Do Friends Improve Female Education? The Case of Bangladesh*

Youjin Hahn† Behrooz Hassani-Mahmooei‡ Asadul Islam§
Eleonora Patacchini¶ Yves Zenou‖

October 11, 2016

Abstract

We randomly assign more than 6,000 students to work on math tests in one of three settings: individually, in groups with random mates, or in groups with friends. The groups consist of four people and are balanced by average cognitive ability and ability distribution. While the achievement of male students is not affected by the group assignment, low-ability females assigned to groups outperform low-ability females working individually. The treatment is particularly effective when low-ability females study with friends. To rule out sorting effects, we show that random groups with identical composition to that of friendship groups do not produce similar effects. Our study thus documents that there are teaching practices where mixing students by ability may improve learning, especially for low-ability female students.

JEL Classifications: E21, I25, J16, O12.

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

*We thank Vesall Nourani, Sangyoon Park, the seminar participants at Monash University, University of Queensland, Yonsei University, and the 24th Annual SJE International Symposium: Human Capital and Economic Development in Seoul for valuable comments. We are grateful for the funding supports from Monash University and AusAID (DFAT). We thank the Department of Primary Education (DPE) in the Ministry of Education of Bangladesh for its supports in conducting this project. Angela Cools, Mahbub Sarkar, Mujahid Islam, Foez Mojumdar provided excellent research assistance.

†Yonsei University, South Korea. E-mail: Youjin.Hahn@yonsei.ac.kr

‡Monash University, Australia. E-mail: behrooz.hassani.mahmooei@monash.edu.

§Monash University, Australia. E-mail: Asadul.Islam@monash.edu.

¶Cornell University, EIEF, IZA and CEPR. E-mail: ep454@cornell.edu.

‖Monash University, Australia, and IFN. E-mail: yves.zenou@monash.edu.

1 Introduction

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., 2015). However, persistently low levels of learning and a large gender gap in educational performance remain. In response to these challenges, many experimental studies have considered interventions to improve learning in developing countries (see Glewwe and Muralidharan, 2016 or Ganimian and Murnane, 2016, for detailed reviews). Most notably, pedagogical schemes based on grouping students by ability produce noticeable effects on learning levels (Duflo et al., 2011). However, a recent battery of randomized control trials implemented in primary schools in India reveal that significant effects of this teaching practice are associated with an involvement of volunteers from non-government organizations (Banerjee et al. 2015).

This paper documents alternative ways of grouping students that may aid learning. In particular, we show that mixing low- and high-ability students in small study groups with common objectives significantly improves the individual performance of low-ability female students. Such teaching practices require no guidance or monitoring from personnel outside the school.

We randomly assign more than 6,000 students from 150 schools to study in one of three settings: individually, in groups with random mates, or in groups with friends. At the beginning of the experiment, each student performs a math test to measure his or her ability. The student is then allocated to work on the math assignment in one of the three settings. The groups with random mates and groups with friends each consist of four students, and are balanced by average cognitive ability. After working for a week in his or her given setting, each student individually takes another math test, which is similar in content to the math group assignment. Our outcome variable is the difference in individual test scores before and after the experiment.

We show that, regardless of their initial ability, the gain (or loss) in math scores for male students is not affected by whether they studied by themselves, with random peers, or with friends. However, for female students, there is a significant and positive gain in math scores for the low-ability students who studied in groups. Moreover, we find that the magnitude of the effect is much higher if a low-ability female works with a group of friends rather than with a group of random peers. We show that random groups with identical composition to that of friendship groups do not produce similar effects. This additional evidence indicates that we are identifying the effects of friendship per se, rather than the effects of observable or unobservable characteristics of people who sort into the same peer group.

One of the biggest difficulties in the experimental literature is the identification of credible mechanisms through which the effects are obtained. The presence of randomized control trials gives us internally valid estimates of the effects of grouping on school performance, but does not enable us to unambiguously associate the evidence with specific drivers of individual behavior. However, our evidence is consistent with the sociology literature, which suggests that females' improvements from group work may be driven by social indispensability (that is by the feeling that people, especially friends, care about the value of their own performance for the group outcome) (see, e.g. Weber et al., 2009). This motivation might prevail in a society such as Bangladesh where women, and in particular low-ability women, may be of lower social status. In addition, psychology 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).[1] Friendship effects, however, may arise for a variety of different mechanisms that are difficult to pinpoint. Although our paper does not provide a definitive answer to the question of mechanism, it moves the literature forward by providing evidence on the effect of friendship ties on cognitive ability for disadvantaged groups, such as low-ability female students in developing countries.

Our analysis contributes to the economic development literature on the gender gap, aiming at evaluating interventions for improving female education. Although the enrollment rates of girls at the primary level have increased rapidly in most developing countries (Banerjee et al., 2015), the gender gap in enrollment and attainment are still very large (Hausmann et al., 2012; Bharadwaj et al., 2016; Muralidharan and Prakash 2016). Policies to improve female educational attainment in developing countries have mainly focused on both increasing the immediate benefits of schooling to families and on reducing the costs of attending school. The most commonly used demand-side intervention to increase female schooling has been giving conditional cash transfers (CCTs) to households for keeping girls enrolled in school. Several well-identified studies of CCT programs have found a positive impact on girls' school enrollment and attainment (for a review, see Fiszbein and Schady, 2009). On the supply side, one of the policy measures has been to improve school access by constructing more schools and thereby reducing the distance cost of attending school. For

[1] There is also a recent literature in economics looking at gender differences in cooperative environments, with mixed results (see Table 3 in Niederle, 2016). The common consensus seems to be that women have a cooperative personality that gives them a comparative advantage in contexts where such skills translate into superior outcomes for all parties (Babcock and Laschever, 2003). This is in line with the finding of our analysis. This literature, however, does not consider frienship effects.

example, it has been shown that placing schools in villages improves school enrollments for girls in Indonesia (Duflo, 2001), Afghanistan (Burde and Linden, 2013) and in Burkina Faso (Kazianga et al., 2013). Moreover, it has also been shown that recruiting female teachers has positive effects on girls' education outcomes in India (Muralidharan and Sheth, 2016).[2] Our study extends 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.

Our paper also fits to a small but rapidely growing literature that focuses on 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) find that, when a student is given monetary incentives to exercise, this student exercises more if a higher is fraction of his or her friends are also given incentives to exercise. In a field experiment setting in which workers are paid a piece rate for fruit picking, Bandiera et al. (2010) find that workers perform better when working with more able friends and perform worse when working with less able friends. Chen and Gong (2016) examine the effect of group formation on performance by randomly assigning 685 students in an undergraduate business course to one of three types of groups: groups that are assigned randomly; groups that are assigned to maximize skill complementarity; and groups that are determined by the students. They show that the members of two last groups outperform members of the first one. Park (2016) finds that workers in a seafood processing plant in Vietnam perform poorer when they work with their friends, suggesting that disruptions might be greater among friends.[3] An important role of friends for children's learning level has been recently uncovered by Lavy and Sand (2016) using administrative data for Israel. They exploit a unique feature of the Israeli school placement system, which assigns peers randomly conditional on school choice.

[2]See also Muralidharan and Prakash (2016) who study a "conditional kind transfer" program in the Indian state of Bihar that has features of both demand and supply-side interventions. Indeed, they examine a program that provided all girls who enrolled in grade 9 with funds to buy a bicycle to make it easier to access schools. They show that this program increased girls' age-appropriate enrollment in secondary school by 32% and reduced the corresponding gender gap by 40%.

[3]Using a field experiment in India, Field et al. (2016) show 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 borrow as a result of the training sessions, and they almost exclusively use the marginal loans for business purposes. More strikingly, four months later, those invited with a friend also report significantly higher household income and expenditures and are less likely to report their occupation as housewife.

Their study look 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.[4] 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.

Finally, our paper is related to the economic literature on tracking in education (for an overview, see Betts, 2011). Proponents argue that tracking can increase the efficiency of schooling by focusing on the needs of distinct groups of students. It is often recommended that low-ability students should be put together and high-ability students should also be put together.[5] For example, using cohorts of entering freshmen at the United States Air Force Academy, Carrell et al. (2013) show that low-ability individuals perform worse when working in groups with high-ability individuals than when working with other low-ability individuals. They show that this result is due to sorting; high-ability individuals do not interact with their low-ability peers assigned to the same group.[6] Our large-scale experiment, on the contrary, provides evidence that mixing students by ability can be beneficial in improving individual performance, especially for low-ability female students. There are three conditions under which mixing students by ability can be effective: (1) small groups, (2) common objectives, and (3) friendship. The results of the tracking literature mentioned above are based on groups that tend to be large (30 people), with no common assignments and without consideration as to the friendship relationships between low-ability and high-ability students.

The remainder of the paper unfolds as follows. In Section 2, we explain the institutional context and our experimental design. Section 3 is devoted to the description of our data. Our main empirical results are displayed in Section 4. Section 5 contains robustness checks. In Section 6, we explore the mechanisms underlying our results. Finally, Section 7 concludes.

[4] In the educational psychology literature, there is a longer 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, even 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).

[5] There is no clear consensus in the literature on whether tracking leads to significant achievement gains (Betts, 2011). However, as noted by Betts (2011), many studies—particularly those using U.S. data—rely on observational designs, which can be easily criticized. Two recent non-U.S. studies (Duflo et al., 2011; Vardardottir, 2013) use more rigorous designs and find positive effects on upper-tracked students.

[6] See also Feld and Zölitz (2016), who estimate peer effects in a university context in which students are randomly assigned to sections. They show that low-achieving students are harmed by high-achieving peers.

2 Institutional context and experimental design

2.1 The context

Bangladesh, like many other countries in South Asia, has traditionally been characterized by low school enrolment and gender disparity in educational achievement. According to the 1991 Census, the net primary school enrollment was 75 percent for girls and 85 percent for boys, while the net secondary school enrollment was 14 percent for girls and 25 percent for boys. 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 conditional on the children's enrollment 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 enrollment (see Hahn et al. 2016).[7]

Over the last decade, enrollment rates in primary schools have increased rapidly, leading to gender parity in enrollment, reduction in dropout, and improvement in completion of the cycle. Indicators of learning, however, remain low, in particular for females. 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 large gender gap.

2.2 The experiment

The experimental design involves within-classroom grouping among grade-four students in rural primary schools. In Bangladesh, each school has only one class for each grade and the class size is large (on average 40 students). The experiment was conducted in 150 randomly chosen schools in two districts (Khulna and Satkhira) in Bangladesh. Figure 1 shows where these regions are located. There are more than 800 primary schools in these two districts. Figure 2 shows the location of the selected schools. In total, we interviewed 6,376 students.

[*Insert Figures* 1 *and* 2 *here*]

The experiment was conducted under the direct supervision of the researchers after pre-testing and piloting in a few schools. The enumerators and the field workers who actually ran the experiments in schools were given a week-long training by the researchers. The project received enormous support from teachers and administrations.

[7] A number of NGOs have successfully implemented large scale programmes to reach out of school children.

Figure 3 shows the timing of our experiment. There are two phases in the experiment. In the first stage, we elicit friendship and household information. In the second stage, groups are formed and assignments are distributed. A cognitive ability test is conducted individually in the first stage. Learning achievements are then tested again after the treatment.

More specifically, in June 2013 (referred to as period $t-1$), we interview all students in the 150 schools. We ask them to nominate up to 10 closest friends from a school roster, and conduct a household survey where parents respond about their education, age, occupation, and other household characteristics. Each student's ability is measured using a math test (*individual pre-experiment math test*, IPEMT). 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. Children have 20 minutes to complete the test. The test is developed by local educators and experts in the field of education. A detailed description of the IPEMT is contained in the online appendix.

[*Insert Figure* 3 *here*]

In July 2013 (referred to as period t), groups of four students are formed in each school. We consider three different groups: (1) the *random-peer group*, where students are randomly allocated to a group of four within a school, regardless of friendship; (2) the *friendship group*, where students are allocated to a group of four based on friendship nominations; and (3) the *individual group*, where students are not grouped at all. We choose at random 80 schools where students are allocated into random-peer groups, 35 schools where students are allocated into friendship groups, and 35 schools where students are not grouped. Both friendship and random-peer groups are designed in a way such that students face the similar ability mix of group mates, that is, the mean and the distribution of student ability is comparable across groups. Specifically, for random-peer groups, 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 for equality in means and variance across groups are performed for three characteristics: cognitive ability (as measured by IPEMT), parental education, and household income. If similarity is confirmed, the grouping is recorded and a new classroom is considered. If one of these test fails, then the grouping is discarded and the algorithm is run again. In all classrooms, groups are formed in fewer than 10 iterations. No information on friendship links is used for the group formation of random groups.

Similarly to the algorithm for random groupings, the algorithm for friendship groupings is also run for each relevant classroom. The difference is that groups are formed using the friendship nominations and concept of *cliques* in network analysis.[8] First, the computer finds a first clique of size four, keeps it and then remove the edges (i.e. links) of the selected clique. Then, the algorithm finds a new clique of size four. It continues until there are no other cliques of size four. For the remaining students, it finds groups for which at least one student is a friend of two other students in that group, and so forth. After a grouping is achieved, the tests mentioned above for differences in terms of peers' ability, parental education, and household income across groups are performed. As in the random group case, if similarity is confirmed, the grouping is recorded and a new class is considered, otherwise the algorithm is run again. As in the case of random groups, friendship groups are formed in fewer than 10 iterations in all classrooms. In our final data, more than 97 percent of groups had four students. Out of 1,176 groups (924 random groups and 252 friendship groups), 29 groups had 3 students and 1 group had 5 students.

Newly formed groups (random and friendship groups) are then asked to solve a group general knowledge test (GGKT), which is performed immediately after groups are formed. Each group works on this test *collectively*. The GGKT consists of 20 multiple choices items that explore students' knowledge on national and international affairs, geography, current affairs, and sports. We allocate 20 minutes for groups to work on the test. Students are not informed about the test or its content before the test is administered. The purpose of this task is to help students learn to work as a group. After the GGKT is performed, each group is given a *group math test* (GMT) to be completed *collectively* outside school time and handed in after one week (referred to as period $t + 1$). This test consists of 10 questions. While the questions reflect the content in Grade 4 mathematics textbook, they are not directly taken from the textbook. To develop the test, we consider international mathematics testing (e.g., NAPLAN) for students of their age. Following NAPLAN, we present the mathematical problems to students as related to their real life contexts. The tests are developed in consultation with retired school teachers and local educational experts. A detailed description of the GGKT and GMT is contained in the online appendix. Students belonging to the individual group work on the GGKT and GMT by themselves.

At the end of the week (i.e. at $t + 1$), after each group (or individual if belonging to the individual group) has handed in its GMT, each student is asked to perform an individual post-experiment math test (IPOMT). The IPOMT is based on the GMT they completed. Although none of the test items is repeated from the GMT, the questions are similar so that it was possible for them to use what they learned from the group project (GMT). A detailed description of the IPOMT is contained in the online appendix. Students are given 1.5 hours to perform this test. Students had been informed at the beginning of the week

[8] 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).

that they would take an individual test after one week. To incentivize students to work together, they are also told that the study effort for the group project will help them to do well on the individual test. At the end of the week, students are asked to complete a short questionnaire on their group/individual study effort. The questions include (1) the number of times students met as a team (extensive margin); (2) how many hours the group met as a team (intensive margin); (3) how many hours a student spent in total doing the group math test.

Students are given prizes based on their group's performance on the different tests. For students belonging to groups (random or friendship), there is a prize for the best performing group in the GGKT. For the math tests, two prizes are given in each class: one prize for the group with the highest average score in the IPOMT (*best performing group*), and another prize for the group with *highest improvement* (between IPEMT and IPOMT) from their group average baseline math test. This prize scheme is chosen to ensure that all the students are incentivized to work together and help each other during the week. Two prizes for the math test are also given in each class for students working by themselves (individual groups): one prize for the student with the highest score in the IPOMT (best performing student), and another prize for the student with the highest improvement (between IPEMT and IPOMT). Thus, the incentive structure across school types (i.e. individual, friendship, and random) is the same.

For the group general knowledge test, the prize is a pencil box scale (ruler) for each student of the best performing group. For the best performing group in the math test IPOMT and for the highest improvement group (between IPEMT and IPOMT), students are given an instrument box (geometry box) or diary and scale. The same prize is given for individuals working by themselves for the best performing student and the highest-improvement in test score. These prizes are set in consultation with teachers and students to make sure they are incentive compatible. The cost of the prize for each student is approximately US$1. If two or more groups (or students) attain the same score, all of them receive the prizes. In our research, all participant children receive gifts (e.g., a pencil/pen) and certificates for their participation. In addition, some children receive more of the same gifts depending on their performance as described above.

3 Data description

The network survey and the household survey are administered to all students in all 150 schools, for a total of 6,376 students. As mentioned above, we ask students to nominate up to 10 closest friends from a classroom/grade roster. Figure 4 reports the distribution of students by the number of same-gender nominations. More than 50% of the students nominate more than eight friends of the same gender. The tendency to nominate mainly same-gender friends does not show, however, marked differences by gender. Gender differences are also minimally present for other drivers of friendship formation. Table 1 shows the percentage of same-type friends for cognitive ability, 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. Observe that they are below 50%, with the only exception for high-ability students. This evidence thus shows no signs of a strong tendency for homophily behaviors (McPherson et al., 2001) in dimensions different from gender.

[*Insert Table* 1 *and Figure* 4 *here*]

Panel (a) in Figure 5 depicts the distribution of students by number of friends, distinguishing between friendship and random groups. As expected, when grouping is random (in blue), most individuals end up in a group where very few students are friends. In more than 50% of the cases, a student has no friend at all. When grouping is based on friendship (in orange), the opposite is true. Panel (b) in Figure 5 shows the distribution of students by the total number of links within a group, distinguishing between random and friendship groups. For a group of 4 people, the maximum total number of links is 12. The figure confirms that, for individual in random groups, few friendship links exist while, for those in the friendship groups, the opposite occurs.

[*Insert Figure* 5 *here*]

Table 2 shows the pre-experiment gender gap in test scores (IPEMT) across group types.[9] Whatever the group, females always perform worse than males. On average, females' IPEMT scores are roughly 0.15 standard deviations below the average, and this gender gap does not close when we control for observable student characteristics such as household income and educational attainment of the parents. The gap seems, however, to be mainly driven by low-ability students. Indeed, Table 3 reports the gender gap in test scores distinguishing between low, middle and high-ability students. Using the distribution of the IPEMT for the whole sample, we define *low-ability students* as those who are at the bottom 33% of the distribution, the *high-ability students* as those who are at top 33% of the distribution, and *middle-ability students* as those in the middle of the distribution. This table reveals that female students perform worse than males, especially when they are of low-ability. Indeed,

[9] We regress the pre-experiment test (IPEMT) on a dummy variable ("Female" in the table) that takes 1 if the student is a female and 0 if it is a male, including a set of controls.

high-ability females tend to perform even better than their male counterparts, although the difference is not statistically significant.

[*Insert Tables* 2 *and* 3 *here*]

Table 4 presents summary statistics, distinguishing between the three types of groups (random, friendship and individual). Many households in this region of rural Bangladesh lack access to electricity and only about 27 percent of the sample students have access to electricity at home. Parental educational attainment is, on average, 5 years.[10] The last columns of the table formally test whether there are statistically significant differences between the different groups in terms of the observed characteristics.[11] It appears that the group types are well balanced for all individual characteristics, except for the percentage of females.[12] The friendship group schools have a slightly higher percentage of female students. In all our regressions, we will therefore control for female share in each group.

[*Insert Table* 4 *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 using individual, friendship and random group schools. The top figures show the IPEMT distributions and the bottom figures depict the IPOMT distributions. The test scores are standardized across the 150 schools so that the average value of the test score is zero with standard deviation equal to one. While male students perform better than female students before the experiment, females studying in friendship groups catch up in the post-experimental math test. The differences before and after the experiment by group type can be seen more clearly in Figure 7, which displays the distributions of the IPEMT and the IPOMT for the three different group types by gender. While the performance of boys is minimally affected by the group-type, both before and after the experiment, the performance

[10] Also, the illiteracy rate is high: about 40 percent of the parents are either illiterate or can only sign. Parental education was measured as the maximum between mother's years of education and father's years of education.

[11] 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 or a random school or the individual group. Standard errors are clustered at the school level. For instance, for the individual versus the friendship group, the p-value of the estimated coefficient on a dummy of friendship group is used when only individual and friendship groups are included in the sample.

[12] Roughly 16 percentage of students miss the IPEMT. We impute it using gender, school fixed effects, and test score of subjects in Bengali, English, Math, and Science that are 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.

of female students is clearly affected by the treatment. The upper panel shows that the pre-experiment performance of females assigned to friendship groups is roughly similar to that of females in the other groups. After the treatment, that is after having interacted for a week with peers, females having worked with friends outperform females working individually or in random groups.

[*Insert Figures* 6 *and* 7 *here*]

In Figure 8, we plot the estimated post-experimental performance against initial levels of ability allowing for non-linear effects.[13] 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- and middle-ability students. In the remainder of this paper, we further investigate these stylized facts using a more rigorous analysis.

[*Insert Figure* 8 *here*]

4 Results

We begin by looking at the effect of belonging to a study group on educational outcomes. We use the following regression model:

$$\Delta y_{irs} = \beta_0 + \beta_1 D_{\text{friend}} + \beta_2 D_{\text{random}} + \beta_3 X_{irs} + \epsilon_{irs} \quad (1)$$

where $\Delta y_{irs} = y_{irs}^{IPOMT} - y_{irs}^{IPEMT}$ is the difference in math scores between the post-experiment test (IPOMT) and 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 zero if he/she studies by him/herself, and D_{random} is a dummy variable that is equal to 1 if student irs belongs to a random group and zero if he/she studies by him/herself. 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.) and ϵ_{irs} is an error term. Standard errors are clustered at the school level.

Table 5 displays the results of estimating equation (1) with OLS. We analyze the results separately for female (panel A) and male (panel B) students, and we also separate students by their initial cognitive ability levels. Table 5 shows that, for male students, there is no effect of grouping on the gain (or loss) in math test scores. In contrast, there is a significant and positive gain in math scores for the low-ability female students who studied in groups. The fact that we find positive gains for low-ability students only is consistent with the idea

[13] We compare the different groups by gender by performing a regression where the dependent variable is the IPOMT while the independent variables are the IPEMT and the square of IPEMT. As a result, the figure depicts the predicted IPOMT for different levels of ability.

that high-ability students have less room for improvement than low-ability ones. This does not explain, however, the fact that only low-ability *female* students receive higher scores from the treatment. Figure 6 shows that the distribution of abilities of males and females is comparable, meaning that the performance of low-ability males is not higher than their female counterparts.

We also find that the magnitude of the effect is much greater if a low-ability female studies with a group of friends rather than with a random group. Indeed, compared to studying alone, studying with a group of friends increases the test scores of low-ability female students by 0.58 of a standard deviation of the IPOMT (which is standardized using the mean and standard deviation of the entire sample of students), while being in a random group increases math scores by only about 0.29 of a standard deviation.

[*Insert Table* 5 *here*]

5 Robustness checks

Our results show that low-ability female students perform better in friendship groups. However, this result must be interpreted with caution because of the endogenous nature of friendship nominations. Suppose that friends are chosen as a function of both observable and unobservable characteristics so that the probability of forming a friendship link is given by:

$$P(g_{irs,jrs} = 1 | X_{irs}, X_{jrs}, \theta_{irs}, \theta_{jrs}) = f(X_{irs}, X_{jrs}, \theta_{irs}, \theta_{jrs}),$$

where $g_{irs,jrs} = 1$ if there is a friendship relationship between individual i belonging to group r in school s and individual j belonging to group r in school s, X_{irs} and X_{jrs} are the observable characteristics of individual i and individual j, respectively, θ_{irs} and θ_{jrs} are the unobservable characteristics of individual i and individual j, respectively.

If there are some peers' characteristics that affect both friendship formation and the outcome (test score), then the correlation between those characteristics (or a function of those characteristics) and the treatment would be different from zero, that is $cor(D_{\text{friends}}, X_{jrs}) \neq 0$ and/or $cor(D_{\text{friends}}, \theta_{jrs}) \neq 0$. In other words, the effects of friendship grouping (β_1) in (1) may then simply capture those effects (spurious correlation).

To address this issue, we first consider the possible presence of *common observable characteristics*, i.e. the fact that $cor(D_{\text{friends}}, X_{jrs}) \neq 0$. Although friendship groups are balanced in terms of peers' ability, parental education and income, they are not balanced by gender (see Table 4). Because students in our context mostly name same-gender friends (see Figure 4), an explanation of our evidence may be related to the literature on single-sex schooling showing that girls do better in single-sex environments. For example, Eisenkopf et. al. (2015) document that single-sex environments may be more effective for females because they boost self-confidence. We explore this possibility in Table 6. We focus on the subsample of students in study groups only and report the results on friendship grouping when we control for the fraction of female friends. The results in Table 6 reveal that the effects of friendship grouping remains unchanged. First, we still find that male students are not affected by friendship grouping. Second, we still find that low-ability female students working in a group of friends gain the most. We also find evidence of gains from studying with friends for middle-ability females. In other words, studying with friends is always better than studying with random peers for all female students, apart from the very high-ability ones for which we find no difference.

[*Insert Table* 6 *here*]

We now address the possible presence of *common unobservable characteristics* (i.e. the fact that $cor(D_{\text{friends}}, \theta_{jrs}) \neq 0$), since one may be worried that there may be unobservable characteristics driving both performance and friendship formation. To address this issue, we conduct the following placebo tests. For each gender and ability level, we create "fake" friendship groups by using random groups students who have a similar empirical distributions of the observable characteristics to that of the friendship groups. We consider four characteristics: the fraction of females, IPEMT scores, parental education, and household income. In other words, we create "fake" friendship groups that have the same characteristics as "real" friendship groups but whose members are not "real" friends. For the fraction of females, which takes a discrete value, we match the exact distribution. For the other characteristics (IPEMT, parental education and household income), which take continuous values, we match the quartiles of the group average distribution. We then focus on the subsample of students in study groups and run the following regression:

$$\Delta y_{irs} = \gamma_0 + \gamma_1 D_{\text{Placebo Friends}} + \gamma_3 X_{irs} + \epsilon_{irs} \tag{2}$$

where $D_{\text{Placebo Friends}}$ is a dummy variable that is equal to 1 if individual irs belong to the group of "fake" friends and zero if he/she belongs to a group of random peers. If our estimates of β_1 in (1) simply capture the unobserved group environment characteristics, then these regressions should show a statistical significant effect for γ_1. If, on the contrary, our estimates capture the effects of friendship, then we should not find any effect of random peers behavior on own outcomes in these placebo regressions. The results of these regressions are displayed in Table 7. One can see that none of the effects is statistically significant, suggesting that our friendship grouping dummy D_{friend} is not simply picking up unobserved friends' characteristics.

[*Insert Table 7 here*]

6 Inspecting the mechanisms

In Table 5, we have found two main results:

(*i*) **Result 1:** *Low-ability female students perform better when studying in groups than when studying by themselves.*

(*ii*) **Result 2:** *Low-ability female students perform better when studying in friendship groups than when studying in random groups.*

Let us now investigate the possible mechanisms underlying these results. As mentioned in the Introduction, this is a difficult task, even in presence of a randomized control trial.

A theory consistent with Result 1 can be found in the sociology 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). This motivation might prevail in a society such as Bangladesh where women may be of lower social status, especially low-ability female students. In addition, psychology research suggests that females may care more than males about collective outcomes, and thus may be less likely to exert less effort when they work in a group than when they work alone (i.e. to engage in social loafing; see e.g. Karau and Williams, 1993). Since groups are balanced by ability, low-ability females benefit the most from being in a group because they interact with higher-ability students. Since groups are small (the size is four), there cannot be sorting in which high-ability students do not talk to low-ability students. Since groups perform common assignments, it is in the interest of the high-ability students that the outcomes of the two common assignments (Group General Knowledge Test (GGKT) and Group Math Test (GMT)) are good.

The literature mentioned above considers how groupings of students potentially changes performance incentives. Group study can 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 with others in his/her 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 classroom setting without groups. 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 not (or even regardless of whether they are in a group). This theory may explain why low-ability female students tend to perform better in friendship groups (Result 2).

An alternative story for Result 2 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 compared to female students in random groups. The post-experiment survey gives us the ability to consider and rule out this possibility. We compare the effort of students working in random groups and in friendship groups using the following regression model:

$$INT_{irs} = \beta_0 + \beta_1 D_{\text{friend}} + \beta_2 F_{irs} + \beta_3 X_{irs} + \epsilon_{irs} \tag{3}$$

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) or how many hours a student has spent in total doing the Group Math Test (HW Hrs). F_{irs} is the fraction of female peers in the group. All the other variables have the same interpretation as in (2). The results are displayed in Table 8. This table shows no differences in frequency of interactions or study time between random and friendship groups, with the exception of high-ability females in friendship groups who study more hours with group members compared to their counterpart in random groups. Therefore, our results showing that low-ability girls tend to perform better with friends than with random peers seem to indicate that improved learning is taking place between friends.

[*Insert Table* 8 *here*]

Given the experimental nature of our data and having ruled out a variety of possible confounding factors, the interpretation of Result 2 rests on different gains from studying with friends between males and females, which are especially high for low-ability females. According to the above-mentioned social psychology literature, the motivation gains when working in groups of females and that of less-capable group members are highest in cohesive groups and when groups have stronger agreements (Karau and Hart, 1998). Turning to our data, we thus gather additional evidence that may be helpful to understand the validity of these theories in our context.

In the social network literature, the number of friends and links in a group can also be considered as a measure of group cohesion (see e.g. Jackson, 2008 or Jackson et al., 2016). In Table 9, we thus investigate whether the number of friends and the number of links in the study group matter for the individual performance of the group members. We consider as additional explanatory variables for Δy_{irs} both the number of friends and the number of links in a group. The results indicate that male students are not affected by these variables while female students are. Interestingly, when we control for both variables (see columns (3) and (6)), the overall environment of friendship (links across all group members) in a group appears to be more important than the number of direct friends in the group. Such additional evidence is in line with the postulated mechanism.

[*Insert Table* 9 *here*]

7 Concluding remarks

Fighting low levels of basic education in developing countries is a priority for economic development. Among the plethora of on-going interventions in many developing countries, the experiment reported here provides evidence on the effectiveness of teaching practices based on a novel grouping scheme in the context of Bangladeshi primary schools. This pedagogical scheme consists of assigning children to study for a common goal in small teams. These teams are balanced by ability and, sometimes, consist of friends. The practice is inexpensive and does not require involvement of personnel outside the school. The results reveal important gender differences in the responsiveness of the children to the treatments. In particular, we identify an important effect of studying with friends on low-ability females, a group that typically performs well below grade level. Our field experiment is potentially of great importance for educational policies in developing countries and shows the possibility for effective improvements in learning through inexpensive teaching practices during class time.

References

[1] Andersson, J. (2001), "Net effect of memory collaboration: How is collaboration affected by factors such as friendship, gender and age?" *Scandinavian journal of Psychology* 42, 367-375.

[2] Andersson, J. and J. Rönnberg (1995), "Recall suffers from collaboration: Joint recall effects of friendship and task complexity," *Applied Cognitive Psychology* 9, 199-211.

[3] Andrabi, T., Das, J., Khwaja, A.I., Vishwanath, T. and T. Zajonc (2007), "Learning and educational achievements in Punjab schools (leaps): Insights to inform the education policy debate," Washington, DC: World Bank.

[4] Babcock, L. and S. Laschever (2003), *Women Don't Ask*, Princeton: Princeton University Press.

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

[6] Bandiera, O., Barankay, I. and I. Rasul (2010), "Social incentives in the workplace," *Review of Economic Studies* 77, 417-458.

[7] Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M. and M. Walton (2015), "Teaching at the right level: Evidence from randomized evaluations in India," Unpublished manuscript, MIT.

[8] Betts, J.R. (2011), "The economics of tracking in education," In: E.A. Hanushek, S. Machin and L. Woessmann (Eds.), *Handbook of the Economics of Education, Vol. 3*, Amsterdam: Elsevier Science, pp. 341-381.

[9] Bharadwaj, P., De Giorgi, G., Hansen, D. and C. Neilson (2016), "The gender gap in mathematics: Evidence from low- and middle-income countries," *Economic Development and Cultural Change*, forthcoming.

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

[11] Carrell, S.E., Sacerdote, B. and J.E. West (2013), "From natural variation to optimal policy? The importance of endogenous peer group formation," *Econometrica* 81, 855-882.

[12] Chen, R. and J. Gong (2016), "Group formation and performance: Field experimental evidence," Unpublished manuscript, National University of Singapore.

[13] Duflo, E. (2001), "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment," *American Economic Review* 91, 795-813.

[14] Duflo, E., Dupas, P. and M. Kremer (2011), "Peer effects and the impacts of tracking: Evidence from a randomized evaluation in Kenya," *American Economic Review* 101, 1739-1774.

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

[16] Epstein, J.L. and D.J. MacIver (1990), *Education in the Middle Grades: National Practices and Trends*, Columbus, OH: National Middle School Association.

[17] Feld, J. and U. Zölitz (2016), "Understanding peer effects – On the nature, estimation and channels of peer effects," *Journal of Labor Economics*, forthcoming.

[18] Field, E., Jayachandran, S., Pande, R. and N. Rigol (2016), "Friendship at work: Can peer effects catalyze female entrepreneurship?" *American Economic Journal: Economic Policy* 8, 125-153.

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

[20] Foot, H. and A.M. Barron (1990), "Friendship and task management in children's peer tutoring," *Educational Studies* 16, 237-250.

[21] Ganimian, A.J. and R.J. Murnane (2016), "Improving educational outcomes in developing countries: Lessons from rigorous impact evaluations," *Review of Educational Research* 86, 719-755.

[22] Glewwe, P. and K. Muralidharan (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*, Amsterdam: Elsevier Science, pp. 653-744.

[23] Hahn, Y., Islam, A., Nuzhat, K., Smyth, R. and H.-S. Yang (2016), "Education, marriage and fertility: Long-term evidence from a female stipend program in Bangladesh," *Economic Development and Cultural Change*, forthcoming.

[24] Harrison, D.A., Mohammed, S., McGrath, J.E., Florey, A.T. and S.W. Vanderstoep (2003), "Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality," *Personnel Psychology* 56, 633-669.

[25] Hausmann, R., Tyson, L.D. and S. Zahidi (2012), *The Global Gender Gap Report 2012*, Geneva: World Economic Forum.

[26] Hertel, G., Kerr, N.L. and L.A. Messé (2000), "Motivation gains in performance groups: Paradigmatic and theoretical developments on the Koehler effect," *Journal of Personality and Social Psychology* 79, 580-601.

[27] Hertel, G., Niemeyer, G. and A. Clauss (2008), "Social indispensability or social comparison: The why and when of motivation gains of inferior group members," *Journal of Applied Social Psychology* 38, 1329-1363.

[28] Jackson M.O. (2008), *Social and Economic Networks*, Princeton, NJ: Princeton University Press.

[29] Jackson, M.O., Rogers, B. and Y. Zenou (2016), "The economic consequences of social network structure," *Journal of Economic Literature*, forthcoming.

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

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

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

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

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

[35] Lavy, V. and E. Sand (2016), "The effect of social networks on student's academic and non-cognitive behavioral outcomes: Evidence from conditional random assignment of friends in school," Unpublished manuscript, University of Warwick.

[36] MacDonald, R., Miell, D. and L. Morgan (2000), "Social processes and creative collaboration in children," *European Journal of Psychology of Education* 15, 405-415.

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

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

[39] Muralidharan, K. and N. Prakash (2016), "Cycling to school: Increasing secondary school enrollment for girls in India," Unpublished manuscript, University of California, San Diego.

[40] Muralidharan, K. and K. Sheth (2016), "Bridging education gender gaps in developing countries: The role of female teachers," *Journal of Human Resources* 51, 269-297.

[41] Niederle, M. (2016), "Gender," In: J.H. Kagel and A.E. Roth (Eds.), *Handbook of Experimental Economics, Vol. 2*, Princeton: Princeton University Press, pp. 481-562.

[42] Park, S. (2016), "Socializing at work: Evidence from a field experiment with manufacturing workers," Unpublished manuscript, Northwestern University.

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

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

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

[46] Vardardottir, A. (2013), "Peer effects and academic achievement: A regression discontinuity Approach," *Economics of Education Review* 36, 108-121.

[47] Weber, B., Wittchen, M. and G. Hertel (2009), "Gendered ways to motivation gains in groups," *Sex Roles* 60, 731-44.

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

	By ability (%)				By parental education (%)				By household income (%)		
Panel A: Entire sample											
	Low	Mid	High		Low	Mid	High		Low	Mid	High
Low	46.75	27.84	25.41	Low	39.71	32.67	27.63	Low	37.43	35.98	26.59
Mid	29.61	33.41	36.98	Mid	29.10	40.95	29.95	Mid	30.56	39.79	29.64
High	15.53	20.84	63.63	High	25.60	31.93	42.48	High	27.36	33.74	38.90
Panel B: Females											
	Low	Mid	High		Low	Mid	High		Low	Mid	High
Low	49.99	27.04	22.97	Low	39.93	31.84	28.24	Low	39.44	35.57	24.99
Mid	32.48	32.34	35.17	Mid	29.08	41.34	29.57	Mid	31.77	42.22	26.01
High	16.44	20.32	63.24	High	26.73	31.70	41.57	High	29.54	35.91	34.55
Panel C: Males											
	Low	Mid	High		Low	Mid	High		Low	Mid	High
Low	42.33	28.93	28.74	Low	39.46	33.61	26.93	Low	34.96	36.48	28.57
Mid	26.78	34.46	38.76	Mid	29.12	40.54	30.35	Mid	29.27	37.19	33.54
High	14.73	21.31	63.97	High	24.36	32.17	43.46	High	25.17	31.56	43.27
Panel D: Friends (friendship group)											
	Low	Mid	High		Low	Mid	High		Low	Mid	High
Low	49.29	26.63	24.08	Low	41.05	30.12	28.83	Low	38.37	40.07	21.57
Mid	30.36	33.09	36.56	Mid	28.93	39.74	31.32	Mid	33.92	40.92	25.16
High	15.84	22.95	61.22	High	25.90	30.66	43.45	High	26.68	39.18	34.14
Panel E: Random (random-peer group)											
	Low	Mid	High		Low	Mid	High		Low	Mid	High
Low	45.99	28.20	25.81	Low	39.37	33.30	27.33	Low	37.18	34.87	27.95
Mid	29.41	33.50	37.10	Mid	29.14	41.25	29.61	Mid	29.58	39.46	30.96
High	15.45	20.27	64.27	High	25.50	32.32	42.17	High	27.51	32.51	39.97

Table 2: Pre-experiment gender gap in test score by group types
(Dependent variable = IPEMT)

	(1)	(2)	(3)
	Panel A: No controls		
Female	-0.153**	-0.146*	-0.169***
	(0.061)	(0.077)	(0.054)
	Panel B: Controls for individual characteristics		
Female	-0.148**	-0.147*	-0.157***
	(0.060)	(0.076)	(0.056)
Observations	1,660	1,005	3,671
Type of group	Individual	Friend	Random

Notes: Panel A controls include only a dummy for female. Panel B controls include a dummy for female and other individual characteristics as defined in the text. Standard errors are clustered at the school level and are in parenthesis. * p<0.10 ** p<0.05 *** p<0.01.

Table 3: Pre-experiment gender gap in test score by cognitive ability
(Dependent variable = IPEMT)

	(1)	(2)	(3)
		Panel A: No controls	
Female	-0.064***	-0.003	0.042
	(0.023)	(0.012)	(0.031)
		Panel B: Controls for individual chars	
Female	-0.058***	-0.003	0.038
	(0.021)	(0.012)	(0.031)
Observations	1,982	2,527	1,827
	Low	Mid	High

Notes: Panel A controls include only a dummy for female. Panel B controls include a dummy for female and other individual characteristics as defined in the text. “Low” indicates the students from bottom 33 percentiles of IPEMT; “Mid” indicates students from 34-72 percentiles of IPEMT; “High” indicates students who are from 72-100 percentiles of IPEMT distribution. Standard errors are clustered at the school level and are in parenthesis. * p<0.10 ** p<0.05 *** p<0.01.

Table 4: Descriptive statistics and balance checks

	Random	Friendship	Individual	p-value of the difference: Friendship vs. random	Individual vs. friendship	Individual vs. random
Individual pre-experiment math test (IPEMT)	-0.0332	-0.0451	0.131	0.937	0.303	0.194
	(1.011)	(1.030)	(0.972)			
Missing IPEMT	0.163	0.152	0.163	0.593	0.601	0.998
	(0.369)	(0.359)	(0.369)			
Female	0.503	0.545	0.504	0.024**	0.038**	0.973
	(0.500)	(0.498)	(0.500)			
Household income per cap	4467.1	4422.5	4507.1	0.605	0.428	0.665
	(1519.5)	(1390.9)	(1469.2)			
Household has electricity	0.275	0.276	0.265	0.999	0.869	0.853
	(0.447)	(0.447)	(0.442)			
Parent education in years	4.923	5.142	4.777	0.494	0.298	0.625
	(3.740)	(3.768)	(3.825)			
Parent age	39.85	40.09	40.44	0.631	0.565	0.252
	(6.910)	(6.444)	(7.001)			
Obs	3,671	1,005	1,660			
Number of groups	924	252				
Number of classrooms (schools)	80	35	35			

Note: * p<0.10 ** p<0.05 *** p<0.01. Standard deviations are shown in parenthesis

Table 5: Do students perform better when studying with peers than studying alone?
Dependent variable is gain in test score (IPOMT-IPEMT)

	(1)	(2)	(3)	(4)	(5)	(6)
	Low		Mid		High	
	Panel A: Females					
Random	0.290*	0.314*	-0.117	-0.126	0.062	0.037
	(0.166)	(0.161)	(0.138)	(0.134)	(0.199)	(0.196)
Friends	0.583***	0.590***	0.233	0.212	-0.025	-0.053
	(0.209)	(0.210)	(0.194)	(0.189)	(0.286)	(0.273)
Observations	1,131	1,131	1,261	1,261	843	843
	Panel B: Males					
Random	0.048	0.085	-0.092	-0.092	0.120	0.116
	(0.158)	(0.149)	(0.154)	(0.148)	(0.205)	(0.203)
Friends	-0.011	0.025	0.049	0.015	0.130	0.127
	(0.187)	(0.180)	(0.184)	(0.172)	(0.286)	(0.285)
Observations	851	851	1,266	1,266	984	984
Included controls	No	Yes	No	Yes	No	Yes

Note: "Low" indicates the students from bottom 33 percentiles of IPEMT; "Mid" indicates students from 34-72 percentiles of IPEMT; "High" indicates students who are from 72-100 percentiles of IPEMT distribution. See text for the included control variables. Standard errors are clustered at the school level and are in parenthesis. * p<0.10 ** p<0.05 *** p<0.01.

Table 6: Do students perform better when studying with friends than studying with randomly assigned peers?

Dependent variable is gain in test score (IPOMT-IPEMT)

	(1) Low	(2) Low	(3) Low	(4) Mid	(5) Mid	(6) Mid	(7) High	(8) High	(9) High
				Panel A: Females					
Friends	0.293	0.278	0.403**	0.351**	0.334**	0.365**	-0.088	-0.094	-0.200
	(0.188)	(0.187)	(0.195)	(0.171)	(0.163)	(0.167)	(0.254)	(0.240)	(0.249)
Fraction of female peers			-0.343***			-0.106			0.320*
			(0.130)			(0.118)			(0.173)
Observations	896	896	896	887	887	887	611	611	611
				Panel B: Males					
Friends	-0.059	-0.060	-0.072	0.141	0.111	0.090	0.010	0.009	0.047
	(0.179)	(0.168)	(0.182)	(0.158)	(0.148)	(0.158)	(0.248)	(0.245)	(0.251)
Fraction of female peers			-0.041			-0.074			0.115
			(0.174)			(0.131)			(0.173)
Observations	661	661	661	916	916	916	705	705	705
Included control	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: “Low” indicates the students from bottom 33 percentiles of IPEMT; “Mid” indicates students from 34-72 percentiles of IPEMT; “High” indicates students who are from 72-100 percentiles of IPEMT distribution. See text for the included control variables. Standard errors are clustered at the school level and are in parenthesis. * p<0.10 ** p<0.05 *** p<0.01.

Table 7: Placebo tests

Dependent variable is gain in test score (IPOMT-IPEMT)

	(1) Low	(2) Mid	(3) High	(4) Low	(5) Mid	(6) High
		Panel A: Female			Panel B: Male	
	Panel A: Random groups with the same group composition on fraction of female students as friendship groups					
Placebo Friends	-0.062	-0.064	0.031	0.135	-0.065	-0.083
	(0.096)	(0.068)	(0.097)	(0.098)	(0.091)	(0.133)
	Panel B: Random groups with the same group average IPEMT as friendship groups					
Placebo Friends	0.036	-0.002	-0.030	0.036	0.010	-0.012
	(0.034)	(0.036)	(0.033)	(0.041)	(0.035)	(0.044)
	Panel C: Random groups with the same group composition on parental education as friendship groups					
Placebo Friends	-0.045	0.025	0.016	-0.005	-0.055	-0.055
	(0.049)	(0.034)	(0.042)	(0.048)	(0.035)	(0.051)
	Panel D: Random groups with the same group composition on household income as friendship groups					
Placebo Friends	0.062	0.017	0.024	-0.030	-0.031	-0.049
	(0.040)	(0.034)	(0.037)	(0.045)	(0.032)	(0.037)

Note: "Placebo Friends" is a dummy variable indicating the students from random group schools selected to create placebo friendship groups. These placebo groups resemble empirical distribution of the friendship groups by the selected criteria, as explained in each panel heading. "Low" indicates the students from bottom 33 percentiles of IPEMT; "Mid" indicates students from 34-72 percentiles of IPEMT; "High" indicates students who are from 73-100 percentiles of IPEMT distribution. We control for individual characteristics as well as fraction of female peers in the group. See text for the included control variables. Standard errors are clustered at the school level and are in parenthesis. * p<0.10 ** p<0.05 *** p<0.01.

Table 8: Potential Channels of Influence in Friendship Grouping

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Num Met	Team Hrs	HW Hrs	Num Met	Team Hrs	HW Hrs	Num Met	Team Hrs	HW Hrs
		Low			Mid			High	
					Panel A: Females				
Friend	-0.062	0.253	-0.001	0.044	0.340	-0.091	0.327	0.629**	0.254
	(0.266)	(0.239)	(0.281)	(0.260)	(0.247)	(0.305)	(0.278)	(0.255)	(0.377)
Observations	896	896	895	887	886	886	611	610	611
					Panel B: Males				
Friend	0.185	0.067	0.216	0.134	0.333	0.112	0.296	0.100	0.591*
	(0.237)	(0.268)	(0.337)	(0.233)	(0.230)	(0.276)	(0.217)	(0.312)	(0.316)
Observations	661	660	660	915	914	916	705	705	705

Notes: The dependent variable for col. (1), (4) and (7) indicates number of times met as a team (Num met); (2), (5) and (8) indicates how many hours the group met as a team (Team Hrs); (3), (6) and (9) how many hours a student spent in total doing the Group math test (HW Hrs); “Low” indicates the students from bottom 33 percentiles of IPEMT; “Mid” indicates students from 34-72 percentiles of IPEMT; “High” indicates students who are from 73-100 percentiles of IPEMT distribution. We control for individual characteristics as well as fraction of female peers in the group. See text for the included control variables. Standard errors are clustered at the school level and are in parenthesis.
* p<0.10 ** p<0.05 *** p<0.01.

Table 9: Alternative definitions of friendship relationships

Dependent variable is gain in test score (IPOMT-IPEMT)

	(1) Low	(2) Low	(3) Low	(4) Mid	(5) Mid	(6) Mid	(7) High	(8) High	(9) High
					Panel A: Females				
Num. of friends in a group	0.123**		-0.006	0.104*		-0.002	-0.042		-0.021
	(0.055)		(0.044)	(0.054)		(0.054)	(0.065)		(0.073)
Num. of links in a group		0.051***	0.053***		0.042**	0.043**		-0.014	-0.009
		(0.018)	(0.019)		(0.018)	(0.020)		(0.025)	(0.031)
Observations	896	896	896	887	887	887	611	611	611
					Panel B: Males				
Num. of friends in a group	-0.014		-0.017	0.013		-0.017	0.023		-0.042
	(0.059)		(0.053)	(0.052)		(0.047)	(0.077)		(0.074)
Num. of group links in a group		-0.003	0.001		0.008	0.012		0.017	0.027
		(0.022)	(0.025)		(0.018)	(0.020)		(0.026)	(0.027)
Observations	661	661	661	916	916	916	705	705	705

Note: "Low" indicates the students from bottom 33 percentiles of IPEMT; "Mid" indicates students from 34-72 percentiles of IPEMT; "High" indicates students who are from 73-100 percentiles of IPEMT distribution. We control for individual characteristics as well as fraction of female peers in the group. See text for the included control variables. Standard errors are clustered at the school level and are in parenthesis. * p<0.10 ** p<0.05 *** p<0.01.

Figure 1: Regions where the experiment was conducted

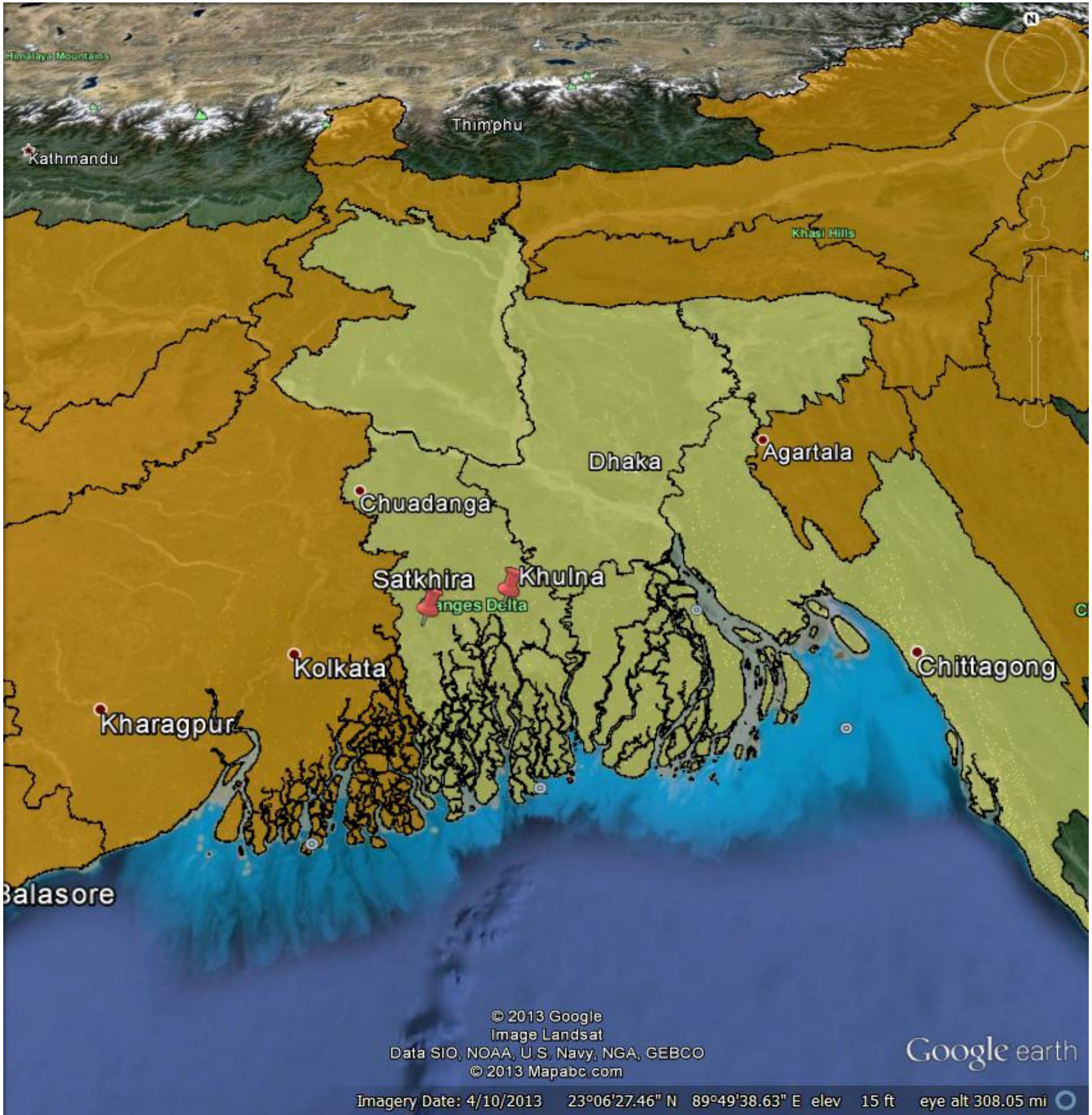


Figure 2: Location of the different schools

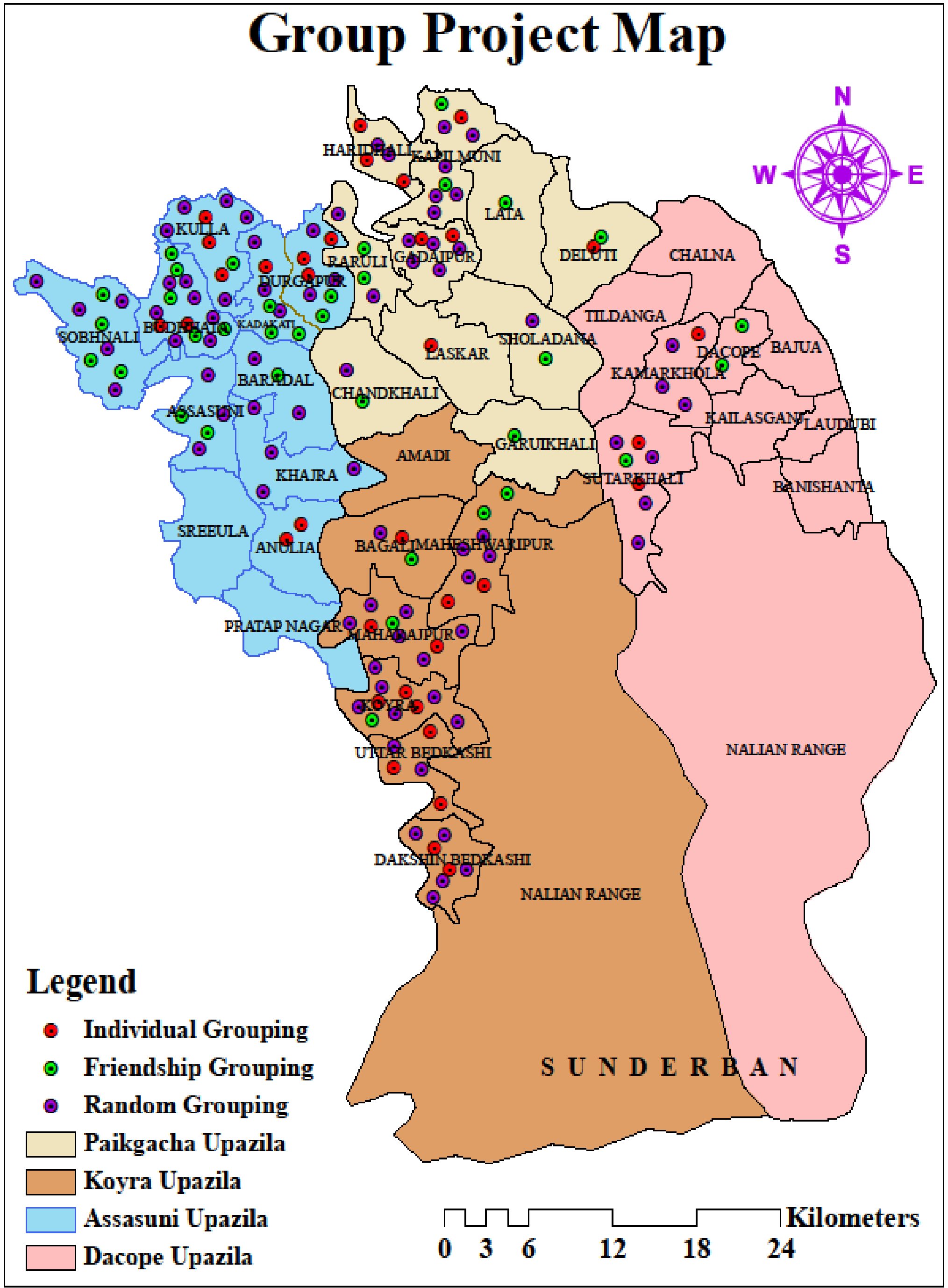


Figure 3: Timeline of the experiment

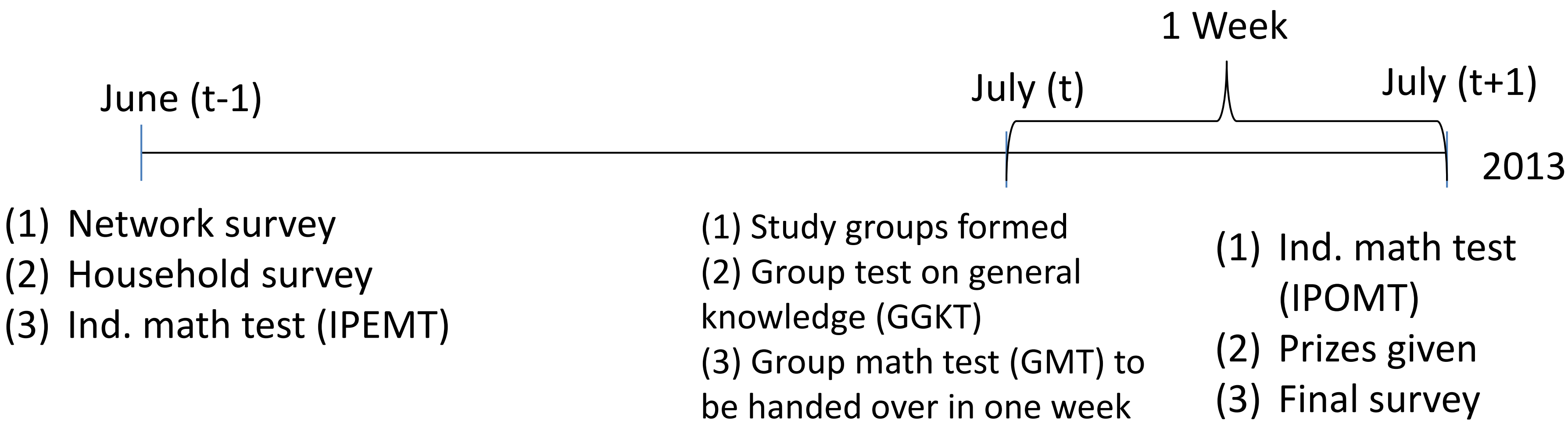


Figure 4: Distribution of students by same-gender friendship nomination

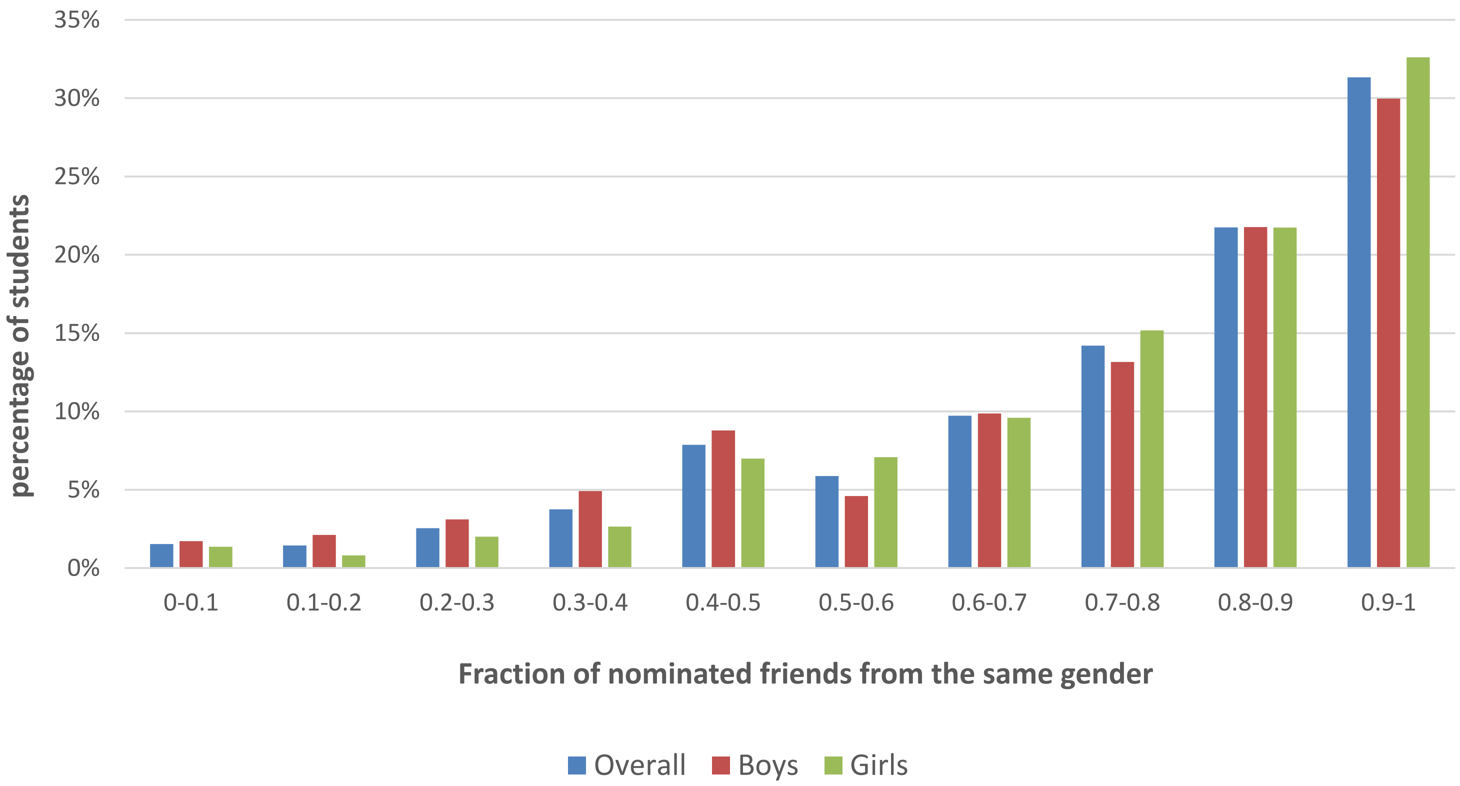


Figure 5: Distribution of students by friendship relationships in a study-group

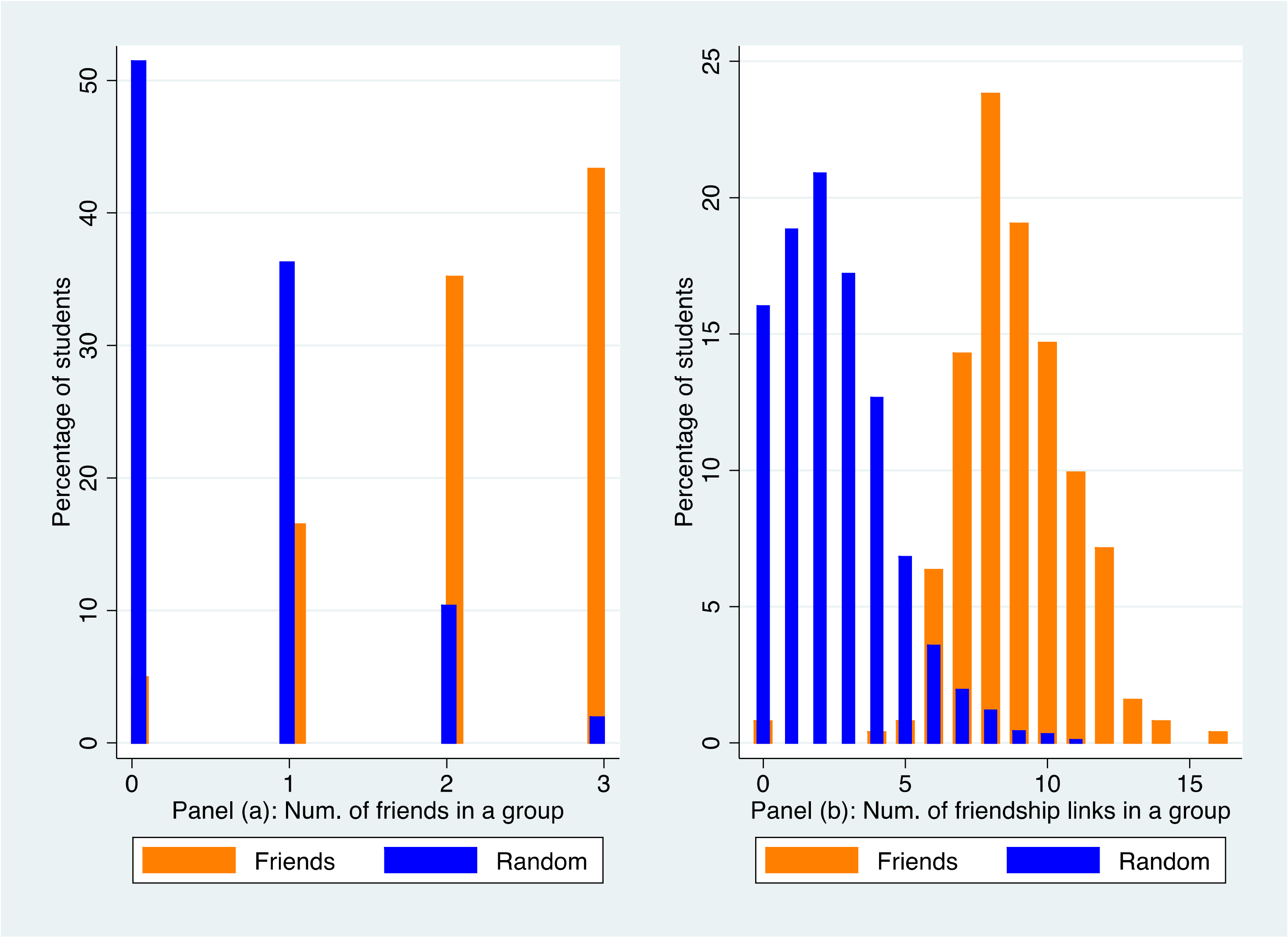
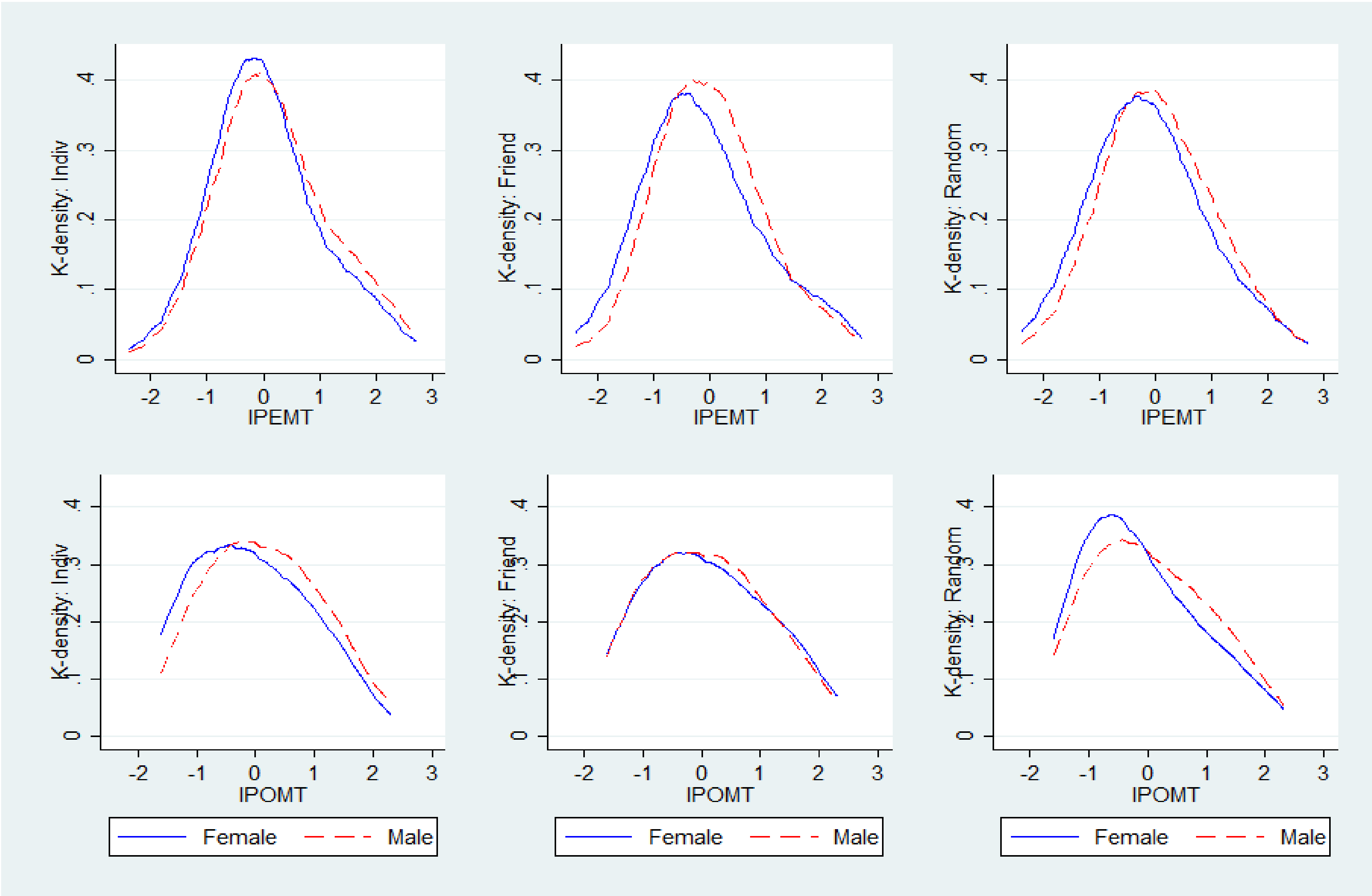


Figure 6: Gender gap before and after the experiment by group type



Note: Left figure is based on individual group; middle is based on friendship; right is based on random

Figure 7: Group-type gap before and after the experiement by gender

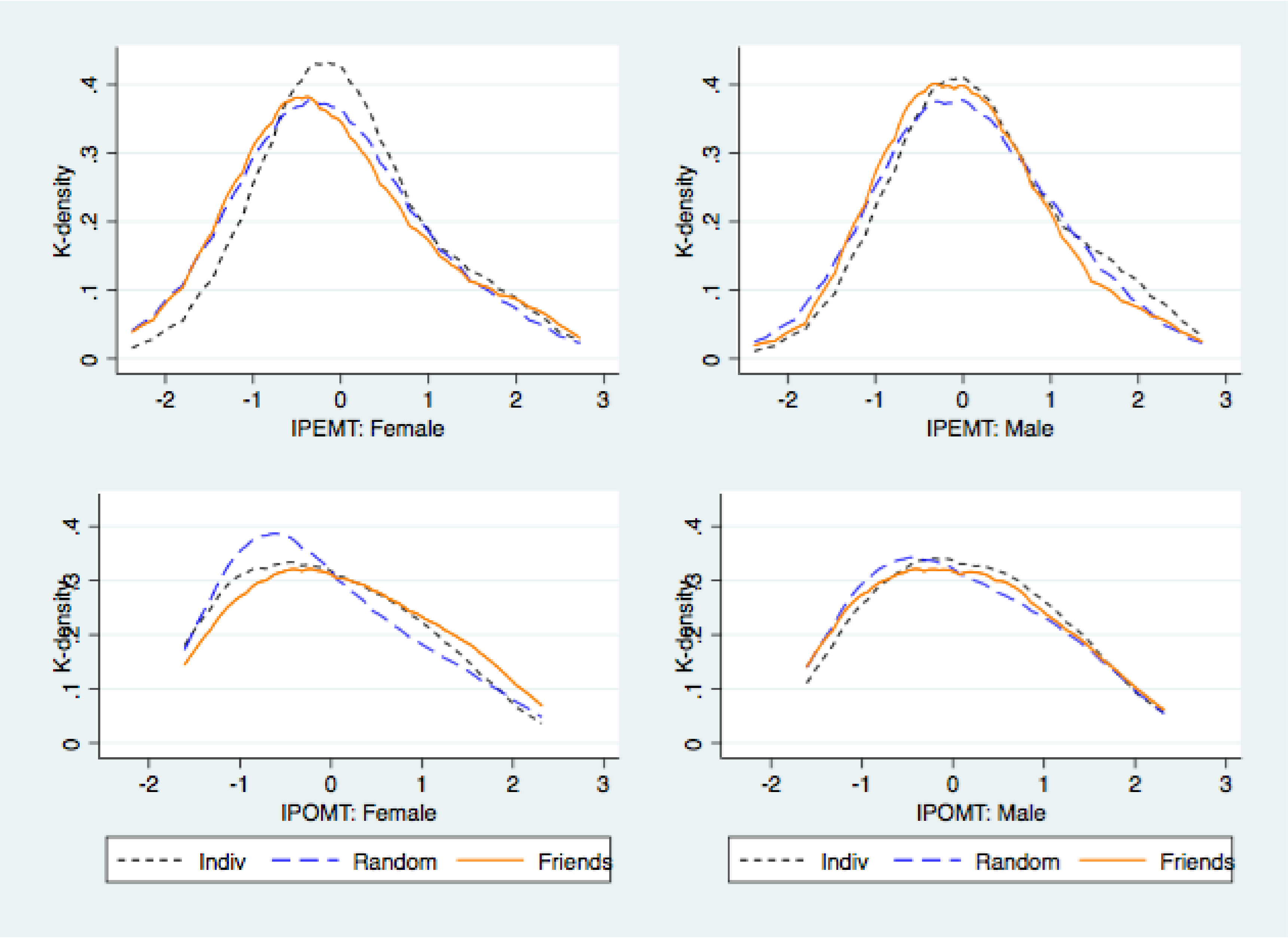


Figure 8: Non linear effects of groupings