Friendship and Female Education: Evidence from a Field Experiment in Bangladeshi Primary Schools*†

Youjin Hahn Asadul Islam Eleonora Patacchini Yves Zenou

Abstract

We randomly assigned 115 primary schools in Bangladesh to one of two settings: children studying in groups with friends and children studying in groups with peers. The groups consisted of four people with similar average cognitive abilities and household characteristics. While the achievement of male students was not affected by the group assignment, low-ability females with friends outperformed low-ability females working with peers by roughly 0.4 standard deviations of the test score distribution. This is not due to the fact that friends tend to be of the same gender or to a higher frequency of interactions among friends.

JEL Classifications: I25, J16, O12.

Keywords: Social interactions, education, gender, learning, friendship.

*Corresponding author: Yves Zenou, Monash University, Department of Economics, Caulfield VIC 3145, Australia. Email: yves.zenou@monash.edu

†We thank the editor, two anonymous referees, Jim Berry, Jaesung Choi, Booyuel Kim, Vesall Nourani, Sangyoon Park, the seminar participants at Monash University, Queensland University of Technology, University of Queensland, University of New South Wales, Yonsei University, the 24th Annual SJE International Symposium: Human Capital and Economic Development, Korean Labor Economic Association Meeting, Korean International Economic Association Winter Meeting, Hanyang University, and the KDI School of Public Policy and Management for valuable comments. We are indebted to Behrooz Hassani-Mahmooei for his help in the early stages of the project. We are grateful for funding support from Monash University, Yonsei University and AusAID (DFAT). We thank the Department of Primary Education in the Ministry of Education of Bangladesh for its support in conducting this project. Angela Cools, Foez Mojumdar, Mujahid Islam, and Mahbub Sarkar provided excellent research assistance.

Methods to improve educational outcomes are of key interest to policy makers, especially in developing countries. Over the last decade, many developing countries have made substantial improvements in primary education. For example, many have achieved gender parity in enrollment, reduced dropout, and/or increased completion of the educational cycle (see, e.g., Andrabi *et al.*, 2007; UWEZO, 2014; Banerjee *et al.*, 2016). However, they show persistently low levels of achievement and a large gender gap in educational performance remains. In response to these challenges, many experimental studies have considered interventions to improve learning in developing countries (see Glewwe and Muralidharan, 2016; Ganimian and Murnane, 2016; and Evans and Popova, 2016, for detailed reviews). For instance, interventions targeting teaching to the right learning levels of students have found large and positive effects on learning (Banerjee *et al.*, 2007; Duflo *et al.*, 2011; Banerjee *et al.*, 2016;).

This paper provides experimental evidence of the effects of an alternative method that may aid learning. In particular, we investigate whether or not it is important to have friends in study groups. We randomly assigned 115 primary schools in Bangladesh to one of two settings: children studying in groups with friends and children studying in groups with peers. The experiment involved all grade-four students in these schools, in total more than 4,600 students. At the beginning of the experiment, each student took a math test to measure his or her cognitive ability. The student was then allocated to work on the math assignment in the setting assigned to his or her own school. The groups with peers and groups with friends each consisted of four students, balanced by average cognitive ability and household characteristics. After working for a week in his or her given setting, each student individually took another math test similar in content to the math assignment. Students were then given prizes based on their performance at the final testing stage. The prize structure across the treatments was the same and followed a tournament scheme based on the individual test scores.

Our analysis revealed that the effects on learning outcomes of combining friends into small study groups with common objectives depended on the gender and abilities of the children. In particular, this intervention significantly improved the individual performance of low-ability females, indicating that studying with peer friends may help close the gender gap in educational performance. The topic of how to increase learning levels among primary-aged children and how to close the gender gap is at the forefront of the political debate in many development countries. In Bangladesh, since the mid-1990s, the government has introduced many education policies targeting female children, including compulsory free primary education and a female stipend program in secondary schools in rural areas. These policies have led to more gender parity in enrolment in both primary and lower-secondary levels (Begum *et al.*, 2017; Hahn *et al.*, 2018), but female students still lag behind their male

counterpart in learning outcomes (Asadullah and Chaudhury, 2015).

The main challenge in interpreting our results is to disentangle the effects of studying with friends from the effects of studying with same-gender peers, since a student's reported friends are mostly of the same gender. Given the traditionally strong differences in the roles of males and females in Bangladesh, one could imagine that same-gender groups have different dynamics than different-gender groups. We addressed this issue by examining whether the outcomes of students allocated to study groups with same-gender peers differed from those of students who studied with same-gender friends. Similarly, we compared the outcomes of students in mixed-gender peer groups and mixed-gender friendship groups. Our evidence shows large gains for low-ability females from working with friends, regardless of the gender composition of the study group. This suggests that it is really friendship that matters and not gender composition.

Our evidence is consistent with the sociological literature, which suggests that females' improvements from group work may be driven by social indispensability (the feeling that people, especially friends, care about the impact of their own performance on the group outcome) (see, e.g. Weber *et al.*, 2009). In addition, psychological research suggests that women may care more than men about collective outcomes, and thus may be more likely to exert more effort when they work in a group than when they work alone (Karau and Williams, 1993). The gains of females in cooperative environments are highest in cohesive groups and when groups have stronger agreement (Karau and Hart, 1998).

Our analysis contributes to the economic development literature on the gender gap. Although the enrolment rates of girls at the primary level have increased rapidly in most developing countries (Banerjee *et al.*, 2016), including Bangladesh (Asian Development Bank, 2017), the gender gap in academic performance is still very large (Bharadwaj *et al.*, 2016). The demand-side interventions in developing countries aimed at improving female educational attainment primarily involve providing conditional cash transfers to households. These cash transfer programs have a generally positive impact on female education (for a review, see Fiszbein and Schady, 2009; Hahn *et al.*, 2018).[1] On the supply side, local governments have constructed more schools to reduce the distance to attend school. This policy has been successful in Indonesia (Duflo, 2001), Afghanistan (Burde and Linden, 2013), and Burkina Faso (Kazianga *et al.*, 2013). Also, recruiting female teachers has had a positive effect on girls' educational outcomes in India (Muralidharan and Sheth, 2016).[2] Our study extends

[1] There are other interventions improving female educational attainment in developing countries such as, among others, "girl-friendly" schools (Kazianga *et al.*, 2013), scholarships (Kremer *et al.*, 2009), school sanitation (Adukia, 2017), and female teachers (Eble and Hu, 2017).

[2] See also Muralidharan and Prakash (2017), who studied a program in India that provided all grade-nine girls funds to buy a bicycle in order to make it easier for them to access schools. This program increased girls' enrollment in secondary school by 32% and reduced the corresponding gender gap by 40%.

this literature by showing that female education in developing countries could potentially be improved within the existing school system by grouping students based on their friendship ties. Such teaching practices require relatively little guidance or monitoring from personnel outside the school. It is also cheaper to implement than other commonly used interventions which usually require extra resources.

Our paper also contributes to the small but rapidly growing literature examining the effects of friendship on performance. The evidence here is mixed. From a theoretical standpoint, working with friends may improve performance if it leads students to place more value on the group outcome or increases motivation to "catch up" with higher-ability peers. At the same time, it may impair performance if socializing with friends inhibits studying. Using an experimental study in a university context, Babcock *et al.* (2015) found that, when given monetary incentives to exercise, a student exercises more if a higher fraction of his or her friends are also given incentives to exercise. In a field experimental setting in which workers were paid a piece rate for fruit picking, Bandiera *et al.* (2010) found that workers perform better when working with more able friends and perform worse when working with less able friends. Chen and Gong (2018) examined the effect of group formation on performance by randomly assigning 685 students in an undergraduate business course at the National University of Singapore to one of three types of groups: groups assigned randomly; groups assigned to maximize skill complementarity; and groups determined by the students. They show that the members of two last groups outperform members of the first. Park (2016) found that workers in a seafood processing plant in Vietnam performed worse when they worked with their friends, suggesting that disruptions might be greater among friends. An important role of friends in children's learning level has recently been discovered by Lavy and Sand (2019) using administrative data for Israel. They exploit a unique feature of the Israeli school placement system, which assigns peers randomly conditional on school choice. Their study looks at the impact of the number of pre-existing friends and their socioeconomic background on students' academic progress from elementary to middle school, finding a positive association. As a result, one should expect that the effects of working or studying with friends on outcomes should depend on the context and the type of task. Our study is among the first to present experimental evidence on the effects of working with friends and social incentives on cognitive outcomes of children.

The remainder of the paper is as follows. In Section 1, we explain the institutional context and our experimental design. Section 2 is devoted to the description of our data. Our results are presented and discussed in Section 3. Section 4 contains robustness checks. Finally, Section 5 concludes.

1 Institutional context and experimental design

1.1 The context

Bangladesh, like many other countries in South Asia, has traditionally been characterized by low school enrolment and gender disparity in educational achievement. In 1993, the government introduced the Food for Education (FFE) program to support poor children in completing primary schooling. Under the FFE program, children from poor, rural families were given wheat rations for regular school attendance. In 2002, the FFE program was replaced by the Primary Education Stipend Project (PESP). The PESP provided cash transfers to households of children in poor areas conditioned upon the children's enrolment in and attendance at school. In addition, a variety of policies - the elimination of official school fees, free textbooks, stipends for girls, and incentives to encourage the participation of vulnerable children - have been recently put in place to encourage school enrolment (see Hahn *et al.*, 2018).

Over the last decade, enrollment rates in primary schools have increased rapidly, leading to gender parity in enrolment, reduction in dropout, and improvement in completion of the cycle. In 2015, the net enrolment rate in primary schools was 98% for girls and 97% for boys, and the net enrolment rate in secondary schools was 67% for girls and 60% for boys. The gender parity index (the school enrolment ratio of girls to boys) increased from 0.84 in 1990 to 1.06 in 2016 in primary schools and from 0.51 to 1.11 in secondary schools (source: http://uis.unesco.org/country/BD). In particular, there has been a significant enrolment growth for poor girls during recent years supported by a range of female stipend programs.

However, despite this, women and girls do not benefit from secondary education as much as men and boys do. For example, the dropout rate at secondary level in 2015 was 33.72% for boys and 45.92% for girls, while the secondary cycle completion rate was 65.98% for boys and 51.62% for girls (see Tables 16 and 17 in Asian Development Bank, 2017). In addition, there still exists a large gender gap in learning outcomes (Asadullah and Chaudhury, 2015). Therefore, a topic at the forefront of the political debate is how to increase learning levels among primary-aged children and how to close the gender gap.

1.2 The experiment

The experimental design involves within-classroom grouping among students in rural primary schools. This is not a common practice. However, students are typically grouped for extra-curricular project activities such as physical exercise, drawing, or occasionally playing in small groups during recess or lunch breaks. The idea of our experiment was to verify the prospect of using group work for educational purposes, a practice which students are familiar

with in different contexts. We randomly selected 115 out of 800 primary schools in two districts (Khulna and Satkhira). Each of our sample schools had only one class for each grade, a single teacher, and a large class size (40 students on average). We chose grade 4 students since students in grade 5 sit for the nationwide competitive exam and follow a strict daily routine, whereas grade 3 students are too young to follow our instructions and perform their homework in group.[3]

Figure 1 shows the location of the selected schools. We can see that the different treatments (peer groups and friendship groups) are reasonably distributed. In total, we interviewed 4,627 students.

[*Insert Figure* 1 *here*]

When designing the experiment, we faced two issues. First, social contacts evolve over time. For our results to be credible, there should not be too much time between the collection of friendship information and the assignment of study peers. We thus elicit friendship nominations less than a month before the grouping of students takes place. Second, the intervention (allocating students into groups) does affect the individual educational performance of students but may also change their friendship relationships. It is indeed well-documented that networks *do* rewire in response to interventions (see, e.g. Comola and Prina, 2015, and Banerjee *et al.*, 2019). To prevent restructuring of the network, we limited the period of our study to one week.

Figure 2 shows the timing of our experiment. There were two phases in the experiment. In the first stage, we elicited friendship and household information and conducted an individual cognitive ability test. In the second stage, we formed study groups and distributed assignments. After the treatment, we again tested students' achievement.

More specifically, in June 2013, we interviewed all grade 4 students in the 115 schools. We asked them to nominate up to 10 closest friends from a school roster and conducted a household survey in which parents reported their education, age, and occupation, as well as other household characteristics. Each student's ability was measured using a math test (*Individual Pre-Experiment Math Test*, IPEMT), which was developed by local educators and experts in the field of education. This is a multiple-choice test which contains 15 questions measuring numbering and number-comparison skills, numeral literacy, mastery of number facts, calculation skills, and understanding of concepts. Questions also include arithmetical reasoning, data addition, deduction, multiplication, and division. The children were given 20 minutes to complete the test. A detailed description of the IPEMT is contained in the Online Appendix.

[3]Grade 4 students are also not old enough to shy away from gender-mixed study groups. Parental concerns related to mixed gender interactions typically start arising after the children move to secondary schools, which starts at grade 6. Prior to the experiment, we received parental consent.

In July 2013, students in the schools were randomly allocated to two different settings: (1) the *peer group*, where students were assigned to a group of four peers within a school regardless of friendship and (2) the *friendship group*, where students were assigned to a group of four based on friendship nominations. We chose 80 schools at random where students were allocated to peer groups and chose 35 schools where students were allocated to friendship groups.[4] Friendship and peer groups were constructed to balance the ability of group members (that is, the mean and the distribution of student ability was comparable across groups).

[*Insert Figure* 2 *here*]

To construct peer groups that have similar characteristics across groups, we use the following methodology. We first rank students according to their IPEMT in each class/school. We then randomly select a student from each quartile of the IPEMT empirical distribution to form a group of size four. At the end of the grouping process, ANOVA tests of equality in means and variance across groups are then performed for three characteristics: cognitive ability (as measured by IPEMT), parental education, and household income. If similarity was confirmed, the grouping was recorded and a new classroom was considered. If one of these tests failed, then the grouping was discarded and the algorithm was run again. In all classrooms, groups were formed in fewer than 10 iterations. No information on friendship links was used for the group formation of peer groups.

The groupings in schools assigned to the friendship treatment were designed to have more friends than the children in peer groups. Thus, friendship groups were formed using the friendship nominations and concept of cliques in network analysis, and we tested for similarities across groups.[5] First, the computer would find an initial clique of size four, keep it, and then remove the edges (i.e. links) of the selected clique. The algorithm then found another clique of size four. This process continued until there were no other cliques of size four. For the remaining students, it would find groups in which at least one student was a friend of two other students in that group, and so forth. After the algorithm finished, we performed the tests mentioned above for differences in terms of peers' ability, parental education, and household income across groups. As in the peer-group case, if similarity was confirmed, the grouping was recorded and a new class considered; otherwise the algorithm

[4] We also had 35 schools in which children were assigned to study individually. However, children studying alone are faced with individual-performance-based-incentives whereas children studying in groups with friends or peers are faced with group-performance-based-incentives. Our analysis focuses on comparing learning improvement of students assigned to different types of peer groups because the structure of incentives is the same.

[5] A clique in a network is a subset of its vertices (i.e., nodes) such that every two vertices in the subset are connected by an edge (i.e., a link).

was run again. As in the case of peer groups, friendship groups were formed in fewer than 10 iterations in all classrooms. In our final data, more than 97 percent of the groups had 4 students. Out of 1,176 groups (924 peer groups and 252 friendship groups), 29 groups had 3 students and 1 group had 5 students.

The newly formed groups (peer and friendship groups) were then asked to take a group general knowledge test (GGKT) immediately after the groups were formed. Each group worked on this test *collectively*. The GGKT consisted of 20 multiple choices items that explored students' knowledge of national and international affairs, geography, current affairs, and sports. We allocated 20 minutes for groups to work on the test. Students were not informed about the test or its content before the test was administered. The purpose of this task was to help students learn to work as a group. After the GGKT was completed, each group was given a *group math test* (GMT) to be completed *collectively* outside school time and handed in after one week. This test consisted of 10 questions. While the questions reflect the content in the grade 4 mathematics textbook, they are not directly taken from it. To develop the test, we considered the international mathematics testing (e.g., NAPLAN) for students of this age. Following NAPLAN, we presented the mathematical problems to students in a form related to their real-life contexts. The tests were developed in consultation with retired school teachers and local education experts. A detailed description of the GGKT and GMT is included in the Online Appendix.

At the end of the week, after each group (or individual) had handed in its GMT, each student was asked to perform an individual post-experiment math test (IPOMT). The students were not aware that the grouping or testing were part of a research project. They completed their exams individually in a regular exam-taking environment where they had been assigned a seat randomly.[6] As mentioned earlier, we only allowed student interactions for one week to avoid the effects of the students in peer groups forming new friendship relationships. The IPOMT was based on the GMT. Although none of the test items were repeated from the GMT, the questions were similar so that students could apply what they had learned from the group project (GMT). A detailed description of the IPOMT is contained in the Online Appendix. Students were given 1.5 hours to complete this test. Students had been informed at the beginning of the week that they would take an individual test after one week. To incentivize students to work together, they were also told that the study effort for the group project would help them to do well on the individual test.[7] At the end of the week, students

[6] While the researchers were involved in the design and in the training of the field workers, they were not present in the study period.

[7] The concern that students (particularly low-ability students) might be embarrassed in a group setting since this setting may reveal their ability through interactions with their peers did not apply in our context. In fact, students already knew each other's academic ability well since teachers routinely give class tests whose results are announced in front of all the students.

were asked to complete a short questionnaire on their group study effort. The questions included (1) the number of times students met as a team (extensive margin) and (2) how many hours the group met as a team (intensive margin).[8]

Students were given prizes following a tournament scheme based on their group's performance on the different tests. For students belonging to peer or friendship group settings, there was a prize for the best performing group on the GGKT. For the math tests, two prizes were given in each class: one prize for the group with the highest average score on the IPOMT (*best performing group*), and another prize for the group with the greatest improvement from their group average baseline math test (that is, the most improvement between the IPEMT and IPOMT).[9] This prize scheme was chosen to ensure that all the students were incentivized to work together and help each other during the week, as well as to mitigate a potential discouragement effect for those who were grouped with low-performing students.

For the GGKT, the prize was a pencil box scale (ruler) for each student of the best performing group. For the best performing group in the IPOMT and for the highest improvement group (between IPEMT and IPOMT), students are given an instrument box (geometry box) or a diary and scale. These prizes were set in consultation with teachers and students to ensure that they were comparable incentives. The cost of the prize for each student was approximately US$1. If two or more groups (or students) made the same score, all of them received the prizes. In addition, all participating children received gifts (e.g. a pencil/pen) and certificates for their participation.

2 Data description

The network survey and the household survey were administered to all grade four students in all 115 schools. As mentioned above, we asked students to nominate up to 10 closest friends from a classroom/grade roster. Figure 3 reports the distribution of students by the number of same-gender nominations. More than 50% of the students nominated more than 80% of the friends with the same gender. The tendency to nominate mainly same-gender friends did not, however, show marked differences by gender. Gender differences were also minimally present for other drivers of friendship formation. Table 1 shows the percentage of same-type friends for cognitive ability (IPEMT), parental education, and family income by gender and group-type. The percentages on the main diagonal indicate the percentage of same-type nominated friends. These percentages are remarkably similar by gender and are generally

[8] There was no mention of a gender focus in any part of the experiment (including the survey).

[9] We designed the incentive scheme based not just on the overall performance of the group but also on the improvement of the group (that is, improvement compared to the average IPEMT of the group members), to avoid a potential discouragement effect among group members who might have considered themselves grouped with low-ability peers.

slightly above 50%. This seems to indicate that there was not a strong tendency toward homophily behaviours (McPherson et al., 2001) for characteristics different from gender.

[*Insert Table* 1 *and Figure* 3 *here*]

Panel (a) in Figure 4 depicts the distribution of students by number of friends, distinguishing between friendship and peer groups. As expected, when grouping is based on peers (in blue, solid), most individuals end up in a group where very few students are friends. In more than 50% of the cases, a student had no friend at all in his or her assigned group. When grouping is based on friendship (in red, dashed), the opposite is true. Panel (b) in Figure 4 shows the distribution of students by the total number of links within a group, distinguishing between peer and friendship groups. The figure confirms that for individuals in peer groups, there were few friendship links, while for those in the friendship groups, the opposite was true.[10]

[*Insert Figure* 4 *here*]

Figure 5 provides the gender composition of friendship (dashed red bars) and peer (solid blue bars) groups. We see that the friendship groups were more homogenous in terms of gender composition than peer groups but still showed great variation. For example, 36% of the friendship groups were composed of only females versus 6% for peer groups. Similarly, 30% of the friendship groups had only male students versus 6% for peer groups. For mixed groups, the overlap between friendship and peer groups was much higher. To summarize, in friendship groups, roughly 1/3 were composed of only female students, 1/3 were only male students, and 1/3 had a mix of male and female students. For peer groups, 1/3 of the groups were composed of half females and half males.[11]

[*Insert Figure* 5 *here*]

Table 2 shows the pre-experiment gender gap in test scores (IPEMT) across treatment types.[12] Regardless of the treatment status, females always performed worse than males. On average, females' IPEMT scores were roughly 0.15 standard deviations below the average,

[10] The total number of links varies between 0 and 11 in peer groups and between 1 to 12 in friendship groups. We eliminated 49 (out of 4,676) students who ended up by chance in groups with no friends in the schools assigned to the friendship grouping treatment.

[11] The gender composition of the groups is not pre-determined by the group formation algorithm. In Figure A1 of the Online Appendix, we show the distribution of peer groups by fraction of females when rerunning the algorithm 1000 times with different seeds. The picture shows that the realized gender composition is one of many possibilities, and it is not an atypical one.

[12] We regressed the pre-experiment test (IPEMT) scores on an indicator for Female, which took the value 1 if the student was a female and 0 if male, with and without a set of controls.

and this gender gap did not close when we controlled for observable student characteristics such as household income and educational attainment of the parents.

[*Insert Table 2 here*]

Table 3 presents summary statistics for the two types of schools (peer and friendship groups). Many households in this region of rural Bangladesh lack access to electricity, and only about 27 percent of the sample students had access to electricity at home. Parental educational attainment, measured as the maximum of mother's and father's years of education, is on average 5 years.[13,14] The last three columns of the table formally test whether there were statistically significant differences between the schools placed in the three settings in terms of the observed characteristics. It appears that all characteristics were well balanced except for the percentage of females, which was slightly higher in the schools assigned to the friendship treatment.[15]

[*Insert Table 3 here*]

Figure 6 shows the gender gap in school performance before (IPEMT) and after (IPOMT) the experiment, distinguishing between group types. From left to right, the figures are plotted for the friendship and peer group schools. The top figures show the IPEMT distributions and the bottom figures depict the IPOMT distributions. The test scores are standardized across schools so that the average value of the test score is 0 with standard deviation equal to 1. While the performance of boys was minimally affected by group-type, the performance of girls was clearly affected by the treatment. Moreover, while male students performed better than female students before the experiment, females studying in friendship groups caught up with them in the post-experimental math test. Finally, this figure shows that the pre-experiment performance of females assigned to friendship groups was roughly similar to that of females in peer groups. However, after the treatment, that is, after having interacted for a week with their peers, females who worked with friends outperformed females working

[13] This generation of parents received education at least more than 25 years ago when the education system in Bangladesh was not developed, and there was no compulsory education system in place. There has been a significant increase in education over the last two decades. As a result, our students did not leave school the year after the experiment.

[14] Roughly 16 percentage of the students missed the IPEMT. We imputed their scores using gender, school fixed effects, and test scores of subjects in Bengali, English, Math, and Science administered at schools. The likelihood of a missing test score was not different across school types, and we control for an indicator of missing IPEMT in our analysis. The results did not change qualitatively when we dropped students with imputed test scores.

[15] All our regression results are presented with and without controlling for the percentage of females in the group.

with peer groups.

[*Insert Figure* 6 *here*]

Figure 7 depicts the estimated post-experimental performance (IPOMT) against the initial levels of ability (as measured by the pre-experiment test, IPEMT) using a local polynomial smoother that allows for non-linear effects. The figure reveals that grouping has an heterogeneous effect across ability types. In particular, the positive gains from studying with friends for females are only present for low-ability students. In the remainder of this paper, we further investigate these stylized facts using a regression analysis.

[*Insert Figure* 7 *here*]

3 Results

We estimate the following regression model:

$$y_{irs}^{IPOMT} = \gamma_0 + \gamma_1 D_{\text{friend}} + \gamma_2 y_{irs}^{IPEMT} + \gamma_3 X_{irs} + \eta_{irs} \quad (1)$$

where y_{irs}^{IPOMT} is the math score of the post-experiment test (IPOMT) and y_{irs}^{IPEMT} is the math score of the pre-experiment test (IPEMT) of individual i belonging to group r in school s. D_{friend} is a dummy variable that is equal to 1 if student irs belongs to a friendship group and 0 if he/she belongs to a peer group. X_{irs} denotes the observable characteristics of individual i belonging to group r in school s (parents' education, household income per capita, access to electricity, etc., as shown in Table 3), and η_{irs} is an error term. Standard errors are clustered at the school level.

Table 4 reports the OLS results in columns (1) to (4) with an increasing set of controls. The challenge in our analysis is that the gender composition in a group is not orthogonal to the treatment. This is because students tend to nominate same-gender friends (Figure 3). Since peer-gender composition may be important in shaping a student performance, gender composition is a possible pathway for the effects of studying with friends. As a result, in column (2), we add the fraction of females as a control.[16] In column (3), we control for individual pre-experimental ability (as measured by IPEMT), and, in column (4), we add additional observable characteristics. The results reveal no statistically significant differences in learning performance between friendship and peer groups in any of our specifications.

[16] This logic (which is also used by Lavy and Schlosser, 2011; Lavy *et al.*, 2012; Bifulco *et al.*, 2011, among others) is akin to a mediation analysis. The results of a formal mediation analysis (Imai *et al.*, 2010) are discussed in footnote 18.

[*Insert Table* 4 *here*]

In Table 5, we investigate whether the effects depend on the gender and ability of the students, with an increasing set of controls (as in Table 4). Using the distribution of the IPEMT for the whole sample, we define low-ability students as those below the median value, whereas high-ability students are those above the median value.[17] The results reveal that, for male students, there was no effect of studying with friends rather than peers on the change in math test scores. In contrast, there was a large and positive gain in math scores for low-ability female students who studied in friendship groups, compared to those studying in peer groups. The effect is large and statistically significant, especially when controlling for the fraction of females in the group. Indeed, compared to studying in peer groups, studying with a group of friends increases the test scores of low-ability female students by roughly 0.4 standard deviations on the IPOMT.[18]

[*Insert Table* 5 *here*]

In order to deal with the potential concern that our sub-group analysis based on ability and gender suffered from type I error, we adjusted the p-values for multiple hypothesis testing. In particular, the Family Wise Error Rate (FWER) was estimated using the free step-down resampling approach of Westfall and Young (1993).[19] These adjusted p-values are reported in italics in the row below the standard error of the friendship effect. We see that by adjusting for multiple hypothesis testing, the p-values increase but the results remain statistically significant at the conventional level.[20]

We continue our analysis by investigating whether it is peer friendship that matters rather than the peer-gender composition. A large literature documents that a single-sex environment may facilitate more effective learning for females. Lessened gender stereotype threat may lead to increased self-confidence (Booth *et al.*, 2014; Eisenkopf 2015; Spencer et al., 1999). For example, using a natural experiment at a high school in Switzerland,

[17] Due to the discrete scoring of the IPEMT, the percentages of students below and above the median are 45 and 55 percent, respectively.

[18] In order to appreciate how much of this effect is, indeed, mediated by gender composition, we conducted a formal mediation analysis using the approach (and Stata code) provided by Imai *et al.* (2010). The estimated mediation effect is about 20%, but it is imprecisely estimated with a confidence interval that includes zero. Importantly, the mediation effect is negative, reflecting an adverse effect of having a higher fraction of females in the group for females in friendship groups. If the fraction of females is excluded from the model, this implies that the treatment effect may be underestimated, as shown in Table 5.

[19] We use the method implemented by Jones *et al.* (2018) to obtain the FWER adjusted p-values.

[20] For each of the subsamples, we also test for balance in pre-determined characteristics between friendship groups and peer groups. The results, displayed in Table A1 of the Online Appendix, show no significance imbalance for all of these characteristics.

Eisenkopf *et al.* (2015) found that single-sex schooling improves female students' performance in mathematics and boosts self-confidence. In the absence of male peers, females may also feel less anxiety pursuing non-stereotypical courses, such as math (Mael *et al.*, 2004). Because friendship groups have a higher fraction of same-gender members, it is important to rule out the possibility that our results simply reflect the exposure to a larger fraction of same-gender peers. We performed three exercises. The results are displayed in Table 6. First, we looked at our (semi-random) peer groups and explored whether the outcomes of students allocated in study groups with the same-gender peers differ from those of students allocated to study groups with mixed-gender peers. The results are shown in panel A. We found no difference in performance for high-ability or low-ability female students.

Second, we considered all groups with a given gender composition and explored whether any gains resulted from being grouped with friends. Panels B and C display the results. In Panel B, we consider the peer groups and friendship groups with students of mixed gender,[21] while, in panel C, we look at peer groups and friendship groups composed of only same-gender peers and estimate the additional effect in performance of being with friends.[22] The results displayed in panel B show a higher performance on the IPOMT for low-ability females in friendship groups. The estimated friendship effects are statistically significant at the conventional levels even when adjusting for inference from multiple hypothesis testing. In panel C, we only consider groups composed of same-gender peers. If same-gender dynamics explained the improvement in performance rather than friendship per se, then we should see no difference in outcomes between students belonging to friendship groups and those in female-only-peer groups. The results show, on the contrary, that the gains from working with friends are substantial. The estimate of the friendship effects is qualitatively similar to the estimate reported in Table 5.[23]

[*Insert Table* 6 *here*]

The evidence in Table 6 indicates that effects stemming from studying in a same-gender environment are not consistent with the large gains revealed in our analysis since (i) all-female groups show no substantial gains compared to mixed-gender groups and (ii) the friendship effect remains strong even when we use the subsample of friendship and peer groups with a similar gender-mix environment. Friends seem to matter the most in explaining

[21] In principle, we could compare friendship groups and peer groups for every possible gender composition, but we did not have sufficient power to do so. Our results were obtained using students in *any* gender-mix environment.

[22] In Table A2 of the Online Appendix, we perform balance checks in observable characteristics for the subsamples used in panel B and panel C. These tests do not reveal any sign of imbalance of characteristics between students in friendship and peer groups for any subsample.

[23] The hypothesis that the friendship effects in Table 5 and 6 are equal cannot be ruled out.

the individual outcomes.

A theory consistent with the importance of studying with friends for females can be found in the sociological literature. Indeed, a number of studies suggest that for women, improvements from group work may be driven by social indispensability, that is by the feeling that people care about the value of their own performance for the group outcome (see, e.g. Weber *et al.*, 2009). Girls surrounded by friends may feel socially accepted and thus more comfortable expressing themselves. Group study can thus indirectly improve performance by increasing the amount of participation in the learning process. In other words, while a classroom setting may encourage passive learning, a small group setting may encourage a student to think more deeply about a given topic because he/she will need to discuss it the with other students in the group. If within-group differences challenge individual participants' thinking (both among high achievers – who have to "teach" the material to others – and among the low achievers, who might find their high-performing peers easier to approach than their teachers), then we would expect to see small groups improve learning. Females might benefit more than males in this context if they are less likely to engage in the learning process in a general classroom setting. Additionally, females may only engage if they are in a group with friends, whereas males may feel comfortable engaging regardless of whether they are with friends (or even regardless of whether they are in a group). This theory is thus in line with the finding that low-ability female students tend to perform better in friendship groups.

An alternative story for our result is that our friendship dummy picks up the frequency of interactions. Female students in friendship groups may meet more often (or study more) during the week for the collective assignments than those in peer groups. Indeed, given the traditionally strong differences in the roles of males and females in Bangladesh, families may feel more comfortable having their young girls interact with other girls outside of school than with boys (so study groups meet more when peers are largely of the same sex). The post-experiment survey gives us the ability to consider and rule out this possibility. We compared the effort of students working in peer groups and in friendship groups using the following regression model:

$$INT_{irs} = \beta_0 + \beta_1 D_{\text{friend}} + \beta_2 ff_{irs} + \beta_3 X_{irs} + \epsilon_{irs} \tag{2}$$

where INT_{irs} is either the number of times the group meets during the week (Num Met) or the number of hours the group meets during the week (Team Hrs). All the other variables have the same interpretation as in (1). The results are shown in Table 7. This table shows no differences in frequency of interactions or study time between peer and friendship groups with the exception of high-ability females in friendship groups, who studied more hours with

group members than did their counterparts in peer groups.

[*Insert Table 7 here*]

Our results suggest that low-ability females achieve increased learning in friendship groups, and that this is not due to the time spent on work together. The social psychology literature suggests additional reasons this may be the case. It has been noted that motivation gains are highest in cohesive groups (Karau and Hart, 1998). In Table 8, we thus investigate whether friendship effects are stronger when studying in groups with all friends by interacting D_{friend} with a dummy variable taking the value of 1 if the student is in a group where all members are friends, and 0 otherwise. The results show that a point estimate of the interaction term is large and positive, but not statistically significant due to a large standard error.[24]

[*Insert Table 8 here*]

Some evidence supporting the mechanism that studying in small groups with friends may improve learning can be found by comparing the distributions of the group outcome for the test performed immediately after the groups were formed (GGKT) and of the group outcome for the test that took place after a week of interactions (GMT) by grouping schemes. If learning is an important factor in enhancing student performance when studying with friends, we would expect to find no differences in the distributions of the GGKT scores between peer and friendship groups because students had no time to interact. In a similar vein, we would expect to find a difference in the distributions of the GMT scores after a week of interactions.

Figure 8 displays the kernel density plots for the GGKT and the GMT, distinguishing between peer and friendship groups. The graphs show that while the two curves almost overlap for the GGKT, the distribution of the GMT scores for friendship groups is shifted to the right. We formally tested these differences using a Kolmogorov–Smirnov test. The test did not reject the null hypothesis that the GGKT has the same distribution between the peer and the friendship groups (p-value equal to 0.375), while it did detect a statistically significant difference in distributions between these two types of groups for the GMT (p-value smaller than 0.001). The results for the GGKT and GMT are consistent with the idea that greater learning takes place within groups of friends than within groups of peers who

[24] We further investigated nonlinear effects by analysing whether the effect is different when studying with one, two, or three friends separately. Due to large standard errors, we were not able to detect nonlinear effects in any of these cases.

are not necessarily friends to each other.[25]

[*Insert Figure 8 here*]

4 Robustness checks and additional evidence

In this section we report on robustness tests that we performed and collect additional evidence. One idea consistent with the fact that we found effects for low-ability females and not for low-ability males is that low-ability males have less room for improvement than low-ability females. This was true of these data. Low-ability male students had, on average, a baseline ability level in initial ability (IPEMT) higher by 0.115 standard deviations than low-ability females. To address this concern, we removed high-scoring low-ability male students from the subsample of low-ability male students (so that their average IPEMT matched that of low-ability female students) and replicated our baseline results in Table 5. Among the 1,000 low-ability male students, we dropped 182 students, so that the difference in IPEMT scores between the remaining low-ability males and the original low-ability females was on average close to 0 (i.e., less than 0.001 standard deviation). Table 9 shows the results after making the subsample of low-ability males comparable to the low-ability female sample. The evidence remains unchanged, suggesting that the lack of friendship effect for low-ability male students is not due to having less room for improvement.[26]

[*Insert Table 9 here*]

In our next robustness check, we added two characteristics of the group environment that might affect students' performance. We started by controlling for the level of ability of the peers in the study group. This is an important check, given that girls on average have lower IPEMT scores, as shown in Table 2. Having more female students in a group may thus capture a lower ability environment. We investigated this issue by controlling for the *average* IPEMT of the peers. The results are reported in columns (1) and (4) of Table 10

[25] The fact that we see an effect for friendship groups is also consistent with the idea that friends mitigate free-riding behaviour. While in groups with strangers the best students do all the work, in groups with friends there is more collaboration and in the end all students learn more. This mechanism, however, should apply to both males and females, whereas in our analysis we found an effect of friendship only for low-ability females. Also, Table 7 shows that it does not seem to be true that groups with friends meet more often than groups with strangers.

[26] Another conjecture is that males tend to have more social interactions outside school than girls. However, to the best of our knowledge, there are no conscious attempts or cultural/religious norms for such 10-11-year-old children to have different social interactions by gender both within and outside the family circle. Our casual observation is that both boys and girls play with friends outside the classroom during recess, lunch breaks, or any other time when the teachers permit, without any noticeable difference.

and show that our evidence on the importance of friends for low-ability females remains true. Perhaps not surprising, the direct effect of the average IPEMT on individual performance of low-ability girls is significantly positive. Another issue is the fact that friendship groups generally have a tighter variance in terms of members' abilities. In columns (2) and (5) of Table 10 we thus investigate whether the dispersion in the ability of the peers affected educational outcomes by including the standard deviation of the IPEMT scores of the peers in the regression. The results show that the evidence remains qualitatively unchanged. In columns (3) and (6) we include in the model both the average and standard deviation of the IPEMT scores of the peers. Our results remain robust to the addition of both factors.

[*Insert Table* 10 *here*]

Finally, we exploited the semi-randomness of the peer group assignment to make a casual inference of the effect of peer ability (as measured by the average IPEMT of peers) on test scores. For this purpose, we focused our analysis on the schools with the (semi-random) peer grouping treatment. We collect the results in Table 11. The table reveals positive and statistically significant peer effects for low-ability students, both males and females, although the results are stronger for females. When considering this evidence in the context of our paper, the results of this exercise support the importance of studying with friends for low-ability female students since they show that peer ability increases the performance of low-ability students among both females and males. Peer friendship, in contrast, shows an effect which is specific to females.

[*Insert Table* 11 *here*]

5 Concluding remarks

Fighting low levels of basic education and improving the gender gap in developing countries is a priority for economic development. In this paper, we explored the efficacy of an intervention that consisted of allocating students into small groups with friends in the context of Bangladeshi primary schools. When compared to students allocated to small groups with peers, we find that friendship groups increase the achievement of low-ability female students, i.e., female students with test scores below the median. This friendship effect does not appear to be due to the gender composition of the groups or to the frequency of interactions.

In the educational psychology literature, there is a long tradition of research on the effect of friendship on various interpersonal and group outcomes. Friendship has been found to affect learning (Kutnick and Kington, 2005; Foot and Barron, 1990) and collaboration (Miell

and MacDonald, 2000; MacDonald *et al.*, 2000; Andersson, 2001) amongst students in the classroom. However, in this literature, some research has suggested a positive effect of friendship on group performance (e.g. Jehn and Shah, 1997; Shah and Jehn, 1993; Harrison *et al.*, 2003), while other research has documented that friendship negatively impacts performance (e.g. Andersson and Rönnberg, 1995; Swenson and Strough, 2008). Our experiment reveals that the effects of friendship may depend on the subjects' gender and ability.

By showing that friends act as facilitators of learning for low-ability female children in Bangladesh, our analysis depicts friendship as a possible channel to help close the gender gap in education in developing countries. The practice of placing students into small friendship groups is relatively inexpensive and only requires teachers to know the ability of students and the friendship composition in each classroom. This contrasts with other interventions in developing countries, where the focus has been on either increasing the immediate benefits of schooling to families by, for example, providing cash transfers to households, or on reducing the costs of attending school by, for example, constructing more schools in order to reduce the geographical distance to attend school.

To the best of our knowledge, few studies have investigated the role of friends in improving the outcomes of women in developing countries. An exception is Field *et al.* (2016), who conducted a field experiment in India and showed that there are substantial differences in borrowing behavior between women who attend business training sessions alone and those who attend with a friend. Only women invited with a friend borrowed as a result of the training sessions, and they almost exclusively used the marginal loans for business purposes. More strikingly, four months later, those invited with a friend also reported significantly higher household income and expenditures and were less likely to report their occupation as housewife.[27] Our results demonstrate that mixing female children with their friends in the classroom can be a cost-effective tool for improving female educational outcomes in developing countries. Such grouping within classrooms can be implemented even in poor rural primary schools where schools receive very limited funding for their operation. Our experimental evidence suggests that fostering a support system of peers may be an effective intervention to improve outcomes for students from underrepresented backgrounds.[28]

Yonsei University, South Korea; Monash University, Australia; Cornell University, USA, EIEF, IZA and CEPR; Monash University, Australia, and IFN.

[27] There is also a (small) literature on gender and peer effects in developing countries, but peers are usually defined at an aggregate level (for example, the classroom or the school) since these studies do not have information about direct friends (see e.g. Duflo *et al.*, 2011).

[28] Those types of interventions are promoted, for example, by the Posse Foundation in the context of higher education (see https://www.possefoundation.org/).

References

[1] Adukia, A. (2017). 'Sanitation and education', *American Economic Journal: Applied Economics*, vol. 9(2), pp. 23–59.

[2] Andersson, J. (2001). 'Net effect of memory collaboration: How is collaboration affected by factors such as friendship, gender and age?' *Scandinavian Journal of Psychology*, vol. 42(4), pp. 367–375.

[3] Andersson, J. and Rönnberg, J. (1995). 'Recall suffers from collaboration: Joint recall effects of friendship and task complexity', *Applied Cognitive Psychology*, vol. 9(3), pp. 199–211.

[4] Andrabi, T., Das, J., Khwaja, A.I., Vishwanath, T. and Zajonc, T. (2007). 'Learning and educational achievements in Punjab schools (leaps): Insights to inform the education policy debate', Washington, DC: World Bank.

[5] Asadullah, M.N. and Chaudhury, N. (2015). 'The dissonance between schooling and learning: Evidence from rural Bangladesh', *Comparative Education Review*, vol. 59(3), pp. 447–472.

[6] Asian Development Bank (2017). *Bangladesh: Gender Equality, Diagnostic of Selected Sectors*, https://www.adb.org/documents/bangladesh-gender-equality-diagnostic-selected-sectors.

[7] Babcock, P., Bedard, K., Charness, G., Hartman, J. and Royer, H. (2015). 'Letting down the team? Social effects of team incentives', *Journal of the European Economic Association*, vol. 13(5), pp. 841–870.

[8] Bandiera, O., Barankay, I. and Rasul, I. (2010). 'Social incentives in the workplace', *Review of Economic Studies*, vol. 77(2), pp. 417–458.

[9] Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M. and Walton, M. (2016). 'Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level in India" ', NBER Working Paper No. 22746.

[10] Banerjee, A., Cole, S., Duflo, E. and Linden, L. (2007). 'Remedying education: Evidence from two randomized experiments in India', *Quarterly Journal of Economics*, vol. 122(3), pp. 1235–1264.

[11] Banerjee, A., Chandrasekhar, A., Duflo, E. and Jackson, M.O. (2019). 'Using gossips to spread information: Theory and evidence from two randomized controlled trials', *Review of Economic Studies*, forthcoming.

[12] Begum, L., Islam, A. and Smyth, R. (2017). 'Girl power: Stipend programs and their effects on the education of younger siblings', *Journal of Development Studies*, vol. 53(11), pp. 1882–1898.

[13] Bharadwaj, P., De Giorgi, G., Hansen, D. and Neilson, C. (2016). 'The gender gap in mathematics: Evidence from Chile', *Economic Development and Cultural Change*, vol. 65(1), pp. 141–166.

[14] Bifulco, R., Fletcher, J.M. and Ross, S.L. (2011). 'The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health', *American Economic Journal: Economic Policy*, vol. 3(1), pp. 25–53.

[15] Booth, A., Cardona-Sosa, L. and Nolen, P. (2014). 'Gender differences in risk aversion: Do single-sex environments affect their development?' *Journal of Economic Behavior and Organization*, vol. 99(C), pp. 126–154.

[16] Burde, D. and Linden, L.L. (2013). 'Bringing education to Afghan girls: A randomized controlled trial of village-based schools', *American Economic Journal: Applied Economics*, vol. 5(3), pp. 27–40.

[17] Chen, R. and Gong, J. (2018). 'Can self selection create high-performing teams?' *Journal of Economic Behavior and Organization*, vol. 148(C), pp. 20–33.

[18] Comola, M. and Prina, S. (2015). 'Treatment effect accounting for network changes', Working Paper, Paris School of Economics.

[19] Duflo, E. (2001). 'Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment', *American Economic Review* , vol. 91(4), pp. 795–813.

[20] Duflo, E., Dupas, P. and Kremer, M. (2011). 'Peer effects and the impacts of tracking: Evidence from a randomized evaluation in Kenya', *American Economic Review*, vol. 101(5), pp. 1739–1774.

[21] Eble, A. and Hu, F. (2017). 'Stereotypes, role models, and the formation of beliefs', Center for Development Economics and Policy, CDEP-CGEG Working Paper No. 43, Columbia University.

[22] Eisenkopf, G., Hessami, Z., Fischbacher, U. and Ursprung, H.W. (2015). 'Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland', *Journal of Economic Behavior and Organization*, vol. 115(C), pp. 123–143.

[23] Evans, D.K. and Popova, A. (2016). 'What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews', *World Bank Research Observer*, vol. 31(2), pp. 242–270.

[24] Field, E., Jayachandran, S., Pande, R. and Rigol, N. (2016). 'Friendship at work: Can peer effects catalyze female entrepreneurship?' *American Economic Journal: Economic Policy*, vol. 8(2), pp. 125–153.

[25] Fiszbein, A., and Schady, N. (2009). 'Conditional cash transfers: Reducing present and future poverty', Working Paper, Washington DC: World Bank.

[26] Foot, H. and Barron, A.M. (1990). 'Friendship and task management in children's peer tutoring,' *Educational Studies*, vol. 16(3), pp. 237–250.

[27] Ganimian, A.J. and Murnane, R.J. (2016). 'Improving educational outcomes in developing countries: Lessons from rigorous impact evaluations', *Review of Educational Research*, vol. 86(3), pp. 719–755.

[28] Glewwe, P. and Muralidharan, K. (2016). 'Improving school education outcomes in developing countries: Evidence, knowledge gaps, and policy implications', in (E.A. Hanushek, S.J. Machin and L. Woessmann, eds.), *Handbook of the Economics of Education, Vol. 5*, pp. 653–744, Amsterdam: Elsevier Science.

[29] Hahn, Y., Islam, A., Nuzhat, K., Smyth, R. and Yang, H.-S. (2018). 'Education, marriage and fertility: Long-term evidence from a female stipend program in Bangladesh', *Economic Development and Cultural Change*, vol. 66(2), pp. 383–415.

[30] Harrison, D.A., Mohammed, S., McGrath, J.E., Florey, A.T. and Vanderstoep, S.W. (2003). 'Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality', *Personnel Psychology*, vol. 56(3), pp. 633–669.

[31] Imai, K., Keele, L. and Yamamoto, T. (2010). 'Identification, inference and sensitivity analysis for causal mediation effects', *Statistical Science*, vol. 25(1), pp. 51–71.

[32] Jehn, K.A. and Shah, P.P. (1997). 'Interpersonal relationships and task performance: An examination of mediation processes in friendship and acquaintance groups', *Journal of Personality and Social Psychology*, vol. 72(4), pp. 775.

[33] Jones, D., Molitor, D. and Reif, J. (2018). 'What do workplace wellness programs do? Evidence from the Illinois workplace wellness study,' National Bureau of Economic Research Working Paper No. 24229.

[34] Karau, S.J. and Hart, J.W. (1998). 'Group cohesiveness and social loafing: Effects of a social interaction manipulation on individual motivation within groups', *Group Dynamics: Theory, Research, and Practice*, vol. 2(3), pp. 185–191.

[35] Karau, S.J. and Williams, K.D. (1993). 'Social loafing: A meta-analytic review and theoretical integration', *Journal of Personality and Social Psychology*, vol. 65(4), pp. 681.

[36] Kazianga, H., Levy, D., Linden, L.L. and Sloan, M. (2013). 'The effects of girl-friendly schools: Evidence from the BRIGHT school construction program in Burkina Faso', *American Economic Journal: Applied Economics*, vol. 5(3), pp. 41–62.

[37] Kremer, M.R., Miguel, E. and Thornton, R. (2009). 'Incentives to learn', *Review of Economics and Statistics*, vol. 91(3), pp. 437–456.

[38] Kutnick, P. and Kington, A. (2005). 'Children's friendships and learning in school: Cognitive enhancement through social interaction?' *British Journal of Educational Psychology*, vol. 75(4), pp. 521–538.

[39] Lavy, V.M., Paserman, D. and Schlosser, A. (2012). 'Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom', *The Economic Journal*, vol. 122(559), pp. 208–237.

[40] Lavy, V.M. and Schlosser, A. (2011). 'Mechanisms and impacts of gender peer effects at school', *American Economic Journal: Applied Economics*, vol. 3(2), pp. 1–33.

[41] Lavy, V. and Sand, E. (2019). 'The effect of social networks on student's academic and non-cognitive behavioral outcomes: Evidence from conditional random assignment of friends in school', *The Economic Journal*, forthcoming.

[42] MacDonald, R., Miell, D. and Morgan, L. (2000). 'Social processes and creative collaboration in children', *European Journal of Psychology of Education*, vol. 15(4), pp. 405–415.

[43] Mael, F. A., Smith, M., Alonso, A., Rogers, K. and Gibson, D. (2004). 'Theoretical arguments for and against single-sex schools: A critical analysis of the explanations', Working Paper, Washington, DC: American Institutes for Research.

[44] McPherson, M., Smith-Lovin, L. and Cook, J.M. (2001). 'Birds of a feather: Homophily in social networks', *Annual Review of Sociology*, vol. 27(1), pp. 415–444.

[45] Miell, D. and MacDonald, R. (2000). 'Children's creative collaborations: The importance of friendship when working together on a musical composition', *Social Development*, vol. 9(3), pp. 348–369.

[46] Muralidharan, K. and Prakash, N. (2017). 'Cycling to school: Increasing secondary school enrollment for girls in India', *American Economic Journal: Applied Economics*, vol. 9(3), pp. 321–350.

[47] Muralidharan, K. and Sheth, K. (2016). 'Bridging education gender gaps in developing countries: The role of female teachers', *Journal of Human Resources*, vol. 51(2), pp. 269–297.

[48] Park, S. (2016). 'Socializing at work: Evidence from a field experiment with manufacturing workers', Working Paper, Northwestern University.

[49] Shah, P.P. and Jehn, K.A. (1993), 'Do friends perform better than acquaintances? The interaction of friendship, conflict, and task', *Group Decision and Negotiation*, vol. 2(2), pp. 149–165.

[50] Spencer, S.J., Steele, C.M. and Quinn, D.M. (1999). 'Stereotype threat and women's math performance', *Journal of Experimental Social Psychology*, vol. 35(1), pp.4–28.

[51] Swenson, L.M. and Strough, J. (2008). 'Adolescents' collaboration in the classroom: Do peer relationships or gender matter?' *Psychology in the Schools*, vol. 45(8), pp. 715–728.

[52] UWEZO (2014). 'Annual learning assessment report', Available at: http://www.uwezo.net.

[53] Weber, B., Wittchen, M. and Hertel, G. (2009). 'Gendered ways to motivation gains in groups', *Sex Roles*, vol. 60(9), pp. 731–44.

[54] Westfall, P.H. and Young, S.S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment*, New York: John Wiley and Sons.

Figure 1: Location of the different schools

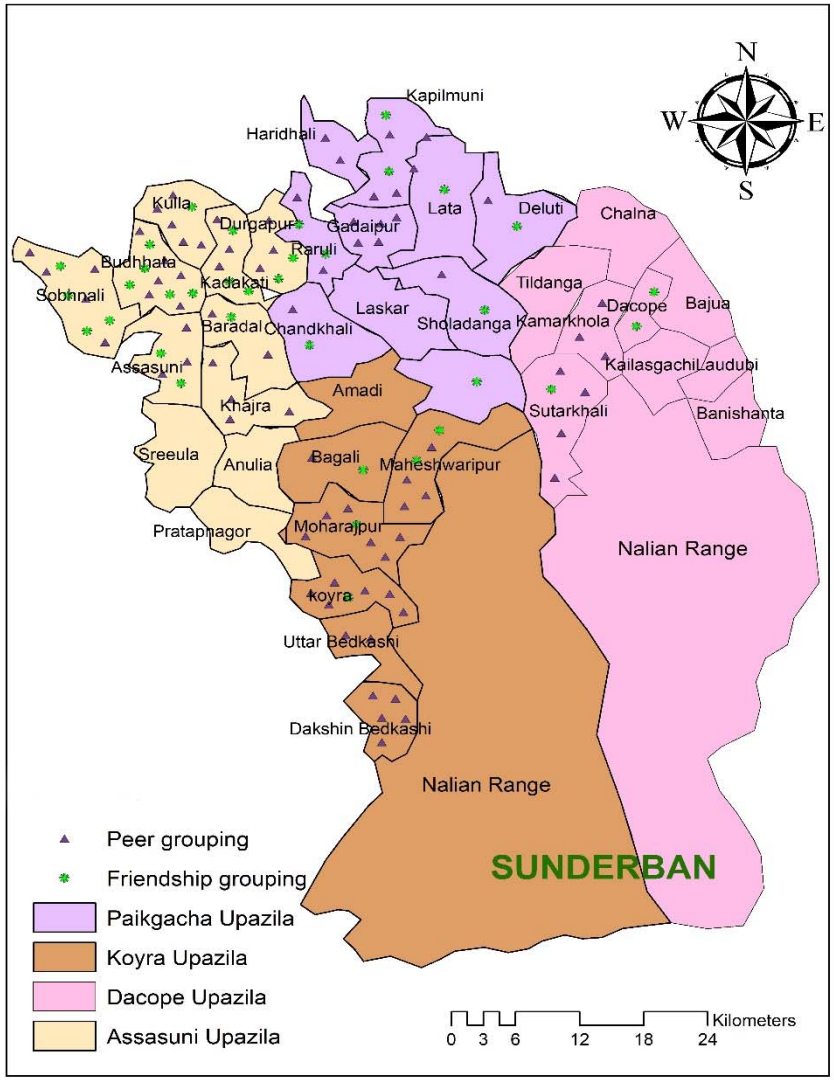


Figure 2: Timeline of the experiment

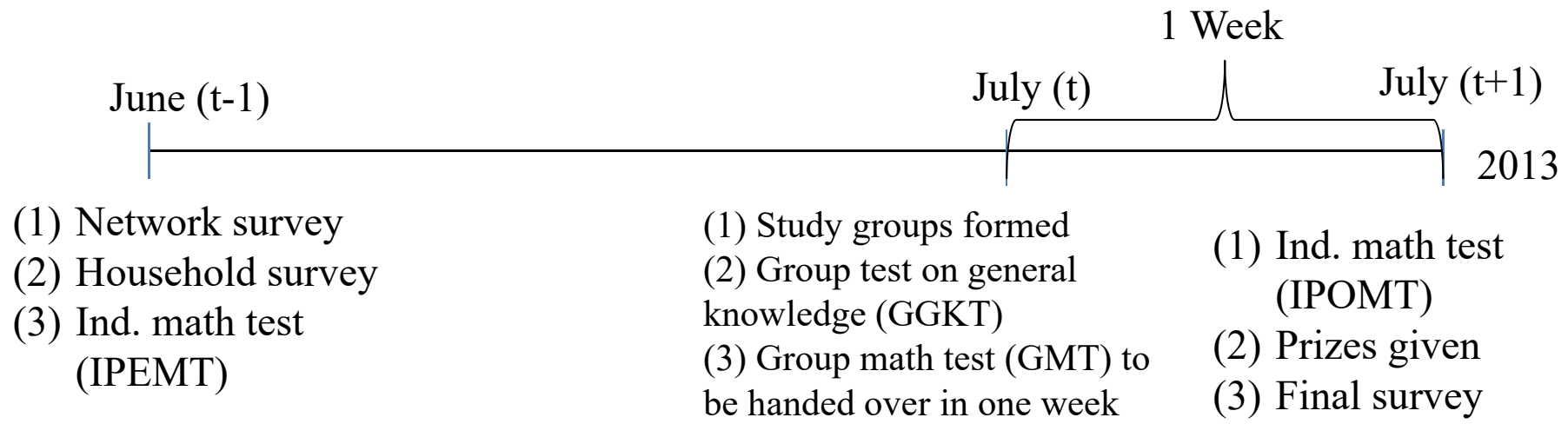


Figure 3: Distribution of students by same-gender friendship nomination

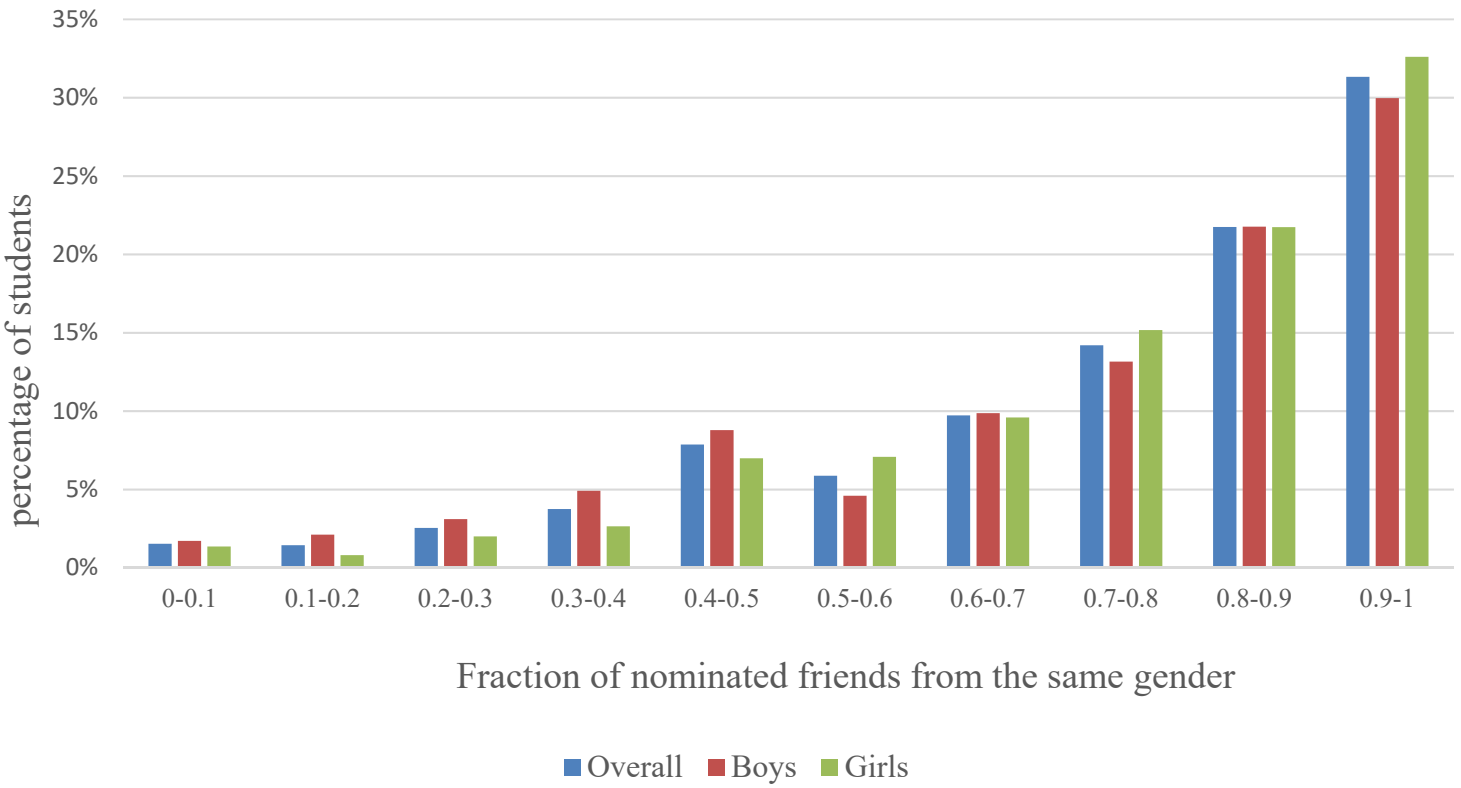
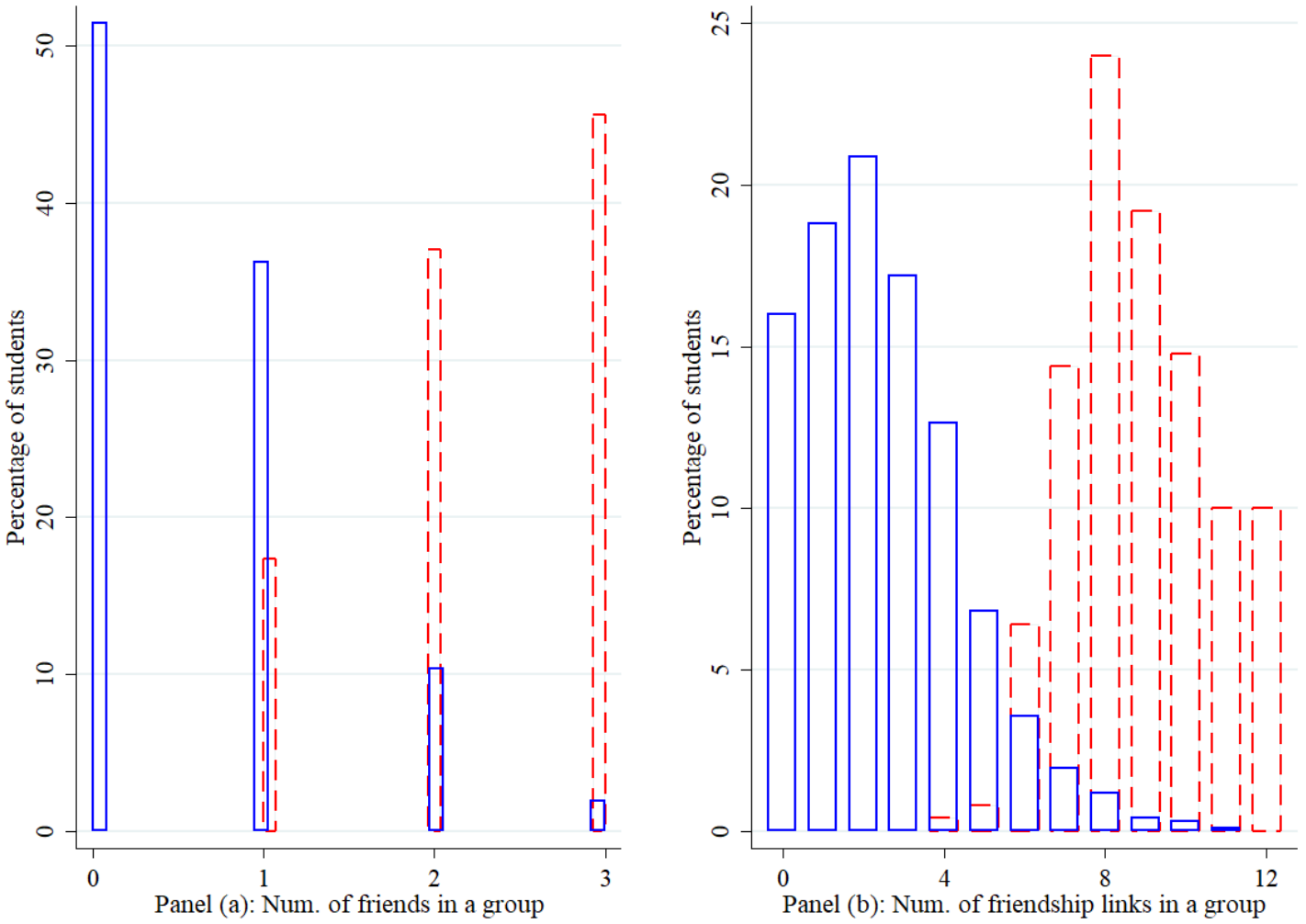
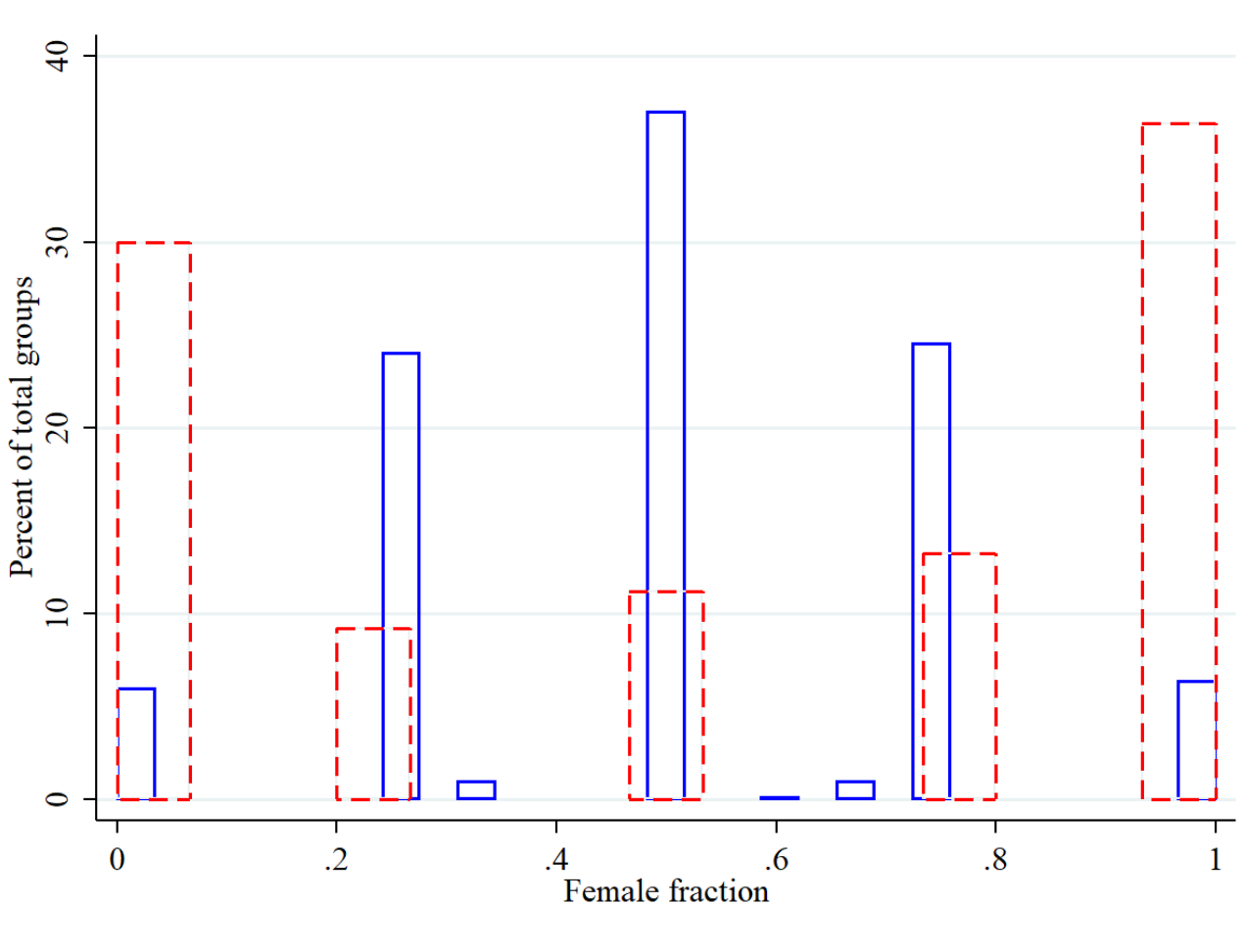


Figure 4: Distribution of students by friendship relationships in a study-group



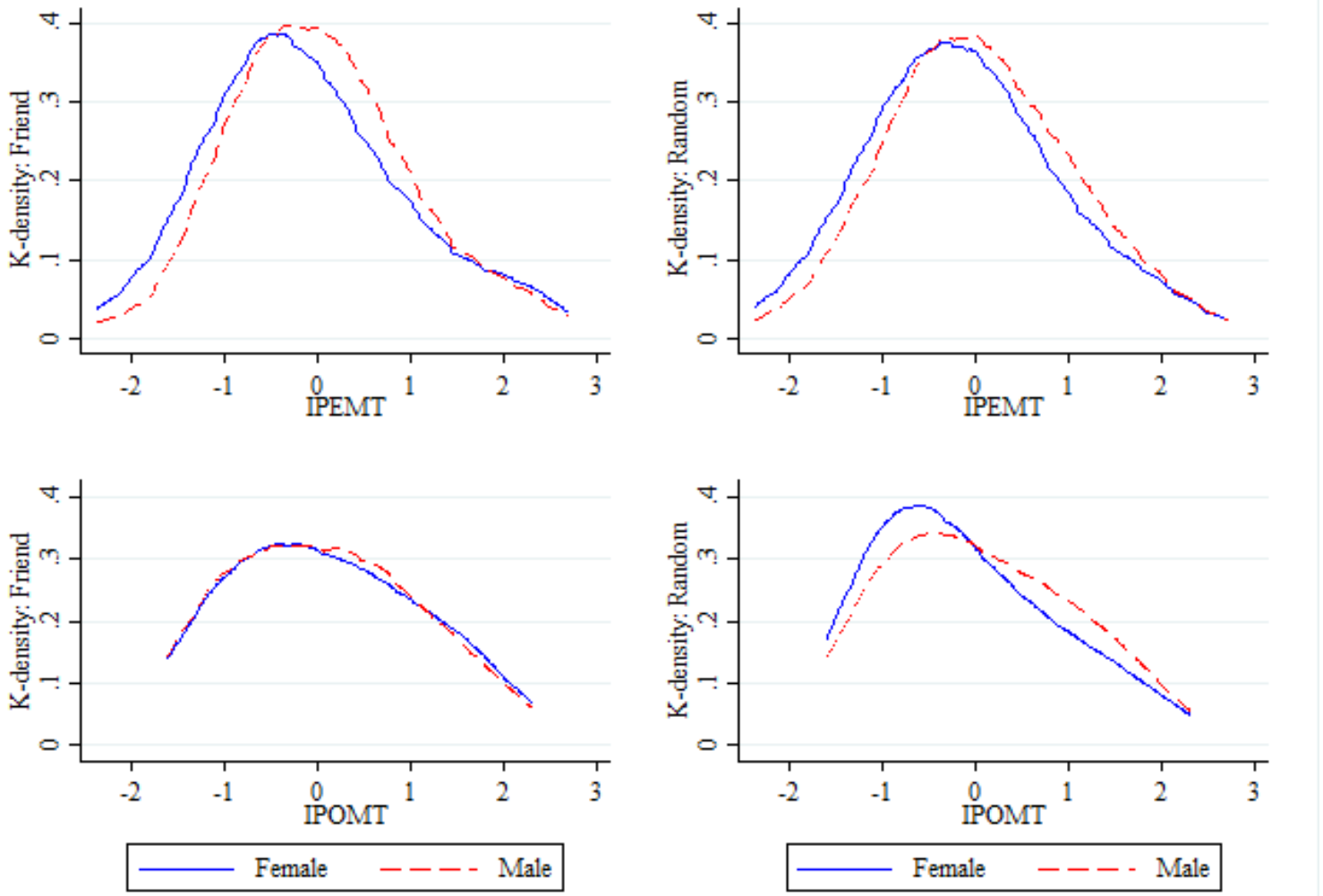
Note: Dashed red bars correspond to friendship groups and solid blue bars correspond to peer groups

Figure 5: Fraction of female students by study group



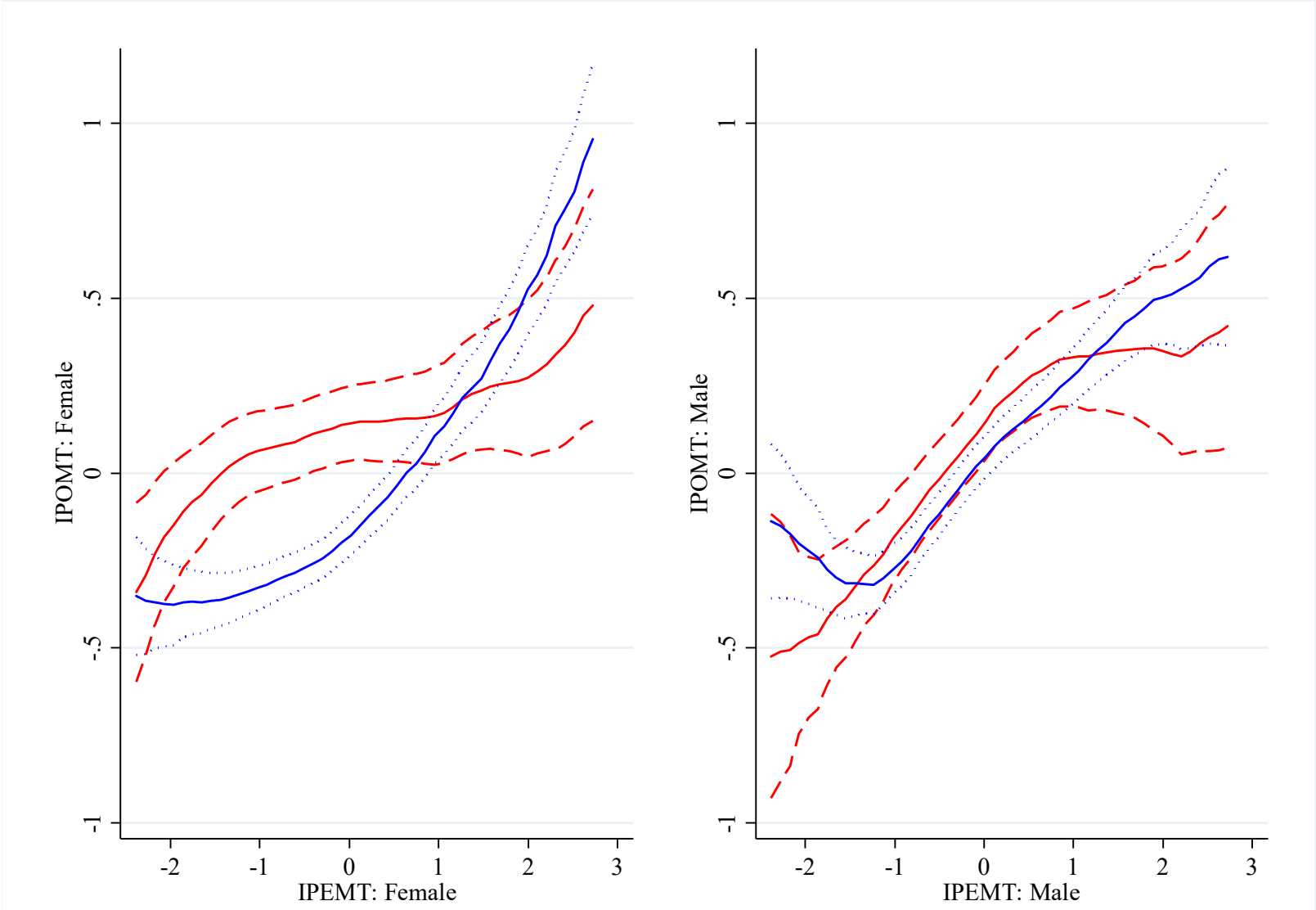
Note: Dashed red bars correspond to friendship groups and solid blue bars correspond to peer groups

Figure 6: Gender gap before and after the experiment by group type



Note: Left figure is based on friendship groups; right is based on peer groups

Figure 7: Non linear effects of groupings



Note: Local polynomial fit with 95% confidence interval is drawn (dashed-confidence interval in red is friendship grouping, dotted-confidence interval in blue is peer grouping)

Figure 8: Effect on Group general knowledge test and group math test

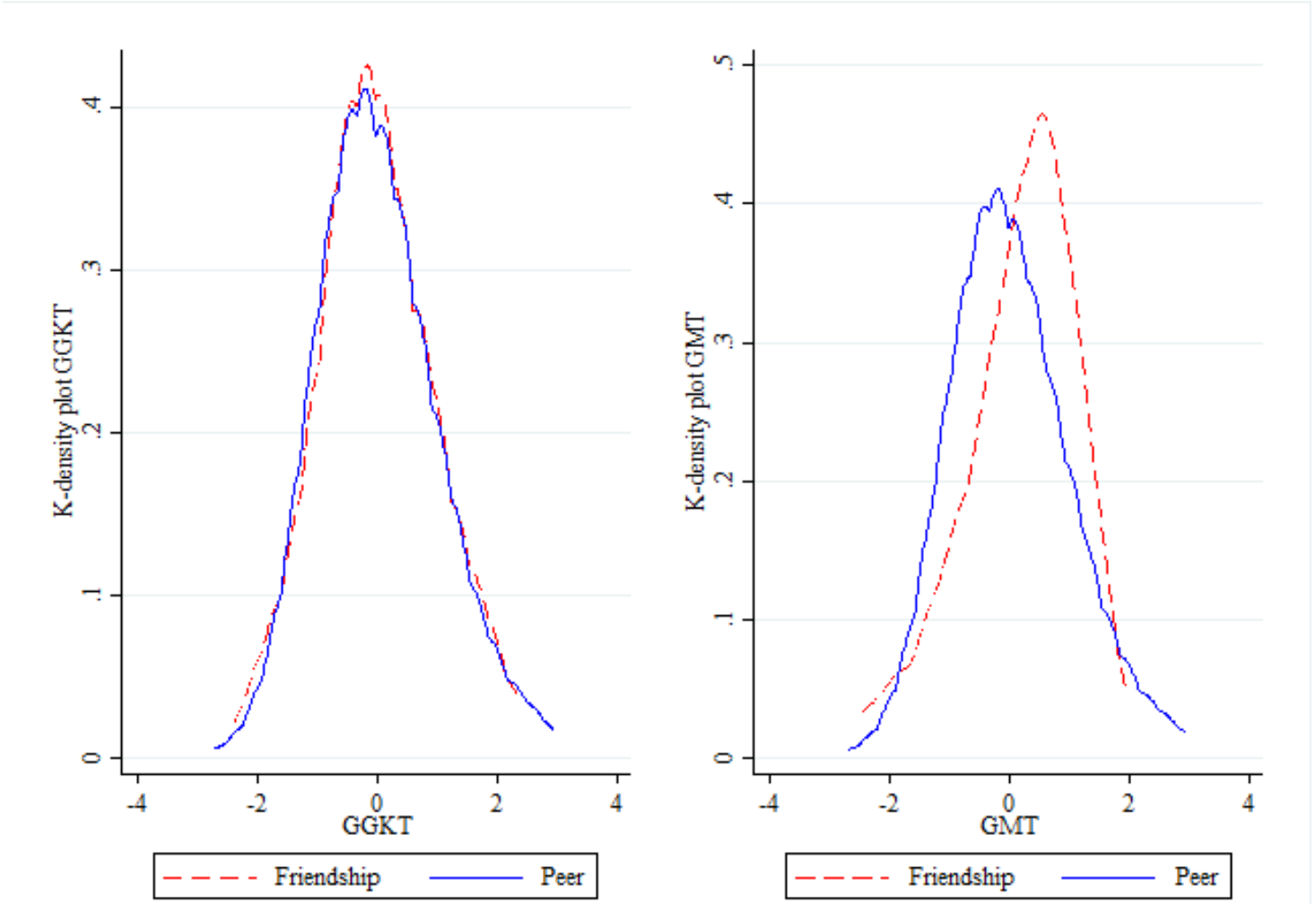


Table 1: Friendship nomination by ability, parental education, and household income

	By ability (IPEMT) (%)			By parental education (%)			By household income (%)	
				Panel A: Entire sample				
	Low	High		Low	High		Low	High
Low	65.51	34.49	Low	53.62	46.38	Low	54.26	45.74
High	32.40	67.60	High	39.78	60.22	High	46.23	53.77
				Panel B: Females				
Low	64.25	35.75	Low	52.05	47.95	Low	51.90	48.10
High	27.54	72.46	High	39.44	60.56	High	43.16	56.84
				Panel C: Males				
Low	65.62	34.38	Low	54.43	45.57	Low	57.39	42.61
High	35.89	64.11	High	38.34	61.66	High	49.19	50.81
				Panel D: Friendship groups				
Low	65.18	34.82	Low	53.95	46.05	Low	49.23	50.77
High	27.55	72.45	High	38.86	61.14	High	43.38	56.62
				Panel E: Peer groups				
Low	64.20	35.80	Low	52.80	47.20	Low	53.89	46.11
High	31.86	68.14	High	39.85	60.15	High	47.37	52.63

Note: For each variable (ability, parental education, and household income), "Low" and "High" indicate students below and above the median (50^{th} percentile) of the distribution.

Table 2: Pre-experiment gender gap in test score by group types
Dependent variable is individual pre-experiment math test (IPEMT)

	(1)	(2)
	Panel A: No controls	
Female	-0.156**	-0.169***
	(0.074)	(0.054)
	Panel B: Controls for individual characteristics	
Female	-0.153**	-0.157***
	(0.074)	(0.056)
Observations	956	3,671
Type of group	Friendship groups	Peer groups

Note: Panel A controls include only a dummy for female. Panel B controls include a dummy for female and household characteristics such as household income, parent education, parent age, and if household has access to electricity. Standard errors are clustered at the school level and are in parenthesis. * p<0.10 ** p<0.05 *** p<0.01.

Table 3: Descriptive statistics and balance checks

	Friendship groups	Peer groups	P-value of the difference
Individual pre-experiment math test (IPEMT)	-0.0429	-0.0332	0.949
	(1.027)	(1.011)	
Missing IPEMT	0.156	0.163	0.733
	(0.363)	(0.369)	
Female	0.540	0.503	0.056*
	(0.499)	(0.500)	
Household income per cap	4.387	4467.1	0.340
	(1.238)	(1519.5)	
Household has electricity	0.276	0.275	0.990
	(0.447)	(0.447)	
Parent education in years	5.078	4.923	0.630
	(3.759)	(3.740)	
Parent age	40.05	39.85	0.695
	(6.445)	(6.910)	
			0.949
Observations	956	3,671	

Note: * p<0.10 ** p<0.05 *** p<0.01. Standard deviations are shown in parenthesis. The reported p-values are based on the estimation of regression models where each characteristic is regressed on a dummy variable indicating whether a student belongs to a friendship school. Standard errors are clustered at the school level.

Table 4: Baseline results
Dependent variable is individual post-experiment math test (IPOMT)

	(1)	(2)	(3)	(4)
Friends	0.131	0.134	0.136	0.129
	(0.141)	(0.141)	(0.136)	(0.131)
Fraction female peers	-	-0.062	-0.050	-0.041
	-	(0.065)	(0.065)	(0.063)
IPEMT	-	-	0.249***	0.222***
	-	-	(0.040)	(0.037)
Observations	4627	4627	4627	4627
Other controls	No	No	No	Yes

Note: The other control variables ("Other controls") are listed in Table 3. Standard errors are clustered at the school level and are in parenthesis. * p<0.10 ** p<0.05 *** p<0.01.

Table 5: Results by gender and ability
Dependent variable is individual post-experiment math test (IPOMT)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low ability				**High ability**			
				Panel A: Females				
Friends	0.360*	0.439**	0.441**	0.433**	0.090	0.083	0.053	0.043
	(0.193)	(0.199)	(0.197)	(0.197)	(0.180)	(0.184)	(0.182)	(0.175)
FWER adjusted p-value	*0.090*	*0.035*	*0.033*	*0.036*	*0.873*	*0.924*	*0.912*	*0.864*
Fraction female peers		-0.232*	-0.244**	-0.248**		0.024	0.058	0.058
		(0.121)	(0.122)	(0.121)		(0.117)	(0.114)	(0.109)
IPEMT			0.116	0.132			0.355***	0.313***
			(0.077)	(0.081)			(0.074)	(0.072)
Observations	1203	1203	1203	1203	1159	1159	1159	1159
Other controls	No	No	No	Yes	No	No	No	Yes
				Panel B: Males				
Friends	-0.018	-0.058	-0.062	-0.082	0.081	0.077	0.084	0.096
	(0.156)	(0.165)	(0.166)	(0.159)	(0.147)	(0.161)	(0.164)	(0.155)
FWER adjusted p-value	*0.887*	*0.924*	*0.912*	*0.864*	*0.873*	*0.924*	*0.911*	*0.864*
Fraction female peers		-0.134	-0.139	-0.119		-0.012	-0.004	0.021
		(0.142)	(0.144)	(0.138)		(0.123)	(0.124)	(0.120)
IPEMT			0.174*	0.197**			0.192***	0.156**
			(0.095)	(0.090)			(0.070)	(0.066)
Observations	1000	1000	1000	1000	1265	1265	1265	1265
Other controls	No	No	No	Yes	No	No	No	Yes

Note: The control variables ("Other controls") are listed in Table 3. Standard errors are clustered at the school level and are in parenthesis. Family-wise error rate (FWER) adjusted p-value uses the free step-down resampling method of Westfall and Young (1993). * p<0.10 ** p<0.05 *** p<0.01.

Table 6: Group gender composition vs group friendship composition among female students

	(1)	(2)	(3)	(4)	(5)	(6)
	Low ability			**High ability**		
			Panel A: Same gender effect in peer groups			
All-females	-0.041	-0.047		0.014	-0.016	
	(0.112)	(0.111)		(0.120)	(0.120)	
FWER adjusted p-value	*0.914*	*0.881*		*0.914*	*0.903*	
IPEMT	0.038	0.056		0.416***	0.363***	
	(0.086)	(0.093)		(0.082)	(0.081)	
Observations	933	933		913	913	
Other controls	No	Yes		No	Yes	
			Panel B: Friendship effect in mixed-gender groups			
Friends	0.514*	0.523*	0.519*	-0.112	-0.107	-0.117
	(0.287)	(0.286)	(0.284)	(0.208)	(0.206)	(0.205)
FWER adjusted p-value	*0.089*	*0.078*	*0.076*	*0.893*	*0.867*	*0.9*
IPEMT	0.017	0.022	0.024	0.368***	0.368***	0.313***
	(0.084)	(0.085)	(0.091)	(0.076)	(0.076)	(0.076)
Fraction female peers		-0.287	-0.268		-0.069	-0.030
		(0.179)	(0.177)		(0.146)	(0.132)
Observations	896	896	896	891	891	891
Other controls	No	Yes	Yes	No	Yes	Yes
			Panel C: Friendship effect in same-gender groups			
Friends	0.379*	0.392**		0.175	0.175	
	(0.194)	(0.193)		(0.213)	(0.205)	
FWER adjusted p-value	*0.128*	*0.092*		*0.586*	*0.553*	
IPEMT	0.413***	0.446***		0.315***	0.298***	
	(0.108)	(0.102)		(0.114)	(0.111)	
Observations	307	307		268	268	
Other controls	No	Yes		No	Yes	

Note: The other control variables ("Other controls") are listed in Table 3. Standard errors are clustered at the school level and are in parenthesis. Family-wise error rate (FWER) adjusted p-value uses the free step-down resampling method of Westfall and Young (1993). * p<0.10 ** p<0.05 *** p<0.01.

Table 7: Potential channels of influence in friendship grouping

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low ability				**High ability**			
	Num Met	Num Met	Team Hrs	Team Hrs	Num Met	Num Met	Team Hrs	Team Hrs
				Panel A: Females				
Friends	0.075	-0.005	0.350	0.333	0.206	0.139	0.534**	0.456*
	(0.241)	(0.252)	(0.221)	(0.232)	(0.230)	(0.236)	(0.216)	(0.231)
FWER adjusted p-value	*0.771*	*0.978*	*0.266*	*0.374*	*0.771*	*0.859*	*0.043*	*0.173*
IPEMT	-0.102	-0.109	0.004	0.002	0.080	0.084	-0.098	-0.093
	(0.131)	(0.131)	(0.127)	(0.128)	(0.105)	(0.105)	(0.085)	(0.085)
Fraction female peers		0.237		0.049		0.209		0.244
		(0.192)		(0.217)		(0.210)		(0.199)
Observations	1203	1203	1203	1203	1159	1159	1157	1157
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
				Panel B: Males				
Friends	0.190	0.162	0.276	0.142	0.214	0.221	0.177	0.261
	(0.216)	(0.228)	(0.229)	(0.239)	(0.283)	(0.192)	(0.175)	(0.245)
FWER adjusted p-value	*0.771*	*0.859*	*0.393*	*0.548*	*0.771*	*0.743*	*0.393*	*0.450*
IPEMT	0.088	0.089	0.221*	0.225*	0.082	0.148	0.148	-0.058
	(0.125)	(0.125)	(0.121)	(0.121)	(0.137)	(0.101)	(0.100)	(0.100)
Fraction female peers		-0.092		-0.447**		0.139		0.007
		(0.208)		(0.222)		(0.188)		(0.203)
Observations	999	999	998	998	1265	1265	1264	1264
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variable 'Num met' indicates number of times met as a team; 'Team Hrs' indicates how many hours the group met as a team. The other control variables ("Other controls") are listed in Table 3. Standard errors are clustered at the school level and are in parenthesis. Family-wise error rate (FWER) adjusted p-value uses the free step-down resampling method of Westfall and Young (1993). * p<0.10 ** p<0.05 *** p<0.01.

Table 8: Are the effects stronger for groups with the strongest links?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low		**High**		**Low**		**High**	
	Females				**Males**			
Friends	0.325*	0.408**	0.033	0.019	-0.055	-0.090	0.073	0.081
	(0.187)	(0.193)	(0.174)	(0.178)	(0.142)	(0.156)	(0.145)	(0.158)
FWER adjusted p-value	*0.122*	*0.051*	*0.942*	*0.911*	*0.942*	*0.911*	*0.942*	*0.911*
Friends*All Friends	0.460	0.424	0.475	0.467	0.093	0.090	0.184	0.189
	(0.396)	(0.407)	(0.415)	(0.417)	(0.315)	(0.317)	(0.257)	(0.256)
IPEMT	0.126	0.133	0.311***	0.312***	0.196**	0.198**	0.155**	0.155**
	(0.081)	(0.081)	(0.071)	(0.072)	(0.089)	(0.090)	(0.066)	(0.066)
Fraction female peers		-0.239*		0.044		-0.119		0.027
		(0.121)		(0.109)		(0.138)		(0.119)
Observations	1203	1203	1159	1159	1000	1000	1265	1265
Other controls		Yes		Yes		Yes		Yes

Note: There was no group in random grouping schools where all members listed the others as friends, thus, the main effect of 'All Friends' is subsumed under 'Friends'. The other control variables ("Other controls") are listed in Table 3. Standard errors are clustered at the school level and are in parenthesis. Family-wise error rate (FWER) adjusted p-value uses the free step-down resampling method of Westfall and Young (1993). * p<0.10 ** p<0.05 *** p<0.01.

Table 9: Robustness check:
Using subsample of low-ability male students whose average IPEMT match those of low-ability female students.

	(1)	(2)	(3)	(4)
	IPOMT	IPOMT	IPOMT	IPOMT
Friends	-0.030	-0.045	-0.050	-0.068
	(0.155)	(0.164)	(0.165)	(0.158)
FWER adjusted p-value	*0.876*	*0.915*	*0.923*	*0.911*
Fraction female peers		-0.046	-0.053	-0.040
		(0.152)	(0.155)	(0.147)
IPEMT			0.113	0.162
			(0.102)	(0.099)
Observations	818	818	818	818
Other controls	No	No	No	Yes

Note: Column (1)-(3) replicates Table 5 results for low-ability males using sub-sample of low-ability males after dropping top 181 students among sample of low-ability male students. The other control variables ("Other controls") are listed in Table 3. Family-wise error rate (FWER) adjusted p-value uses the free step-down resampling method of Westfall and Young (1993). * p<0.10 ** p<0.05 *** p<0.01.

Table 10: Results by gender and ability - Robustness check
Dependent variable is individual post-experiment math test (IPOMT)

	(1)	(2)	(3)	(4)	(5)	(6)
	Low			**High**		
			Panel A: Females			
Friends	0.514***	0.508**	0.533***	0.058	0.131	0.134
	(0.193)	(0.195)	(0.195)	(0.175)	(0.170)	(0.170)
FWER adjusted p-value	*0.008*	*0.01*	*0.005*	*0.911*	*0.76*	*0.755*
Average IPEMT of peers	0.270***		0.232**	-0.069		-0.025
	(0.100)		(0.095)	(0.079)		(0.080)
Std. dev. IPEMT of peers		0.417*	0.171		0.391**	0.380**
		(0.226)	(0.220)		(0.170)	(0.178)
Fraction female peers	-0.220*	-0.233*	-0.218*	0.048	0.062	0.058
	(0.124)	(0.121)	(0.123)	(0.110)	(0.108)	(0.108)
IPEMT	0.082	0.187**	0.112*	0.334***	0.243***	0.253***
	(0.075)	(0.075)	(0.063)	(0.066)	(0.061)	(0.055)
Observations	1203	1203	1203	1159	1159	1159
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
			Panel B: Males			
Friends	-0.041	0.001	0.002	0.107	0.127	0.132
	(0.163)	(0.172)	(0.172)	(0.159)	(0.164)	(0.165)
FWER adjusted p-value	*0.911*	*0.999*	*0.994*	*0.851*	*0.76*	*0.755*
Average IPEMT of peers	0.161*		0.099	-0.044		-0.035
	(0.082)		(0.085)	(0.076)		(0.078)
Std. dev. IPEMT of peers		0.374*	0.266		0.109	0.094
		(0.204)	(0.221)		(0.178)	(0.184)
Fraction female peers	-0.096	-0.124	-0.108	0.016	0.021	0.017
	(0.139)	(0.134)	(0.134)	(0.115)	(0.118)	(0.115)
IPEMT	0.188**	0.253***	0.231***	0.171***	0.142**	0.156**
	(0.084)	(0.077)	(0.067)	(0.060)	(0.066)	(0.061)
Observations	1000	1000	1000	1265	1265	1265
Other controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: The other control variables ("Other controls") are listed in Table 3. Standard errors are clustered at the school level and are in parenthesis. Family-wise error rate (FWER) adjusted p-value uses the free step-down resampling method of Westfall and Young (1993). * p<0.10 ** p<0.05 *** p<0.01.

Table 11: Exploring peer effects using students in peer groups only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Females				**Males**			
	Low ability		**High ability**		**Low ability**		**High ability**	
Avg IPEMT of peers	0.313***	0.311***	-0.012	-0.013	0.179*	0.173*	-0.059	-0.056
	(0.117)	(0.117)	(0.097)	(0.097)	(0.093)	(0.094)	(0.093)	(0.092)
FWER adjusted p-value	*0.006*	*0.007*	*0.891*	*0.886*	*0.14*	*0.155*	*0.719*	*0.737*
IPEMT	0.007	0.016	0.366***	0.366***	0.156	0.160	0.215***	0.214***
	(0.089)	(0.091)	(0.076)	(0.077)	(0.095)	(0.097)	(0.072)	(0.072)
Fraction female peers		-0.199*		-0.003		-0.213		0.101
		(0.118)		(0.122)		(0.153)		(0.127)
Observations	933	933	913	913	805	805	1020	1020
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The other control variables ("Other controls") are listed in Table 3. Standard errors are clustered at the school level and are in parenthesis. Family-wise error rate (FWER) adjusted p-value uses the free step-down resampling method of Westfall and Young (1993). * p<0.10 ** p<0.05 *** p<0.01.