the effects of the Metco Program, which assigned minority students to schools in affluent suburbs of Boston, and find modest and short-lived peer effects. Gould et al. (2009) show that the overall presence of immigrants in a grade adversely affects students' academic achievement.

It is important to note that peer effects could be context and culture related. In the Chinese context, students usually spend significant amounts of time with the same set of peers, i.e., his or her classmates: They follow the same class schedules and take all lectures together; moreover, peers participate as a group in self-study sessions, extracurricular activities, and field trips. Therefore, peer effects might be particularly pronounced in Chinese schools. Using Chinese school settings, Lu and Anderson (2014) and Hu (2015) also find positive effect on test scores from having female peers; Ding and Lehrer (2007), Park et al. (2015) and Ma and Shi (2014) show significant benefits for academic achievement of having high-ability peers. Our study, by using representative Chinese survey data, confirms the gains in test scores and further documents new findings in the domain of noncognitive outcomes, and also investigate potential channels.

Our findings regarding the mechanisms—classroom environment, teacher behavior, and student effort—in particular echo prior studies that also combine test scores and survey data to investigate how peer effects operate. For instance, Stinebrickner and Stinebrickner (2006) use administrative and survey data from Berea College and find that peers' actions and beliefs may change a student's effort in studying and his/her use of time and beliefs. Booij et al. (2017) manipulate the composition of undergraduate tutorial groups and find improved academic achievements among low- and medium-ability students by switching from ability mixing to tracking groups. Feld and Zölitz (2017) use randomly assigned university class sections and find that low-achieving students are harmed by high-achieving peers. Both of these studies highlight the channel of student interactions and involvement in the classroom, through which peer ability influences student outcomes. Our findings add another piece of evidence that peer interaction and student effort are important avenues by which peer effects operate in classrooms, and we identify that teachers' behavior also change with students' gender composition and has great explanatory power for student outcomes. Understanding and decomposing the mechanisms can shed light on practical and affordable opportunities to improve student outcomes without actually changing the fraction of female students. For example, when there are fewer female students than desired, instructors may consider behaving more patiently and responsibly in their interactions with students. School administrators may also consider other instruments to make the classroom friendlier (e.g., encouraging group activities within and after classes) and boost student motivation (e.g., strengthen incentives) to achieve benefits similar to those of having more female peers.

II. Data and Variables

Our main data source is the 2014 China Education Panel Survey (CEPS). This is a nationally representative survey that includes middle schools from 28 counties and city districts and collects rich information from students, teachers, parents, and school principals using questionnaires.[3]

We exploit a novel question on the survey that asks school principals and teachers how students are assigned to classes, and restrict our estimation sample to schools that randomly assign students. The refined sample includes 8,988 students across 208 classrooms in 67 schools.[4] Table 1 presents summary statistics for our main variables: students' academic

[3]The CEPS is the first and largest nationally representative survey in China to focus on secondary school students and teachers. The survey started in 2013 and applied a stratified sampling design: 28 counties/city districts are chosen nationwide, and four middle schools and multiple (but not all) classrooms within each school are chosen to represent a given county/city district. For a given classroom that is chosen, the survey covers all the students, the head teacher and the main-subject teachers.

[4]The CEPS is a longitudinal survey starting with 7th and 9th graders in the 2013-2014 academic year. We use the first wave for this paper, as the second wave had not been released when we conducted this research. Moreover, the second wave has a low retention rate, as it loses track of all the grade 9 students (who have graduated from middle school) and around 15% of the grade 7 students. Some classrooms are reported to have had changes in student composition since the first wave, which may contaminate our estimation. As

and noncognitive outcomes, their own and their peers' gender, and basic demographics.

We measure peers' gender by the proportion of female peers in the same class. Typically, in middle schools in China, students are assigned to classes at the beginning of the 7th grade and take the same courses throughout their three years in middle school. Peers in the same class interact extensively for both academic and nonacademic purposes. During a regular school day, students remain in the same classroom all day and teachers come to deliver lectures in each subject. They also participate in a variety of exercises and activities together, such as self-study sessions, sports events, and field trips. As shown in Table 1, approximately 49% of the students in the sample are female; not surprisingly, this is also the average proportion of female peers that a given student has. Appendix Figure 1A and 1B plot the original and conditional distribution of the proportion of females respectively, and suggest a sufficient variation of gender composition across classrooms.[5]

Students' academic performance is measured by their test scores (provided by schools' administrative offices) in three core courses of Chinese, mathematics and English. These subjects are compulsory for all middle school students and are the main components of the high school entrance examination (*zhongkao*). Within a school, all teachers of a given course use a similar syllabus and give the same exams during a common testing period.[6] Therefore, test scores in the core courses are consistent and reasonable measures of students' academic achievement for students in the same grade in the same school. In addition, we supplement the test scores with students' self-assessed performance scores. Specifically, they are asked to report whether they have difficulties in learning each subject by using a scale of 1 (*a lot*) to 4 (*not at all*). As shown in Table 1, the sample mean is 81.2 for students' test scores and 2.47 for students' self-assessment scores. It is worth noting that both measures have large standard deviations, which suggests wide dispersion among students. In our regression analyses, to facilitate interpretation, we normalize scores within each subject-grade-school level to obtain a mean of zero and a standard deviation of one.

---

such, we use the first wave of the survey to ensure consistency and accuracy.

[5]In addition, the corresponding $1 - R^2$ from this regression (i.e., that regress peer female proportion on school-grade fixed effect and all control variables) equals to 0.243, suggesting a sufficient variation of gender composition across classrooms.

[6]Exams are graded in a rigorous and consistent manner. During the grading process, each student's name, class, and ID are hidden from the graders. Within a grade in the same school, teachers divide the grading work so that the same question is typically graded by the same teacher by using a consistent rubric.

Measures of noncognitive outcomes are obtained from students' responses to eight survey items. Four questions ask about their mental stress; students are asked to report the frequency, during the previous 7 days, of the four feelings on a scale from 1 (*never*) to 5 (*always*): (1) depressed; (2) blue; (3) unhappy; and (4) life is meaningless. Two questions ask about their general satisfaction in school—students are asked to rate how much they agree with the following statements on a scale from 1 (*strongly agree*) to 4 (*strongly disagree*): (5) "School life is boring." and (6) "I feel confident about my future." Finally, two questions asked about their social acclimation— how frequently they participate in various activities on a scale from 1 (*never*) to 6 (*always*): (7) going to museums, zoos or science parks with classmates from school and (8) going to movies, plays, or sporting events with classmates from school.[7]

We follow Autor et al. (2003) for an aggregation method to obtain the overall effect of peers' gender on students' noncognitive outcomes. Specifically, we first conduct a principal component analysis (PCA) to classify the eight survey items into two categories: (1) the level of mental stress, and (2) the level of social acclimation and satisfaction in school.[8] Next, we create an overall index for each category. This aggregation improves statistical power to detect effects that are consistent across specific outcomes while each individual outcome also has idiosyncratic variation. We estimate the overall effect in the main analysis and report the effects on each individual noncognitive measure in the Appendix. To facilitate interpretation, we normalize each index to have a mean of zero and a standard deviation of one.

[Insert Table 1 here]

The data also contain a rich set of of students' predetermined characteristics, such as their age, ethnicity, local residency status, whether they are the only child, whether they

[7]These measures are comparable to indicators used in the literature. For example, Lavy (2016) measures students' satisfaction and social adjustment in school by survey questions "I feel well-adjusted socially in my class" and "I am satisfied in school".

[8]For the level of mental stress, weights on the four components are: 25.2% on depressed, 26.8% on blue, 26.3% on unhappy, and 21.7% on pessimistic. For the level of social acclimation and general satisfaction, weights on the four components are: 8.6% on feeling fulfilling about school life, 14.2% on confident about the future, 39.2% on public enrichment, and 38% on private recreation.

attended preschool, whether they skipped a grade in primary school, whether they repeated a grade in primary school, their parents' education, and their (baseline) noncognitive measure during primary school.[9] We use these predetermined characteristics to conduct a balancing test, and also include them in our regressions as further controls.

III. Estimation Strategy

To investigate how gender composition affects student outcomes, we implement the linear-in-means model, which has been widely adopted in the literature (e.g., Sacerdote 2001; Guryan et al. 2009; Lu and Anderson 2015). Specifically, we use the following regression model:

$$Y_{ics} = \alpha + \beta_1 Peerfem_{-ics} + \beta_2 Female_{ics} + \phi X_{ics} + \tau W_{cs} + \lambda_{sg} + \varepsilon_{ics}, \qquad (1)$$

where Y_{ics} are the measures of academic and noncognitive outcomes for student i in class c of school s, $Peerfem_{-ics}$ is the proportion of females in student i's class, excluding student i; $Female_{ics}$ indicates whether the student i is female; X_{ics} includes student i's predetermined characteristics and teacher controls; W_{cs} are peers' ability controls, including baseline academic ability for male and female peer separately.[10] λ_{sg} is school-grade fixed effect; and ε_{ics} is the error term. We cluster standard errors at the class level, accounting for correlation in outcomes for students in the same class.

The coefficient of interest is β_1, which captures gender peer effects on students' academic and noncognitive outcomes. An unbiased estimation requires that conditional on all of the controls in the equation (student and teacher characteristics, peers' ability, school-grade fixed effects), our regressor of interest, $Peerfem_{-ics}$ (proportion of female peers), is uncorrelated

[9] Noncognitive measures during primary school include: *Express Opinions Clearly in Primary School* (a self-reported score from 1 (*disagree*) to 5 (*agree*) regarding whether they expressed opinions clearly in primary school), *Respond Quickly in Primary School* (a self-reported score from 1 (*disagree*) to 5 (*agree*) regarding whether they responded quickly in primary school), and *Learn New Stuff Quickly in Primary School* (a self-reported score from 1 (*disagree*) to 5 (*agree*) regarding whether they learned new material quickly in primary school)

[10] Student characteristics are the same set of demographic variables as described in Section II. Teacher controls include teachers' gender, age, years of schooling, experience, professional job title, marital status, and whether they graduated from a normal college. Peers' ability control is measured by whether they repeated grades or skipped grades in primary school. We include male peers' ability and female peers' ability separately.

with the error term ε_{ics}. A possible threat to the identifying assumption is that students may select into classes through unobservable factors, and therefore β_1 may reflect the sorting of students with certain characteristics rather than the effect of peers' gender.

To address this concern, we focus on schools that randomly assign students to classes, in the same spirit as Sacerdote (2001), Zimmerman (2003), Carrell et al. (2009), and Shue (2013). Gong et al. (2018) use the same research setting to investigate the effect of teacher's gender on student outcomes. In the next two subsections, we provide institutional information regarding how students are assigned to classes and perform validity checks on random assignment. Another concern is endogenous school choice. While random class assignment is conducted within schools, students' school choices may not be random. To address possible nonrandom matching between students and schools, we include school-grade fixed effects λ_{sg} in all specifications; therefore, the identification comes from within each school-grade unit and across randomly assigned classrooms. Lastly, students' and teachers' predetermined characteristics, X_{ics}, further improve the balance between classrooms and our estimation efficiency.

A. Class Assignment and Estimation Sample

Our key research question concerns the effect of peers' gender on student outcomes. Understanding how students are assigned to classrooms is vital to our estimation and analysis. Middle schools in China use different strategies to assign students: Some schools have placement exams prior to enrollment and use students' scores and/or rankings to assign them to classrooms. There are also schools that assign students based on whether they are local residents or migrants. In addition, some schools assign students by the primary schools that they attended.

More recently, an increasing number of primary and secondary schools have begun to use random assignment to classrooms. This approach is heavily encouraged by the Ministry of Education and local governments to ensure equal and fair opportunities for all students during their compulsory education years. Schools that adopt random assignment typically rely on a computer program to implement the randomization. Alternatively, when the enrollment size is relatively small and manageable, parents of incoming students are invited to draw lots to determine their child's class placement. After students are assigned to classes, teachers draw lots to determine which classes they will teach and manage.

The CEPS asks school principals and teachers about class assignment, allowing us to identify and focus on schools in which students are randomly assigned to classrooms. In particular, we restrict the estimation sample to schools that satisfy three conditions: First, the school principal reports that students are randomly assigned to classrooms; second, once class assignment is determined prior to the beginning of the 7th grade, the school does not rearrange classrooms for the following three years; and third, all head teachers in a grade report that students in their grade are *not* assigned by test scores.[11] Following this criteria, our refined sample contains 67 schools, 208 classrooms, and 8,988 students, which accounts for approximately 59.8% of the original CEPS sample.[12]

To the extent that students in our estimation sample are all randomly assigned to classrooms and remain with the same peers for the next three years, our sample should mitigate any potential concerns regarding self-selection of students to classrooms and/or peers. Nevertheless, we provide further validity checks in the next subsection.

B. Verifying Random Class Assignment

To verify that students in our sample are randomly assigned to classrooms, we conduct a balancing test among students with varying proportions of female peers. If class assignment is indeed random, students who have different proportions of female peers should be similar in terms of their observed characteristics. We regress student's predetermined characteristics—

[11]Criteria are based on responses in the principal and teacher questionnaire. First, all school principals were asked to report which of the following assignment rules they used to place students: (a) based on pre-enrollment test scores, (b) based on students' residential status, (c) random assignment, or (d) based on other factors. We restrict our sample to schools that use (c). Second, the same principals were asked whether their schools rearrange classrooms for grades 8 and 9; we exclude those that do so. Finally, each head teacher was asked whether students in the grade level taught are assigned by test scores; again, we drop the entire grade if any head teacher answers yes.

[12]Limiting the sample to randomized classrooms may raise concern of external validity of our findings. To this end, we compare school-level characteristics between randomized and nonrandomizes classrooms, such as public or private, the share of rural students, share of local versus migrant students, share of teachers with professional title, the school principal's education background and working experience, and average age of schools (how long they have operated) in the district. Results show very similar statistics across random and nonrandom samples, suggesting that our sample restriction will not severely affect the generality of our findings.

gender, age, minority, local residence, only child, whether attended preschool, whether repeated or skipped a grade in primary school, baseline noncognitive measures during primary school, and parents' education—on the proportion of female peers in his or her classroom.[13]

[Insert Table 2 here]

Table 2 presents results of the balancing test. Column 1 reports the unconditional estimates and column 4 reports the conditional estimates with school-grade fixed effects. The unconditional estimates show that some student predetermined characteristics—such as whether they repeated grade in primary school and their parents' education—vary with the proportion of female peers. Yet most of the differences become much smaller and statistically insignificant after we control for school fixed effects in column 4. The only exceptions are being an only child and predetermined noncognitive measures, but the magnitudes of the differences are very small. For example, while the estimate on *onlychild* is significant, the coefficient implies that a 10 percentage point increase in the fraction of female peers is associated with only a 2.7 percentage points increase in the likelihood that the focal student is the *onlychild* in the family.

It is worth noting that the balancing test for student i's own gender may encounter a potential bias caused by sampling peers without replacement. Because a student cannot be assigned to herself, the sampling of peers is conducted without replacement. In our setting, an immediate problem is that the peers of a female student are chosen from a group with fewer females than the peers of a male student from the same class. Guryan, Kroft and Notowidigdo (2009) discuss this issue, and we follow their proposed solution to further control for the mean of the sampling pool, i.e., the proportion of female peers in the same grade at the same school excluding student i. Results are reported in the lower panel of Table 2; the estimate of $Peerfem_{ics}$ is very small (0.1) and statistically insignificant.

Moreover, we follow literature (Lim and Meer 2017, 2018; Carrell and West 2010; Carrell et al. 2013) to implement a permutation test with a resampling approach. First, for each

[13] The balancing test exhibit fewer observations than the summary statistics (Table 1), due to missing values for some of the student characteristics. We test sample attrition in Section IV and do not find any correlation between attrition and gender composition in the classroom.

classroom within a school, we randomly draw 10,000 synthetic classrooms of the same size from the sample of all students in the school-grade block. We do this for all student characteristics (i.e., gender, age, minority, local residence, only child, whether attended preschool, whether repeated a grade in primary school, whether skipped a grade in primary school, baseline noncognitive measures during primary school, and parents education.). Second, for each student characteristic, and for each classroom within a school-grade, we calculate the average value for each characteristic within a classroom. We then obtain an empirical p-value, that is, the proportion of the 10,000 resampled classrooms with lower statistics for the corresponding characteristic (for example, female student dummy) within the observed classrooms. Last, we find that for all 13 predetermined characteristics including student's own gender, the distribution of p-values are uniformly distributed from a χ^2 test . Overall, we do not find evidence of nonrandom placement of students into classrooms by the predetermined characteristics including baseline academic ability and own gender.

Altogether, results from balancing test suggest that student characteristics are well balanced across classrooms with different fractions of female peers, lending further support to our identification assumption that students in our sample were randomly assigned to classrooms.

IV. Main Results

A. Gender Peer Effects on Academic Performance

We first examine the gender peer effect on students' academic outcomes using regression model (1). Table 3 reports the estimated effects of female peer proportion on students' test scores in core courses. To facilitate interpretation, we normalize test scores by school, grade, and subject to obtain a mean of zero and standard deviation of one. All regressions include subject and school-grade fixed effects.

[Insert Table 3 here]

As shown in Table 3, column 1, the coefficient for the proportion of female peers is positive and statistically significant, which suggests that on average, when a student has more female peers in the class, he or she tends to achieve higher grades. After controlling for predetermined characteristics of the focal student (column 2), the teachers (column 3), and the academic ability of female and male peers (column 4), we find that the effect is consistently positive and statistically significant at the 1% level.

To appreciate the economic significance of the effects, we use the more conservative estimate from column (4), which controls for student and teacher characteristics and peers' ability. The coefficient, 1.019, suggests that a 10-percentage-point (approximately 1.25 standard deviation) increase in the proportion of female classmates raises a student's test score by 10.19% of a standard deviation.

In addition to test score, which is an objective measure of academic performance, we also examine a subjective measure of academic performance, i.e., students' self-assessed scores regarding their learning effectiveness. Appendix Table 1 presents the results. Contrary to the positive effect on test scores, gender peer effects on self-assessment scores are very small and statistically insignificant. Taking the two sets of results together, our findings suggest that having more female classmates improves students' academic performance, but not necessarily their perceived performance and/or confidence in learning.

B. Gender Peer Effects on Noncognitive Outcomes

To examine gender peer effects on students' noncognitive outcomes, we focus on two indices generated from eight items on the student questionnaire (Section II details construction of the indices). One index measures students' mental stress and the other index relates to their social acclimation and general satisfaction in school. Both indices are normalized to have a mean of zero and standard deviation of one. By definition, lower scores for mental stress and higher scores for social acclimation and general satisfaction indicate better outcomes.

Table 4, columns 1 to 3, present the estimated effects on students' mental health. Across the specifications, the estimated impact on mental stress is small in magnitude and statistically insignificant, which suggests that having more female peers does not appear to influence students' mental stress levels.

[Insert Table 4 here]

Table 4, columns 4 to 6, report the estimated effects on students' social acclimation and general satisfaction in school. Overall, we find a positive effect of having more female classmates on students' outcomes along this dimension. The effect remains robust after controlling for student and teacher characteristics, as well as for peers' ability.

We also present estimated effects on the eight noncognitive variables used to construct the indices in Appendix Table 2. Generally, the findings are consistent with the baseline effect. For instance, having more female peers in the classroom renders students to feel that school life is fulfilling and increases social interactions among students. The effects on mental stress, such as feeling blue or unhappy, are very small and statistically not different from zero.

Overall, our results consistently suggest that having a higher proportion of female peers in the classroom improves students' social acclimation and general satisfaction in school.

C. Robustness Checks

In this section, we conduct several empirical exercises to test for random assignment, check whether our results are mainly driven by spillover from female students' academic advantage, or by teachers' differential teaching and grading when more female students are present, examine sample attrition, and explore students' behavioral outcomes that may be related to our noncognitive measures.

A further test for randomization. Our identification strategy relies on the random assignment of students to classrooms. We selected the sample using strong criterion for random assignment, i.e., cross-checking principals' report with those of the respective teachers. A balance test of student baseline characteristics also provides reassurance in this regard. Nevertheless, we conduct a further test to examine whether our regression sample might be contaminated by schools that in fact adopt nonrandom assignment rules, and therefore bias the estimates.

In this empirical exercise, we randomly drop schools from the sample and see whether regression results change dramatically. If our baseline sample contains mostly randomized classrooms, estimates using the reduced sample should not seriously deviate from our baseline estimates. To maintain sufficient sample size, we drop two schools each time, and conduct a total of 2,211 (C^2_{67}) regressions for each outcome variable. Appendix Figure 2 plots the distribution of estimates for test scores, mental stress, and social acclimation and satisfaction separately. We find that all distributions are centered around the respective baseline estimates. Upper and lower bounds also lie in the same direction as the baseline estimates. These findings suggest that our baseline results are unlikely to be severely biased by the possible inclusion of nonrandomized classrooms.

Effects from female students' ability spillover. Our main results document the overall effect of having female peers. One concern is that the effects may come from the spillover of female students' academic ability and performance, given that the literature has established girls' advantage in test scores during primary and middle school.

We address this issue from various angles and provide evidence that the effects are unlikely to be solely driven by girls' academic ability. First, we compare female and male student characteristics and baseline academic ability in Appendix Table 3. Not surprisingly, there are some gender differences, but the magnitudes are economically small and the pattern of academic performance before middle school is mixed: while male students are more likely to repeat grades, they are also more likely to skip grades. Second, when we control for the academic ability of female and male peers, the main results remain similar to the baseline and statistically significant (e.g., Table 3, column 4), which suggests that peers' ability does not explain all of the effects of having more female peers.

Third, we examine the effects on test scores by subject. The premise is that if academic peer effects can explain our findings, then the subject in which female students demonstrate greater advantage should also show stronger effects from having more female peers. In Appendix Table 5, the coefficients on female dummy show a gender gap in test scores for each subject, which suggests that girls lead boys by 0.58 standard deviation in Chinese, 0.539 in English, and only 0.148 in math. In contrast, the benefit of having more female peers—the coefficients on proportion of female peers—is largest for math. In other words, the pattern of academic peer effects goes against that of gender peer effects. As such, it seems unlikely that our findings can be entirely explained by academic peer effects that are correlated with gender.[14]

Teacher assignment, differential teaching and grading. One concern about the effect on test score is that it may not reflect better academic achievement, but rather differential teaching and grading by teachers. For instance, if teachers grade more leniently, or use a different syllabus when there are more female students in the classroom, we would also observe a positive effect on test scores associated with more female peers.

First, we conduct a balancing test in Appendix Table 4 on teachers characteristics, i.e., regressing teacher pre-determined characteristics (gender, education, certificate, experience, title and tenure, etc.) on female peer proportion and controlling for school-grade fixed effect. We find that most estimates are small and statistically insignificant, suggesting no strong correlation between teachers observable characteristics and the percentage of female students in the classroom. We also include these teacher controls in all regressions and our estimates remain stable.

Second, while it is difficult to verify the grading and teaching policies of each school in our sample, we provide some anecdotal evidence of consistent teaching and grading across teachers and classrooms of the same grade in the same school. As part of the compulsory education, middle schools curriculum is designed and enforced by the Ministry of Education at the national level. All schools are required to follow the curriculum plan, and teachers cannot arbitrarily change the courses, difficulty level, teaching hours, or scheduled outlines on their own. Education administrators at the province and city level also strictly enforce the implementation and management of coursework, and usually recommend group preparation for teachers who teach the same subject within a school and grade. Group preparation is organized in regular meetings, in which members receive a detailed plan that includes teaching materials, assignments, and tests, and revise as needed in a collective manner.

[14] Previous studies also discuss this issue, and our findings are consistent with theirs. Lavy and Schlosser (2011) argue that it seems unlikely that all gains in achievement are generated solely by girls ability spillover, as they also find positive gender peer effects in subjects in which girls' achievement is lower than boys. Hoxby and Weingarth (2005) show that even after controlling for peers' lagged achievement, the positive gender peer effects is still robust.

Moreover, teachers of the same subject are required to grade midterm and final exams as a group. In a few cities, schools in the same district organize uniform examinations and grading for the same subject and grade.

Third, we further offer suggestive evidence by examining differences across subjects. Of the three core subjects, i.e., math, Chinese and English, math presumably has more objective components and grading rubrics than the other two. Our premise is that if teachers were to grade differently when there are more female students in the classroom, such a bias is more likely to affect test scores in Chinese and English. As shown in Appendix Table 5, we observe a positive effect of having more female peers in all subjects, and strongest in Math, with an estimated coefficient of 1.299 (significant at the 1% level). This empirical finding appears to contradicts teachers differential grading across student gender composition.

Sample attrition. There are missing values in student outcome variables and predetermined charactersitics. Here, we address the sample attrition problem and check whether peer gender is correlated with the likelihood of missing variables, which could result in biased estimates of gender peer effects. We regress the attrition dummy (whether a variable is missing) on peer gender, student gender, and school-grade fixed effects. As shown in Appendix Table 6, the coefficients on peer gender (proportion of female peers) are all close to zero and statistically insignificant, which indicates that our main results are not driven by sample attrition.

Related behavioral outcomes. In constructing the index for noncognitive outcomes, we focus on four variables for mental stress and four variables for social acclimation and satisfaction in school. We also identify two behavioral outcomes that may relate to students' noncognitive factors: frequency of being late for school and dropping classes. We estimate the effect of having more female peers on these two behavioral outcomes, and find a lower likelihood of being late for school or dropping classes (Appendix Table 7, columns 1-2). These findings are consistent with improved noncognitive outcomes.[15]

[15] Another survey question related to our noncognitive measure it students' level of feeling grief. We excluded it from the construction of mental stress index, as the variable may capture short-run, drastic changes in the environment rather than the student's noncognitive factors. Nevertheless, the estimated effect after including "feeling grief" is similar to the baseline findings.

V. Mechanisms

We find positive and significant effects of having female peers on students' academic and noncognitive outcomes. In this section, we explore potential mechanisms and, in particular, focus on how teacher behavior, classroom environment, and student behavior may change when there are more female students in the classroom. We are aware that it is difficult to exhaust all relevant mechanisms, or rule out the possibility that other mechanisms are in play. Accordingly, we conduct a decomposition analysis, which shows that these channels can explain a great deal of the gender peer effect.

A. Teacher Behavior: Teaching Style and Effort

Here we examine how teacher behavior, such as teaching style and effort exerted to work vary by the gender composition of students in the classroom.

First, it is possible that teachers tailor their teaching style and communication strategies, or provide feedback differently, according to student gender—which, in turn, affects student outcomes. To assess the relevance of this mechanism, we construct an index of teacher feedback using PCA by the following two questions from student survey: (1) "The teacher always praises me." and (2) "The teacher always asks me to answer questions in class." Students are asked to rate to what extent they agree with the statement on a scale from 1 (*strongly disagree*) to 4 (*strongly agree*). Similarly, we use two items on the teacher questionnaire to construct the teaching style index: (1) "I introduce discussion among students in lectures."and (2) "I interact with students in lectures."Teachers are asked to rate how often they adopt these methods in class on a scale from 1 (*never*) to 5 (*always*). We also include two variables on teacher behavior and effort: Parents are asked to rate the head teacher of the classroom based on whether he or she is patient and responsible, and teachers are asked to report how many hours they spend teaching and grading homework.

Table 5 reports the estimation results. Note that the columns use student-level (column 1), parent-level (column 2), and teacher-level (columns 3 and 4) data, and therefore the number of observations varies across specifications. Results show that when there are more female students, teachers are more patient with and responsible for students; they also tend to spend more time on teaching and grading, and adopt a more interactive teaching style—i.e., by inviting students to engage in discussions among themselves and with the teacher during class. The effect on teaching style is large in magnitude but not precisely estimated, possibly due to lack of power. We do not find a significant impact on how teachers give feedback to students. Overall, there is evidence that teachers behave differently when there are more female students in the classroom; they adopt a more interactive teaching style and exert more effort.

[Insert Table 5 here]

B. Classroom Environment

A second possible mechanism is that students' gender composition affects the general environment in the classroom, which influences students' academic achievements, social acclimation, and satisfaction in school. To investigate this potential channel, we construct an index of classroom environment using two survey items from the student questionnaire: (1) "I feel that my classmates are friendly to me." and (2) "I feel that our classroom has a satisfying atmosphere." Students are asked to rate the extent to which they agree with the statements on a scale from 1 (*strongly disagree*) to 4 (*strongly agree*). We normalize responses with a mean of zero and a standard deviation of one and fit regression equation (1).

The results, as shown in Table 6 column 1, demonstrate that when more female peers are present in the classroom, students report a significantly more friendly and satisfying classroom environment. The improved classroom environment may render learning more effective and enjoyable, and thus benefit students' academic achievement. A friendlier environment can also facilitate student interaction and support a feeling of being well adjusted among school peers. Our findings also echo prior studies, such as Booij et al. (2017) and Feld and Zölitz (2017), which find that peer composition affects student interaction and involvement in the classroom.

[Insert Table 6 here]

C. Student Behavior: Learning Effort

Last, we analyze how peers' gender affects student behavior. In particular, peers' gender might affect students' motivation and effort exerted in learning, which in turn influence their academic outcomes. There is evidence that peers affect student effort into studying and his/her use of time (Stinebrickner and Stinebrickner 2006). On the questionnaire, students are asked to report how many hours they spend each week on homework and tutorials. We use this information to investigate how gender composition in the classroom affects students' effort in learning.

Table 6 column 2 reports the estimation results, which suggest that students spend more time on homework and tutorials when they have more female peers. The effects are economically and statistically significant. We also notice that female students tend to exert greater effort than male students. A possible explanation is that when there are more female peers, students feel greater peer pressure to work hard as well.

D. Decomposition of Mechanisms

Our findings show that gender peer effects may work through teachers' teaching style and effort, classroom environment, and student effort, which in turn influence student outcomes. To further understand how much each channel explains gender peer effects, as well as their combined explanatory power, we following Heckman et al. (2013) and Gelbach (2016) to exploit a decomposition method. In particular, we denote m_{icb}^{j} as the mechanism variable j and consider the following estimation specification:

$$m_{ics}^{j} = \alpha_1^j Peerfem_{-ics} + \alpha_2^j Female_{ics} + X_{ics}'\phi + \lambda_s + \eta_g + \varepsilon_{ics} \tag{2}$$

Next, we include all relevant mechanism variables into (1) and consider the following specification:

$$Y_{ics} = \zeta_1 Peerfem_{-ics} + \zeta_2 Female_{ics} + X_{ics}'\phi + \sum_j \gamma^j m_{ics}^{j} + \lambda_s + \eta_g + \epsilon_{ics} \tag{3}$$

Gelbach (2016) shows that

$$\hat{\beta}_1 = \hat{\zeta}_1 + \sum_j \hat{\gamma}^j \hat{\alpha}_1^j. \tag{4}$$

This suggests that mechanism j's component is $\hat{\gamma}^j \hat{\alpha}_1^j$ and the remaining unexplained part is $\hat{\zeta}_1$. For each mechanism, we compute its explanatory power for gender peer effect by $\hat{\gamma}^j \hat{\alpha}_1^j / \hat{\beta}_1$.[16]

Figure 1A plots the estimated decomposition of gender peer effects on academic outcomes into teachers' teaching styles, time spent on teaching and grading, and patience and responsibility; students' effort; classroom environment; and other factors. We find that for the overall effect on test scores, teachers' effort (time spent on teaching and grading) explains approximately 3.9% of the effects, teachers' patience and responsibility explains around 7.9%, classroom environment explains 2.8%, and student effort explains 8.6%. They jointly explain 23.2% of gender peer effects on test scores. The remainder is unexplained by these abovementioned mechanisms.

[Insert Figure 1 here]

Figure 1B presents the decomposition of gender peer effects on social acclimation and general satisfaction. We find that similar to its effect on test scores, teachers' behavior—which includes teaching style, time allocation, and responsibility, in total, explains approximately 10% of the effect on students' social acclimation, classroom environment accounts for 8.9% of the effect and student effort explains a smaller share (2.7%) of the improvement in social acclimation and satisfaction in school.

Overall, we find evidence that having more female students in the classroom motivates teachers to allocate more time to teaching and grading and lecture more interactively, increases student effort, and improves the classroom environment. These channels explain large share of gender peer effects. Our findings are consistent with those of Lavy and Schlosser (2011), who find that an increased proportion of female peers reduces the level of disciplinary problems, improves inter-student and teacher-student relationships, and reduces teacher fa-

[16]Note that if there are some unmeasured mechanisms correlated with the observed mechanisms and/or if the observed mechanisms are measured with error, γ^j might be biased. Therefore, the decomposition results should be interpreted with caution.

tigue. Our analysis innovates by measuring not only students' perceptions of the school environment, but also individual-level behaviors such as students' effort in learning, teachers' time spent on working and teaching style.

Understanding the mechanisms of gender peer effects is important for policy design. To the extent that the number of female students in a school is fixed, the benefit of having more female peers in one class could be offset by the cost of having fewer female peers in another class. Understanding the sources of gender peer effects sheds light on more practical and affordable opportunities—in particular, teacher behavior, classroom environment, and student effort—to improve student outcomes. For instance, when there are fewer female students than desired, instructors may consider behaving more patiently and responsibly toward students and adding more discussion during lectures. In assigning teachers to classrooms, principals can take student gender composition into account, in that teachers who tend to be more patient and to actively engage students may be able to compensate for the lower proportion of female students in the classroom. In the same vein, head teachers may consider other instruments to make the classroom friendlier (e.g., encourage group activities within and after classes) and boost student motivation (e.g., strengthen incentives) to achieve benefits similar to those of having more female peers.

VI. Heterogeneity in Gender Peer Effects

Finally, we explore how gender peer effects vary by student characteristics—i.e., own gender and parents' education, and teacher gender and experience. We include interaction terms between female peer proportion and the corresponding variable and report results in Table 7.

We find differential peer gender effects between female and male students. While a higher proportion of female peers improves test scores for both female and male students, the effect is much stronger for male students (Table 7, column 1). The effect size suggests that a 10 percentage point increase in the proportion of female students increases average test scores of boys and girls by 14.0% and 6.5% of a standard deviation, respectively. For noncognitive outcomes, girls appear to benefit more from having female peers than boys: The effect on boys mental stress is not statistically significant, while girls are less likely to suffer from mental stress (Table 7, column 5). The effect on social acclimation and general satisfaction is positive for both female and male students; female students tend to benefit more, although the difference is not statistically significant (Table 7, column 9). This is consistent with Lavy and Schlosser (2011), who find that when more female peers are present, girls tend to report better inter-student relationships and social adjustment in class.

The heterogeneity in the effects on test scores may reflect certain differences in the returns to increased effort of study. As shown in Appendix Tables 8A and 8B, when we decompose the mechanisms of peer effects for male and female students separately, a major difference is the channel through effort: When more female peers are present in the classroom, both male and female students allocate more time to study, but the increased effort translates into better academic performance only for boys. A possible explanation is that compared with male students, female students spend more time in the baseline (on average 27.5 hours on homework and tutorials per week, versus 24 hours for male students in our sample), and the marginal benefit from additional effort may diminish.

There is also evidence of gender peer effects interacting with teacher's gender. Table 7, column 3 suggests that while there are overall positive effects from female peers, the gain is larger when students have a male teacher. This suggests that having a female teacher might be a substitute for having more female peers. We do not find heterogeneous effects by teachers' experience or parents' education on either students' test scores or their noncognitive outcomes.

[Insert Table 7 here]

VII. Conclusion

This paper uses a nationally representative survey of middle school students to investigate gender peer effects on students' academic performance and noncognitive outcomes. By employing information about classroom assignment within schools, we are able to restrict the sample to schools that randomly assign students to classrooms and therefore estimate the causal relationship between peer gender composition and student outcomes.

Our results show that having a higher proportion of female peers in the classroom significantly raises students' test scores and improves social acclimation and general satisfaction in school. By exploring the potential mechanisms through which peers' gender plays a role, we find evidence of teachers behaving more patiently and responsibly towards students and spending more time on teaching and grading, improved classroom environment, and greater students effort exerted in learning. These mechanisms collectively explain a significant fraction of the identified peer gender effects.

Our findings make several contributions to the literature and have some policy implications. First, while most previous literature focuses on the effects of the school environment on students' academic outcomes, our study provides more evidence on the impact on students' noncognitive outcomes, which are important factors in explaining academic achievement, labor market success, and other significant life outcomes (Heckman and Rubinstein 2001; Heckman et al. 2013; Segal 2013; Bertrand and Pan 2013). Second, we provide rich evidence for the mechanisms that drive these effects: teacher behavior and teaching style, classroom environment, and student effort. Understanding these mechanisms sheds light on educational policies designed to improve student outcomes. For example, our decomposition exercise shows that teacher behavior and classroom environment explain a considerable amount of gender peer effects on test scores. One implication could be that to compensate for the small share of female students in certain classes, schools could assign teachers by considering their work style, attitudes, and motivation/effort; also, improving the classroom environment might achieve outcomes similar to those of having more female peers.

References

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

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

Arcidiacono, Peter and Sean Nicholson, "Peer Effects in Medical School," *Journal of Public Economics*, 2005, *89* (2), 327–350.

Autor, David H, Frank Levy, and Richard J Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly journal of economics*, 2003, *118* (4), 1279–1333.

Bertrand, Marianne and Jessica Pan, "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior," *American Economic Journal: Applied Economics*, 2013, *5* (1), 32–64.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes, "Under Pressure? The Effect of Peers on Outcomes of Young Adults," *Journal of Labor Economics*, 2013, *31* (1), 119–153.

Booij, Adam S, Edwin Leuven, and Hessel Oosterbeek, "Ability Peer Effects in University: Evidence from a Randomized Experiment," *The review of economic studies*, 2017, *84* (2), 547–578.

Carrell, Scott E and James E West, "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," *Journal of Political Economy*, 2010, *118* (3), 409–432.

— , **Bruce I Sacerdote, and James E West**, "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 2013, *81* (3), 855–882.

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

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR," *The Quarterly Journal of Economics*, 2011, *126* (4), 1593–1660.

Ding, Weili and Steven F Lehrer, "Do Peers Affect Student Achievement in China's Secondary Schools?," *The Review of Economics and Statistics*, 2007, *89* (2), 300–312.

Epple, Dennis and Richard Romano, "Peer Effects in Education: A Survey of the Theory and Evidence," *Handbook of Social Economics*, 2011, *1* (11), 1053–1163.

Feld, Jan and Ulf Zölitz, "Understanding Peer Effects: on the Nature, Estimation, and Channels of Peer Effects," *Journal of Labor Economics*, 2017, *35* (2), 387–428.

Gelbach, Jonah B, "When do Covariates Matter? And Which Ones, and How Much?," *Journal of Labor Economics*, 2016, *34* (2), 509–543.

Gong, Jie, Yi Lu, and Hong Song, "The Effect of Teacher Gender on Students Academic and Noncognitive Outcomes," *Journal of Labor Economics*, 2018, *36* (3), 743–778.

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

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

Heckman, James J and Yona Rubinstein, "The Importance of Noncognitive Skills: Lessons from the GED Testing Program," *The American Economic Review*, 2001, *91* (2), 145–149.

Heckman, James, Rodrigo Pinto, and Peter Savelyev, "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," *The American Economic Review*, 2013, *103* (6), 1–35.

Hoxby, Caroline, "Peer Effects in the Classroom: Learning from Gender and Race Variation," Technical Report, National Bureau of Economic Research 2000.

Hoxby, Caroline M and Gretchen Weingarth, "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects," Technical Report, Working paper 2005.

Hu, Feng, "Do Girl Peers Improve Your Academic Performance?," *Economics Letters*, 2015, *137*, 54–58.

Jencks, Christopher et al., "Who Gets Ahead? The Determinants of Economic Success in America," 1979.

Kremer, Michael, Esther Duflo, and Pascaline Dupas, "Peer Effects, Teacher Incentives, and the Impact of Tracking," *American Economic Review*, 2011, *101* (5).

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

Lim, Jaegeum and Jonathan Meer, "The Impact of Teacher–Student Gender Matches Random Assignment Evidence from South Korea," *Journal of Human Resources*, 2017, *52* (4), 979–997.

_ and _ , "How do Peers Influence BMI? Evidence from Randomly Assigned Classrooms in South Korea," *Social Science & Medicine*, 2018, *197*, 17–23.

Lu, Fangwen and Michael L Anderson, "Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups," *Journal of Labor Economics*, 2014, *33* (1), 91–122.

_ **and** _ , "Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups," *Journal of Labor Economics*, 2015, *33* (1), 91–122.

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

Ma, Mingming and Xinzheng Shi, "Magnet Classes and Educational Performance: Evidence from China," *Economic Development and Cultural Change*, 2014, *62* (3), 537–566.

Manski, Charles F, "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542.

Park, Albert, Xinzheng Shi, Chang tai Hsieh, and Xuehui An, "Magnet High Schools and Academic Performance in China: A Regression Discontinuity Design," *Journal of Comparative Economics*, 2015, *43* (4), 825–843.

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

Segal, Carmit, "Misbehavior, Education, and Labor Market Outcomes," *Journal of the European Economic Association*, 2013, *11* (4), 743–779.

Shue, Kelly, "Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers," *Review of Financial Studies*, 2013, *26* (6), 1401–1442.

Stinebrickner, Ralph and Todd R Stinebrickner, "What Can be Learned about Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds," *Journal of public Economics*, 2006, *90* (8-9), 1435–1454.

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

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

Tables and Figures

Table 1. Summary Statistics

	Mean	SD	Obs
Outcome variables:			
Academic outcomes:	(1)	(2)	(3)
Test score	81.21	28.4	26209
Self-assessment	2.47	0.92	26746
Non-cognitive outcomes:			
Index 1: mental stress	4.13	1.72	8682
Depressed	2.24	1	8772
Blue	1.98	1.06	8743
Unhappy	2.28	1.05	8762
Pessimistic	1.75	1.07	8734
Index 2: social acclimation and general satisfaction	4.99	1.41	8479
School life is fulfilling	3.38	0.86	8852
Confident about future	3.26	0.72	8924
Social activity: Public enrichment	2.02	1.04	8686
Social activity: Private recreation	2.44	1.28	8653
Regressor of interest:			
Female peer	0.49	0.08	8910
Predetermined characteristics:			
Female	0.49	0.5	8910
Age	13.94	1.35	8815
Minority	0.11	0.31	8968
Local residence	0.8	0.4	8811
Only child	0.51	0.5	8986
Preschool attendance	0.82	0.39	8912
Repeater	0.11	0.32	8988
Skip grade	0.02	0.12	8966
Pre noncognitive measure	3.06	0.67	8485
Mother education (college attendance)	0.18	0.38	8988
Father education (college attendance)	0.21	0.41	8988

Table 2. Balancing Test for Predetermined Characteristics

	Unconditional test			Conditional test		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Coefficient	SE	Obs	Coefficient	SE	Obs
Age	-0.195	(1.011)	8,742	0.178	(0.125)	8,742
Minority	-0.183	(0.307)	8,891	-0.007	(0.037)	8,891
Urban residence	0.299	(0.204)	8,735	0.003	(0.116)	8,735
Only child	0.579**	(0.238)	8,908	0.277***	(0.104)	8,908
Pre-school attendance	0.052	(0.145)	8,835	0.148	(0.099)	8,835
Repeat grade in primary school	-0.346***	(0.130)	8,910	0.017	(0.062)	8,910
Skip grade in primary school	-0.023	(0.027)	8,889	0.001	(0.018)	8,889
Non-cognitive measure	0.243**	(0.118)	8,415	0.899***	(0.262)	8,415
Mother education	0.304**	(0.140)	8,910	0.009	(0.088)	8,910
Father education	0.368**	(0.155)	8,910	0.065	(0.092)	8,910
Test balance of own gender using methods by Guryan, Kroft and Notowidigdo (2009)						
Female (student's own gender)				0.010	(0.031)	8,910

Note: Each cell represents a separate regression which regress the corresponding pre-determined characteristic above on peer female proportion. Conditional estimates are obtained from regression that include school-grade fixed effects. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

Table 3. Gender Peer Effect on Test Score

	(1)	(2)	(3)	(4)
Proportion female peers	1.259***	1.119***	1.101***	1.019***
	(0.332)	(0.310)	(0.310)	(0.278)
Female student	0.436***	0.426***	0.425***	0.420***
	(0.023)	(0.023)	(0.023)	(0.023)
Subject fixed effects	Yes	Yes	Yes	Yes
School-grade fixed effects	Yes	Yes	Yes	Yes
Student controls	No	Yes	Yes	Yes
Teacher controls	No	No	Yes	Yes
Peer ability controls	No	No	No	Yes
Observations	22,405	22,405	22,405	22,405
R-squared	0.051	0.084	0.085	0.090

Notes: Test score is normalized by subject, grade and school, to obtain a mean of zero and standard deviation of one. Student controls include student's age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten, repeated a grade and skipped a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

Table 4. Gender Peer Effect on Noncognitive Measure

	Mental stress index			Social acclimation and general satisfaction index		
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion female peers	0.028	0.045	0.052	0.817***	0.758***	0.767***
	(0.293)	(0.296)	(0.288)	(0.242)	(0.236)	(0.233)
Female student	0.028	0.028	0.028	0.050**	0.049**	0.049**
	(0.026)	(0.025)	(0.025)	(0.021)	(0.020)	(0.020)
School-grade fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	No	Yes	Yes	No	Yes	Yes
Peer ability controls	No	No	Yes	No	No	Yes
Observations	7,616	7,616	7,616	7,418	7,418	7,418
R-squared	0.089	0.090	0.091	0.313	0.314	0.314

Notes: Student controls include student's age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten, and repeated a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

Table 5. Mechanism: Teacher Behaviors

	Praise and question	"Teacher is responsible"	Discussion in lecture	Time spent in teaching and grading (log)
	(1)	(2)	(3)	(4)
Proportion female peers	-0.085	1.004***	0.897	0.379*
	(0.069)	(0.278)	(0.583)	(0.219)
Female student	0.017**	0.031	NA	NA
	(0.007)	(0.024)		
School-grade FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes
Peer ability controls	Yes	Yes	NA	NA
Observations	7,913	7,437	555	459
R-squared	0.205	0.155	0.333	0.623

Notes: Columns 1 and 2 are analyzed at student level, columns 3 and 4 are analyzed at teacher level. Student controls in columns 3 and 4 are aggregated in classroom level. "Proportion female peers" in columns 3 and 4 are measured by the proportion of female in the class because the analysis is at the classroom level. Student controls include student's age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten, and repeated a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

Table 6. Mechanism: Classroom Environment and Student Effort

	Classroom environment	Student time allocated to study per week (log)
	(1)	(2)
Proportion female peers	0.774*	1.165***
	(0.440)	(0.183)
Female student	0.185***	0.150***
	(0.024)	(0.020)
Subject fixed effects	Yes	Yes
School-grade fixed effects	Yes	Yes
Student controls	Yes	Yes
Teacher controls	Yes	Yes
Peer ability controls	Yes	Yes
Observations	7,731	7,019
R-squared	0.157	0.242

Notes: The analysis is conducted at the student level. Student controls include student's age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten, and repeated a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

Table 7. Heterogeneous Effects

	Test score				Mental stress				Social acclimation and general satisfaction			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Proportion female peers *Female student	-0.750***				-0.676**				0.256			
	(0.269)				(0.318)				(0.206)			
Proportion female peers *Mother low education		0.040				-0.014				0.073***		
		(0.042)				(0.012)				(0.010)		
Proportion female peers *Female teacher			-0.369**				-0.180				0.151	
			(0.187)				(0.505)				(0.400)	
Proportion female peers *Teaching experience				-0.008				0.001				0.005
				(0.011)				(0.004)				(0.003)
Proportion female peers	1.399***	0.613	1.279***	1.158***	0.390	0.182	0.182	0.151	0.635**	0.034	0.598*	0.621**
	(0.328)	(0.491)	(0.245)	(0.370)	(0.345)	(0.299)	(0.413)	(0.290)	(0.244)	(0.257)	(0.324)	(0.244)
Female student	0.787***	0.417***	0.420***	0.420***	0.359**	0.029	0.028	0.023	-0.076	0.046**	0.048**	0.044**
	(0.136)	(0.023)	(0.023)	(0.023)	(0.157)	(0.025)	(0.025)	(0.026)	(0.102)	(0.020)	(0.020)	(0.021)
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-grade fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer ability controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,405	22,373	22,405	22,405	7,616	7,608	7,616	7,308	7,418	7,409	7,418	7,112
R-squared	0.091	0.092	0.090	0.090	0.091	0.091	0.091	0.094	0.314	0.320	0.316	0.310

Notes: Student controls include student's age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten, and repeated a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

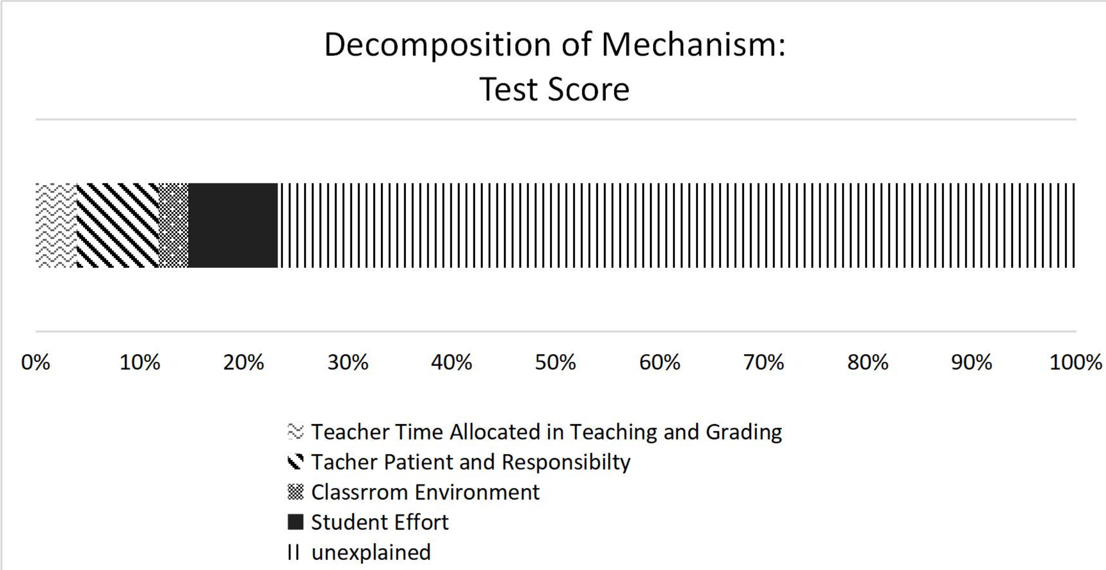


Figure 1A. Decomposition of mechanism behind gender peer effects on test scores

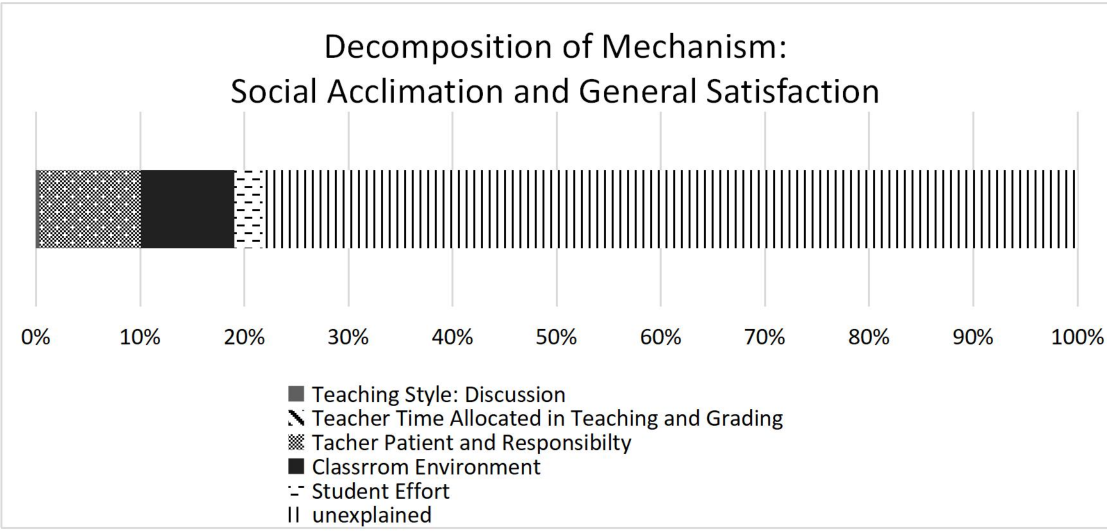


Figure 1B. Decomposition of Mechanism behind Gender Peer Effects on Noncognitive Outcomes

Appendix

Appendix Table 1: Gender Peer Effect on Academic Self-Assessment Scores

	(1)	(2)	(3)	(4)
Proportion female peers	0.325	0.153	0.134	0.136
	(0.205)	(0.183)	(0.181)	(0.179)
Female student	0.179***	0.188***	0.188***	0.187***
	(0.019)	(0.019)	(0.019)	(0.019)
Subject FE	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes
Student controls	No	Yes	Yes	Yes
Teacher controls	No	No	Yes	Yes
Peer ability controls	No	No	No	Yes
Observations	22,914	22,914	22,914	22,914
R-squared	0.009	0.049	0.050	0.050

Notes: The dependent variable is students' self-assessment of academic performance. Specifically, students were asked to report whether they have difficulty in learning each subject on a scale from 1 (a lot) to 4 (not at all). The rating is normalized by subject, grade and school, to obtain a mean of zero and standard deviation of one. Student controls include student's age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten, repeated a grade and skipped a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

Appendix Table 2. Robustness Check: Gender Peer Effects on Individual Noncognitive Factors

	Mental stress				Social acclimation and general satisfaction			
	Depressed	Blue	Unhappy	Pessimistic	Fulfilling of life	Confident abt future	Public enrichment	Private recreation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proportion	0.281	0.038	-0.042	-0.100	0.537**	-0.072	0.341	0.938***
female peers	(0.260)	(0.234)	(0.244)	(0.277)	(0.267)	(0.178)	(0.288)	(0.236)
Female student	0.156***	-0.031	0.042*	-0.073***	0.092***	-0.030	0.048**	0.028
	(0.026)	(0.025)	(0.024)	(0.022)	(0.025)	(0.023)	(0.021)	(0.022)
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer ability controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,616	7,616	7,616	7,616	7,418	7,418	7,418	7,418
R-squared	0.079	0.075	0.084	0.060	0.098	0.139	0.238	0.239

Notes: Dependent variables are normalized to have a mean of zero and standard deviation of one. Student controls include student age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten and repeated a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

Appendix Table 3. Comparison of Student Characteristics by Gender

Predetermined characteristics:	Male			Female			Difference	
	Mean	SD	Obs	Mean	SD	Obs	Difference	P-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	13.99	1.35	4462	13.88	1.34	4280	0.11	0.000***
Minority	0.11	0.31	4558	0.11	0.32	4333	0	0.367
Local residence	0.8	0.4	4487	0.81	0.39	4248	-0.01	0.306
Only child	0.54	0.5	4570	0.48	0.5	4338	0.06	0.000***
Preschool attendance	0.81	0.39	4519	0.82	0.38	4316	-0.01	0.151
Repeater	0.13	0.34	4572	0.1	0.29	4338	0.03	0.000***
Skip grade	0.02	0.14	4554	0.01	0.11	4335	0.01	0.006***
Pre noncognitive measure	3.08	0.69	4277	3.05	0.65	4138	0.03	0.04**
Mother college attendance	0.18	0.38	4572	0.18	0.39	4338	-0.17	0.306
Father college attendance	0.2	0.4	4572	0.22	0.41	4338	-0.02	0.033**

Notes: this table present the summary statistic of pre-determined characteristics for male and female students separately. Column 7 presents the difference of coefficients, and column 8 presents the corresponding p-value.

Appendix Table 4. Balancing Test of Teacher Characteristics

VARIABLES	Coefficient	SE	Observations
Head teacher	(1)	(2)	(3)
Female	0.737	(0.642)	208
Education	0.139	(0.414)	207
Graduation from normal college	-0.025	(0.305)	207
Professional title	-0.052	(1.100)	206
Experience	0.455	(1.206)	205
Tenure status	-0.251	(0.290)	208
Teaching main subject	-0.423	(0.639)	208
Subject teacher			
Female	0.009	(0.159)	617
Education	0.311	(0.235)	613
Graduation from normal college	0.066	(0.175)	614
Professional title	0.372	(0.515)	614
Experience	-0.022	(0.390)	611
Tenure status	0.138*	(0.081)	607

Notes: Each cell represents a separate regression which regress the corresponding teacher characteristic on peer female proportion, controlling for school-grade FE. The specifications for subject teacher also control for subject fixed effects. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

Appendix Table 5. Gender Peer Effect by Subject

	Chinese	Math	English
	(1)	(2)	(3)
Proportion female peers	0.808***	1.299***	1.207***
	(0.303)	(0.392)	(0.298)
Female student	0.580***	0.148***	0.539***
	(0.026)	(0.031)	(0.026)
School-grade fixed effects	Yes	Yes	Yes
Student controls	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes
Peer ability controls	Yes	Yes	Yes
Observations	7,400	7,312	7,305
R-squared	0.129	0.064	0.132

Notes: The table presents the gender peer effect on test score by subject. Test score is normalized by subject, grade and school, to obtain a mean of zero and standard deviation of one. Student controls include student's age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten, and repeated a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

Appendix Table 6. Robustness Check: Sample Attrition

	Test score	Self-assessment	Depressed	Blue	Unhappy	Pessimistic	Fulfilling of life	Confident abt future	Public enrichment	Private recreation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Prop female peers	0.046	-0.018	-0.021	-0.016	-0.022	-0.012	-0.045	-0.045	-0.055	-0.046
	(0.065)	(0.058)	(0.067)	(0.068)	(0.067)	(0.068)	(0.057)	(0.056)	(0.065)	(0.062)
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26,730	26,730	8,910	8,910	8,910	8,910	8,910	8,910	8,910	8,910
R-squared	0.060	0.042	0.044	0.047	0.043	0.044	0.044	0.046	0.047	0.047

Notes: the dependent variable in each cell is a dummy which equals to one if the corresponding measurement is missing. All regressions include school-grade fixed effects. Regressions in column 1 and column 2 further include subject fixed effects.

Appendix Table 7. Robustness Check: Related Behavioral Outcome Variables

	Late for school	Drop classes	Feel Grief
	(1)	(2)	(3)
Proportion female peer	-0.325*	-0.251	0.084
	(0.175)	(0.173)	(0.347)
Female student	-0.094***	-0.100***	0.063**
	(0.029)	(0.025)	(0.026)
Subject FE	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes
Student controls	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes
Observations	7,758	7,752	7,671
R-squared	0.090	0.085	0.077

Notes: Dependent variables are normalized to have a mean of zero and standard deviation of one. Student controls include student age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten, and repeated a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

Appendix Table 8A: Mechanism Analysis by Gender: Female Peer Proportion and Channels

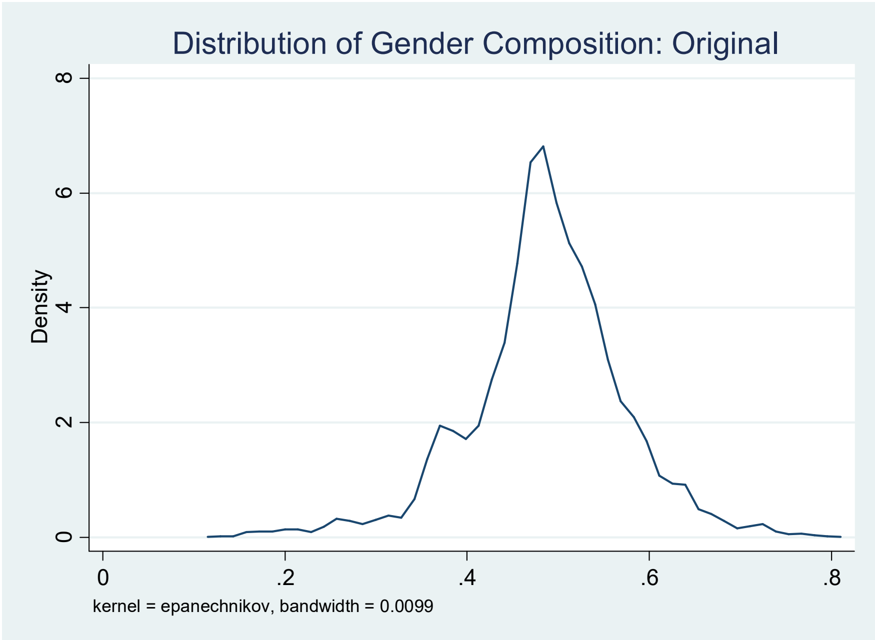
	Discussion in lecture	Teacher effort	"Teacher is responsible"	Classroom environment	Student effort
Female sample					
	(1)	(2)	(3)	(4)	(5)
Proportion female	0.897	0.379*	0.675*	0.163	1.511***
peers	(0.583)	(0.219)	(0.357)	(0.328)	(0.308)
Observations	555	459	11,055	11,305	10,632
R-squared	0.333	0.623	0.177	0.191	0.256
Male sample					
Proportion female	0.897	0.379*	0.762*	0.298	1.174***
peers	(0.583)	(0.219)	(0.422)	(0.348)	(0.230)
Observations	555	459	10,909	11,490	10,862
R-squared	0.333	0.623	0.164	0.149	0.201
School-grade FE	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes
Peer ability controls	NA	NA	Yes	Yes	Yes

Notes: Estimates are obtained from specifications that regress each channel on proportion female peers, with controlling for school-grade FE and other controls. Estimates for "Discussion in lecture" and "Teacher effort" are same for male and female sample because they are obtained from teacher level analysis. Student controls include student age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten, and repeated a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

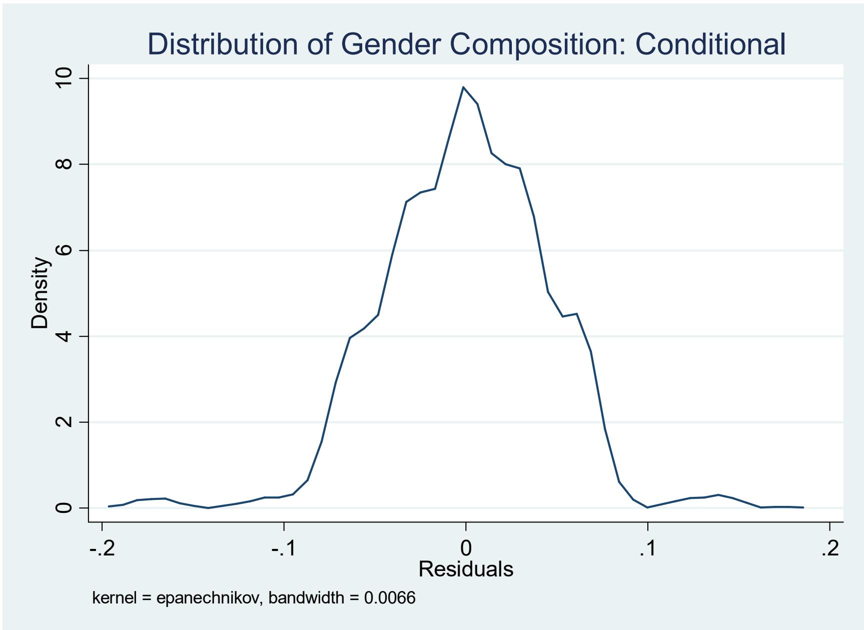
Appendix Table 8B: The Percentage that Each Channel Attributed to Test Score, by Gender

	Female	Male
	(1)	(2)
Teaching style:	-0.014	-0.008
Discussion in lecture	(0.012)	(0.015)
Teacher effort	0.050	0.140***
	(0.039)	(0.044)
Teacher is responsible	0.109***	0.106***
	(0.018)	(0.021)
Classroom environment	0.096***	0.135***
	(0.022)	(0.022)
Student effort	-0.000	0.110***
	(0.022)	(0.026)
Proportion female peers	0.629**	1.403***
	(0.301)	(0.379)
School-grade FE	Yes	Yes
Student controls	Yes	Yes
Teacher controls	Yes	Yes
Peer ability controls	Yes	Yes
Observations	7,430	7,299
R-squared	0.111	0.105

Notes: The estimates are obtained from specification that regress test score on above channels, with controlling for proportion female peer, school-grade FE and other controls. Student controls include student age, baseline noncognitive measurements, mother's education, father's education, and dummy variables indicating minority, local residence, rural residence, only child in family, attended kindergarten, and repeated a grade in primary school. Teacher controls include gender, age, years of schooling, experience, professional job title, and dummy variables indicating marital status and graduated from a normal college. Peer ability controls include male and female peers' average baseline academic performance in primary school, including whether repeated a grade and whether skipped grades. Standard errors are clustered at class level and reported in parentheses. ***significant at the 1% level, **5% level, *10% level.

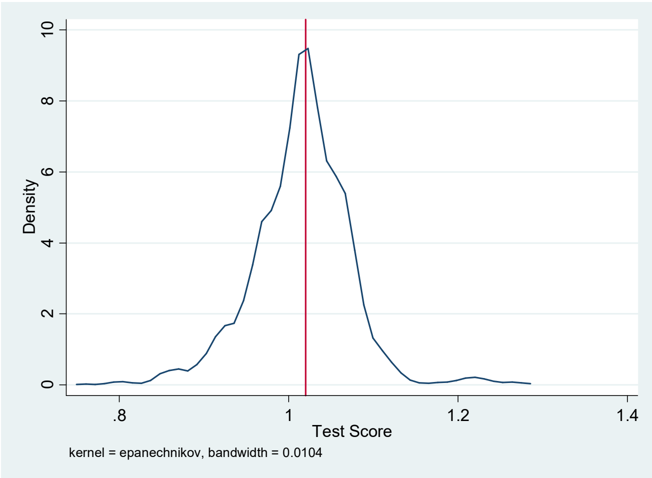


Appendix Figure 1A. Original Distribution of Gender Composition

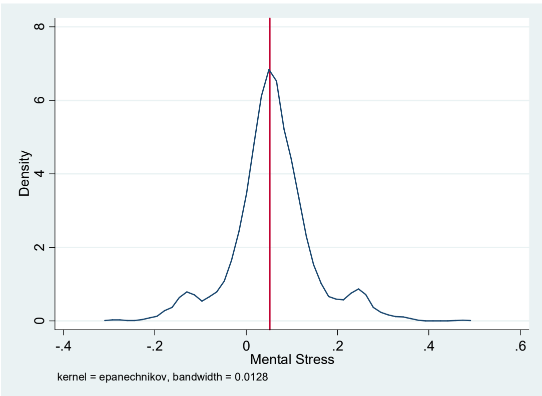


Appendix Figure 1B. Conditional Distribution of Gender Composition

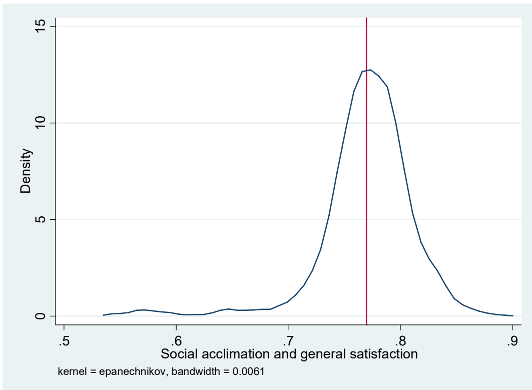
Notes: The figure plots the original and conditional distribution of female peer in class. The conditional distribution is the distribution of residuals that obtained from regress female peer proportion on school-grade FE and controls. The corresponding (1-R^2) from this regression (i.e., that regress peer female proportion on school-grade fixed effect and all control variables) equals to 0.243.



Appendix Figure 2A



Appendix Figure 2B



Appendix Figure 2C

Appendix Figure 2. Robustness check:
Distribution of the Coefficients after Randomly Dropping Two Schools

Notes: The figures plot the distributions of the coefficients from 2,211 regressions that each time randomly drop two schools from the sample. The vertical lines indicate our baseline estimates for the respective student outcome.