Do Friendship Networks Improve Female Education?*

Youjin Hahn† Asadul Islam‡ Eleonora Patacchini§ Yves Zenou¶

August 23, 2017

Abstract

We randomly assign more than 6,000 students from 150 primary schools in Bangladesh to work on math assignments in one of three settings: individually, in groups with random schoolmates, 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 compositions to those of friendship groups do not produce similar effects. Our study thus documents that placing students into study groups with their friends may improve learning, especially for low-ability females.

JEL Classifications: I25, J16, O12.

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

*We thank Jim Berry, Jaesung Choi, Vesall Nourani, Sangyoon Park, the seminar participants at Monash University, Queensland University of Technology, University of Queensland, Yonsei University, the 24th Annual SJE International Symposium: Human Capital and Economic Development, Korean Labor Economic Association Meeting, and Korean International Economic Association Winter Meeting for valuable comments. We are indebted to Behrooz Hassani-Mahmooei for his help at the early stages of the project. 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 support in conducting this project. Angela Cools, Foez Mojumdar, Mujahid Islam, Mahbub Sarkar provided excellent research assistance.

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

‡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 achievement 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 the involvement of volunteers from non-government organizations (Banerjee et al. 2015).

This paper documents an alternative method that may aid learning. In particular, we show that combining friends into small study groups with common objectives significantly improves the individual performance of low-ability females. Such teaching practices require no guidance or monitoring from personnel outside the school.

We randomly assign more than 6,000 students from 150 primary schools in Bangladesh to work on mathematics assignments 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 cognitive 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. The objective of our analysis is to investigate whether the individual test scores improve after the experiment.

Our ex ante question was how to design interventions to help close the gender gap in education in Bangladesh. In the mid-1990s, the government introduced many education policies targeting female children, including compulsory free primary education and a stipend program in secondary schools in rural areas. These policies have led to gender parity in enrolment in both primary and lower secondary levels (Begum et al., 2017; Hahn et al., 2017). However, there has been little progress on learning outcomes, with boys still outperforming girls. Using data from the nationally representative 2005 Bangladesh Adolescent Survey, Amin and Chandrasekhar (2009) document that only 10 percent of girls who completed primary school passed the secondary school certificate (SSC) exam, compared to 25 percent of boys.[1] These findings are also corroborated by the fact that there is persistent gender imbalance in household educational expenditure favoring boys (Shonchoy and Rabbani, 2015).

Significant and persistent gender gaps in education are common across many developing countries. Improving the learning outcomes of female students, especially those with low abilities, is thus an important challenge not only for Bangladesh but also for much of the developing world. Actual progress in this respect, however, requires testing teaching practices that can be easily and inexpensively implemented in real settings. In our experiment, we focus on assessing the effectiveness of a simple teaching practice: the sorting of children into study groups of friends to work outside of class time. This practice is novel for primary school children in Bangladesh. In Bangladesh, as in many other developing countries, the teacher-pupil ratio in primary education is large (about 1 to 40 on average in 2015).[2] Occasionally, children can be grouped for extra-curricular activities during class time such as recitation, dictation, and singing (Rahman et al., 2004), but interpersonal interactions between peers after school are informal and typically not related to learning objectives.

When designing this 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 affects not only the outcome but may also alter friendship relationships, which may contaminate our results. It is indeed well-documented that networks *do* rewire in response to interventions (see, e.g. Comola and Prina, 2015, and Banerjee et al., 2016). To prevent restructuring of the network, we limit our period of study to one week.

The results of our experiment show that, regardless of their initial ability level, the gain (or loss) in math scores for male students is not affected by whether they study 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 study in groups with friends. We show that random groups with identical composition to that of friendship groups do not produce similar effects. This 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

[1] Data from a survey sponsored by the South Asian International Education Studies Network of Economic Research Institutes (SANEI) reveal an important gender test score gap in 2003, and note that fewer girls achieve top GPA in SSC exam (45% of girls as opposed to 55% for boys); see Huq and Rahman (2008).

[2] Source: Bangladesh Bureau of Educational Information and Statistics (BANBEIS), Education Database 2015.

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 (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). 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).[3]

Friendship effects, however, may arise from a variety of different mechanisms that are difficult to pinpoint. In our data, a student's reported friends are almost exclusively of the same sex. This implies that friendship study groups are overwhelmingly of the same sex, while randomly assigned study groups are, on average, composed of roughly half males and half females. In addition, the same-sex students in a random group are quite likely to be friends (because a student lists 10 friends and they are mostly of the same sex). Thus, students in the randomly assigned study-groups have smaller numbers of friends in their study groups (Figure 4) largely because these groups typically contain a lot more students of the opposite sex. The issue of whether differences in outcomes across the friendship study groups and the other study groups are due to differences in friendships per se or are due to differences in the number of same sex peers seems to be of first order importance, especially given the traditionally strong differences in the roles of males and females in Bangladesh. We control the fraction of female peers in our regression. Overall, our results led us believe that girls do well in a study group consisting of friends because low-ability girls do well in a study group consisting of other females since girls are more comfortable expressing themselves (and, therefore, learn more) when they are around high-ability girls.

Our analysis contributes to the economic development literature on the gender gap, aim-

[3]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). In particular, females, as opposed to males, appear to do worse when facing competitive incentive schemes (Gneezy and Rustichini, 2004) but are more attracted by cooperative incentive schemes (Kuhn and Villeval, 2015). Some have argued that differences in preferences and confidence in one's own relative abilities (for overviews, see Eckel and Grossman, 2008, and Croson and Gneezy, 2009) are key in explaining such gender-specific attitudes. This is in line with the finding of our analysis. This literature, however, does not consider frienship effects.

ing to evaluate 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 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).[4] 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 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) find that, when a student is given monetary incentives to exercise, this student exercises more if a higher 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

[4]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%.

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 worse when they work with their friends, suggesting that disruptions might be greater among friends.[5] 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. 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.[6] 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 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.

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 enrollment and a 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

[5] 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 that their occupation is a housewife.

[6] 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).

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., 2017).

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. There are more than 800 primary schools in these two districts. Figure 1 shows the location of the selected schools. In total, we interviewed 6,376 students.

[*Insert Figure* 1 *here*]

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

Figure 2 shows the timing of our experiment. There are two phases in the experiment. In the first stage, we elicit friendship and household information and conduct an individual cognitive ability test. In the second stage, we form study groups and distribute assignments. After the treatment, we again test students' achievement.

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 in which parents report their education, age, and occupation, and report other household characteristics. Each student's ability is measured using a math test (*individual pre-experiment math test*, IPEMT), which has been 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. Children have 20 minutes to complete the test. A detailed description of the IPEMT is contained in the online appendix.

[*Insert Figure* 2 *here*]

In July 2013 (referred to as period t), students are placed into settings for the treatment. We consider three different settings: (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* setting, 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. Friendship and random-peer groups are constructed to balance the ability of group members (that is, the mean and the distribution of student ability is comparable across groups). To achieve similarity across groups, we use the following methodology. 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.

A similar algorithm is used to construct friendship groups. The difference is that groups are formed using the friendship nominations and concept of *cliques* in network analysis.[7] First, the computer finds an initial clique of size four, keeps it, and then removes 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 the algorithm has finished, we perform the tests mentioned above for differences in terms of peers' ability, parental education, and household income across groups. As in

[7] 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).

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 have 4 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 the 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 this 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 education 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) has handed in its GMT, each student is asked to perform an individual post-experiment math test (IPOMT). As mentioned in the Introduction, we only allow for one week student interactions to prevent students in random peer groups from forming new friendship relationships. The IPOMT is based on the GMT. Although none of the test items is repeated from the GMT, the questions are similar so that it is possible for students to use what they have 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 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 most improvement from their group average baseline math test (that is, the most improvement between the IPEMT and IPOMT). 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 setting): 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 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 most improvement in test score. These prizes are set in consultation with teachers and students to ensure that 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 addition, all participant children receive gifts (e.g., a pencil/pen) and certificates for their participation.

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 3 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 (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 slightly above 50%. This seems to indicate that there is not a strong tendency toward homophily behaviors (McPherson et al., 2001).

[*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 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 4 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 individuals in random groups, few friendship links exist while, for those in the friendship groups, the opposite occurs.

[*Insert Figure* 4 *here*]

Table 2 shows the pre-experiment gender gap in test scores (IPEMT) across group types.[8] 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.

[*Insert Table* 2 *here*]

Table 3 presents summary statistics, distinguishing between the three types of schools (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.[9] The last columns of the table formally tests whether there are statistically significant differences between the schools placed in the three settings in terms of the observed characteristics.[10] It appears that all characteristics are well balanced, with the exception of the percentage of

[8] 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, with and without including a set of controls.

[9] 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.

[10] 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.

females which is slightly higher in the school assigned to the friendship treatment.[11] In our regression analysis, we will therefore control for female share in each group.

[*Insert Table* 3 *here*]

Figure 5 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 peer 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 the performance of boys is minimally affected by the group-type, the performance of female students is clearly affected by the treatment. Moreover, while male students perform better than female students before the experiment, females studying in friendship groups catch up in the post-experimental math test. Finally, this figure also shows that the pre-experiment performance of females assigned to friendship groups is roughly similar to that of females in the other groups. However, 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 Figure* 5 *here*]

In Figure 6, we plot the estimated post-experimental performance against initial levels of ability allowing for non-linear effects.[12] 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 more rigorous analysis.

[*Insert Figure* 6 *here*]

[11] 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. The results do not change qualitatively when we drop students with imputed test scores.

[12] 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.

4 Results

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

$$y_{irs}^{IPOMT} = \beta_0 + \beta_1 D_{\text{friend}} + \beta_2 D_{\text{random}} + \beta_3 y_{irs}^{IPEMT} + \beta_4 X_{irs} + \epsilon_{irs} \quad (1)$$

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 zero if he/she studies by him/herself or belongs to a random group, 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 or belongs to a friendship 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.) and ϵ_{irs} is an error term. Standard errors are clustered at the school level.

Table 4 reports the results of this regression for the entire sample in columns (1) to (3) with an increasing set of controls. The results suggest that, for the full sample, there is no effect of grouping on individual math test scores after the experiment. The other columns of Table 4 show the results when considering students with different levels of ability. Using the distribution of the IPEMT for the whole sample, we define low-ability students as those who are below the median value whereas high-ability students are those above the median value.[13] As in the previous case, there is no significant effect of grouping on test scores, even though the effect for low-ability students is positive and greater in magnitude than the effect for the high-ability students.

[*Insert Table* 4]

Let us now turn our attention to the ex ante question of the experiment, which is whether such practices are effective for females, especially low-ability females. Table 5 collects the results. They reveal that, for male students, there is no effect of grouping on the change in math test scores. In contrast, there is a large and positive gain in math scores for the low-ability female students who studied in groups. The effect is larger and significant if a low-ability female studies with a group of friends rather than with a random group of peers. Indeed, compared to studying alone, studying with a group of friends increases the test scores of low-ability female students by 0.45 of a standard deviation of the IPOMT (which is standardized using the mean and standard deviation of the entire sample of students),

[13] Due to discrete scoring of IPEMT, the percentage of students below and above the median is 45 and 55 percent, respectively.

while being in a random peer group increases math scores by only about 0.14 of a standard deviation (and the effect is not statistically significant). When we test whether the two effects (belonging to random groups or friendship groups) are equal for low-ability female students, we reject the null at the 10 percent level. The fact that we find positive gains for low-ability students only is consistent with the idea 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 (and not low-ability male students) obtain higher scores from the treatment.[14]

[*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 irs and individual jrs, respectively, θ_{irs} and θ_{jrs} are the *unobservable characteristics* of individual irs and individual jrs, respectively.

If there are some peer 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 $cor(D_{\text{friends}}, X_{jrs}) \neq 0$. Our data indicate that a student's reported friends are almost exclusively of the same sex (Figure 3). This implies that friendship study groups are overwhelmingly of the same sex, while randomly assigned study groups are, on average, composed of roughly half males and half females. Thus, students in the randomly assigned study-groups have smaller numbers of friends in their study groups (Figure

[14] We also perform our analysis using gains in test scores as the dependent variable, $\Delta y_{irs} = y_{irs}^{IPOMT} - y_{irs}^{IPEMT}$. The results remain qualitatively unchanged, although the magnitude of the effects is larger and the coefficient on random peer group becomes statistically significant for low-ability female at the 10 percent level. These results are available upon request.

4) largely because these groups typically contain a lot more students of the opposite sex. As a result, we believe that girls do well in a study group consisting of friends, who are mainly females, because girls are more comfortable expressing themselves (and, therefore, learn more) when they are around other girls. In other words, friendship groups make girls feel comfortable asking questions and expressing their opinions to other girls. This is 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.

Using the subsample of students belonging to friendship or random peer groups, we address this issue in Table 6 by controlling for the fraction of females in the peer group (excluding the student him/herself). Across all specifications, we find that male students are not affected by friendship groupings. We also find that low-ability female students working in groups of friends gain the most. In other words, studying with friends is always better than studying with random peers for low-ability female students, even after controlling for the fraction of females in the group. However, we also find that, when the fraction of female increases in a group, the performance of low-ability female students decrease. This does not contradict the fact that female students feel more comfortable interacting with other female students but indicate that low-ability female students improve their individual performance when they are able to interact with high-ability female students, which should be mechanically true since groups are balanced by ability.

[*Insert Table* 6 *here*]

To further understand these mechanisms, there are two other characteristics of the group environment that might affect students' performance. First, the level of ability of the peers that are randomly assigned to the study group, which could be a vital issue given that girls, on average, have lower IPEMT. Having more female students in a group may thus capture a lower ability environment as shown in Table 6. We investigate this issue by controlling for the *average* IPEMT of the peers. The results are reported in columns (1) and (3) of Table 7 and show that our evidence on the importance of friends for low-ability females remains true. Also, the direct effect of average IPEMT on individual performance of low-ability girls is significantly positive. Second, the *dispersion* of the ability level of the random peers may also be an important factor in explaining individual performance. Table 7 (columns (2) and (4)) show that the evidence remains qualitatively unchanged when controlling for the standard deviation of the IPEMT of the peers. To summarize, low-ability female perform better in friendship groups because they seem to be more comfortable expressing themselves and, therefore, learn more when they are around high-ability girls.

[*Insert Table* 7 *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 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. We match one characteristic at a time. 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 run the following regression:

$$y_{irs}^{IPOMT} = \gamma_0 + \gamma_1 D_{\text{Placebo Friends}} + \gamma_2 y_{irs}^{IPEMT} + \gamma_3 X_{irs} + \epsilon_{irs} \quad (2)$$

where $D_{\text{Placebo Friends}}$ is a dummy variable that is equal to 1 if an individual student i belongs to the random peer groups in which empirical distributions resemble those of friendship groups, and zero if he/she belongs to the original sample 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 8. 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* 8 *here*]

6 Inspecting the mechanisms

As shown in Table 6, we have 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 following our discussion in the previous section.

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. Since friendship study groups are overwhelmingly of the same sex while randomly assigned study groups are not, our results indicate that low-ability female are more comfortable expressing themselves and, therefore, learn more when they are around high-ability girls.

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 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 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).

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 expect to find no differences in the distributions of the GGKT between random and friendship groups when students had no time to interact and a difference in the distributions of the scores after a week of interactions for the GMT. Figure 7 displays the kernel density plots for the GGKT and the GMT distinguishing between random and friendship groups. The graphs show that while the two curves are almost overlapping for the GGKT, the distribution of the GMT for friendship groups is shifted to the right. We formally test these differences using a Kolmogorov–Smirnov test. The test cannot reject the null hypothesis that the GGKT has the same distribution between the random and the friendship groups (p-value equals 0.375), while it detects a statistically significant difference in distribution between these two types of groups for the GMT (p-value is smaller than 0.001). This evidence suggests that greater learning is taking place within a group of friends than within a group of random peers.

[*Insert Figure* 7 *here*]

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. 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 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 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 (1). The results are displayed in Table 9. 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 counterparts in random groups.

[*Insert Table* 9 *here*]

Our results suggest that low-ability females achieve increased learning in friendship groups, and this is not due to the time spent on work. Social psychology literature suggests some additional reasons why this may be the case. The literature notes that motivation gains are highest in cohesive groups (Karau and Hart, 1998), and the number of friends and links in a group can also be considered a measure of group cohesion (see e.g. Jackson, 2008 or Jackson et al., 2017). In Table 10, 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 alternative explanatory variables for test score 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 low-ability female students are. Such additional evidence is in line with the postulated mechanism of learning.

[*Insert Table* 10 *here*]

7 Concluding remarks

Fighting low levels of basic education in developing countries is a priority for economic development. The experiment reported here provides evidence on the effectiveness of placing students in small study groups in the context of Bangladeshi primary schools. In our experiment, 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 a positive impact of studying with friends on low-ability females, a group that typically performs well below grade level. Our field experiment shows the possibility for effective improvements in learning through inexpensive teaching practices. As such, it is potentially of great importance for educational policies in developing countries.

References

[1] Amin, S. and S. Chandrasekhar (2009), "Looking beyond universal primary education: Gender differences in time use among children in rural Bangladesh," Working Paper No. 17, Population Council.

[2] BANBEIS. (2015). "Educational database 2015." Bangladesh Bureau of Educational Information and Statistics (BANBEIS), Ministry of Education, Government of the People's Republic of Bangladesh, Dhaka.

[3] 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.

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

[5] 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.

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

[7] 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.

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

[9] 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.

[10] Banerjee, A., Chandrasekhar, A., Duflo, E. and M.O. Jackson (2016), "Gossip: identifying central Individuals in a social network," Unpublished manuscript, Stanford University.

[11] Begum, L., A. Islam and R. Smyth (2017), "Girl power: Stipend programs and their effects on the education of younger siblings," *Journal of Development Studies*, forthcoming.

[12] Bharadwaj, P., De Giorgi, G., Hansen, D. and C. Neilson (2016), "The gender gap in mathematics: Evidence from Chile," *Economic Development and Cultural Change* 65, 141-166.

[13] 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.

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

[15] Comola, M. and S. Prina (2015), "Treatment effect accounting for network changes," Unpublished manuscript, Paris School of Economics.

[16] Croson, R. and U. Gneezy (2009), "Gender differences in preferences," *Journal of Economic Literature* 47, 448-474.

[17] 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.

[18] 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.

[19] Eckel, C.C. and P.J. Grossman (2008), "Differences in the economic decisions of men and women: Experimental evidence," In: C.R. Plott and V.L. Smith (Eds.), *Handbook of Experimental Economics*, *Vol. 1*, Amsterdam: Elsevier, pp. 509-519.

[20] 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.

[21] 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.

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

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

[24] 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.

[25] 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.

[26] Gneezy, U. and A. Rustichini (2004), "Gender and competition at a young age," *American Economic Review* 94, 377-381.

[27] Hahn, Y., Islam, A., Nuzhat, K., Smyth, R. and H.-S. Yang (2017), "Education, marriage and fertility: Long-term evidence from a female stipend program in Bangladesh," *Economic Development and Cultural Change*, forthcoming.

[28] 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.

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

[30] Huq, M. and Rahman, P.K. (2008), "Gender disparities in secondary education in Bangladesh," *International Education Studies* 2, 115-128.

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

[32] Jackson, M.O., Rogers, B. and Y. Zenou (2017), "The economic consequences of social network structure," *Journal of Economic Literature* 55, 49-95.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] Kuhn, P. and M.C. Villeval (2015), "Are women more attracted to cooperation than men?" *Economic Journal* 125, 115-140.

[38] 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.

[39] 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.

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

[41] 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.

[42] 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.

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

[44] 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.

[45] 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.

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

[47] Rahman, A., Spaulding, S. and K. Tietjen (2004). "Bangladesh Educational Assessment, Time to Learn: Teachers' and Students' Use of Time in Government Primary Schools in Bangladesh." Creative Associates International, Inc. Washington, DC Prepared for Basic Education and Policy Support Activity, USAID.

[48] 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.

[49] Shonchoy, A.S. and M. Rabbani (2015), "The Bangladesh gender gap in education: Biased intra-household educational expenditures," *mimeo.*

[50] 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.

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

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

Figure 1: Location of the different schools

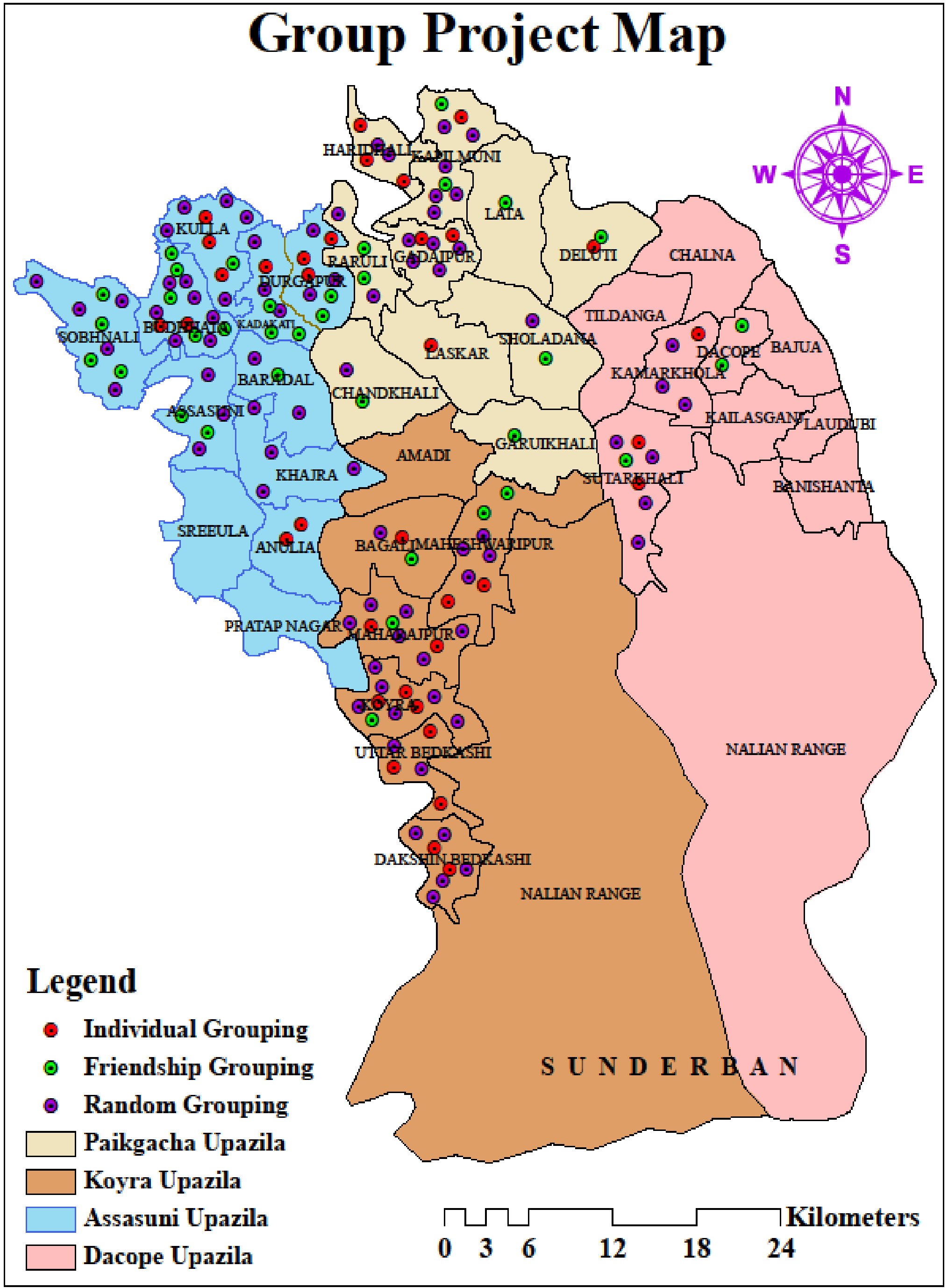


Figure 2: Timeline of the experiment

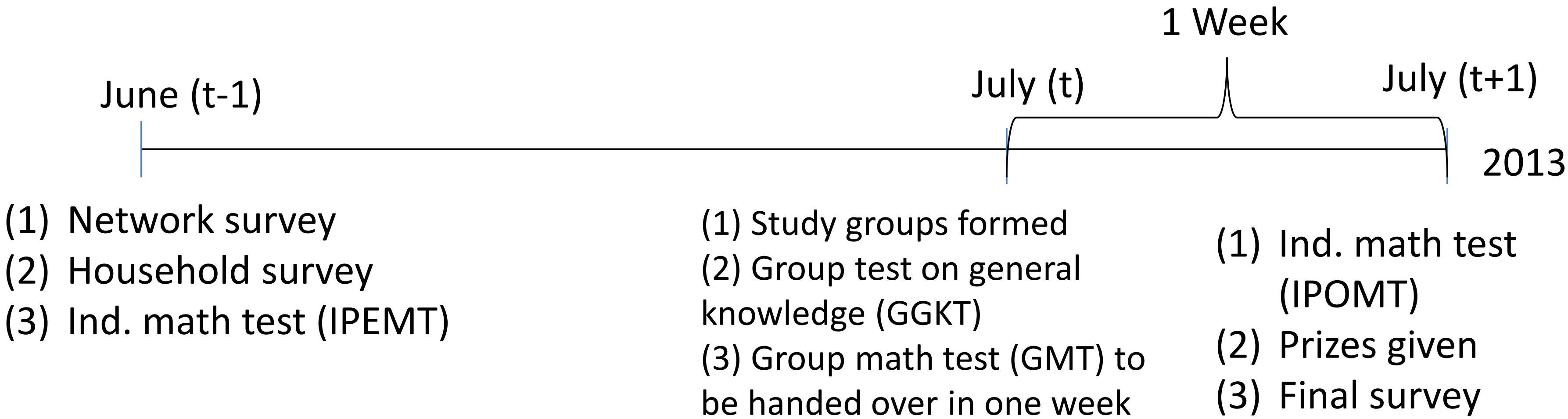


Figure 3: Distribution of students by same-gender friendship nomination

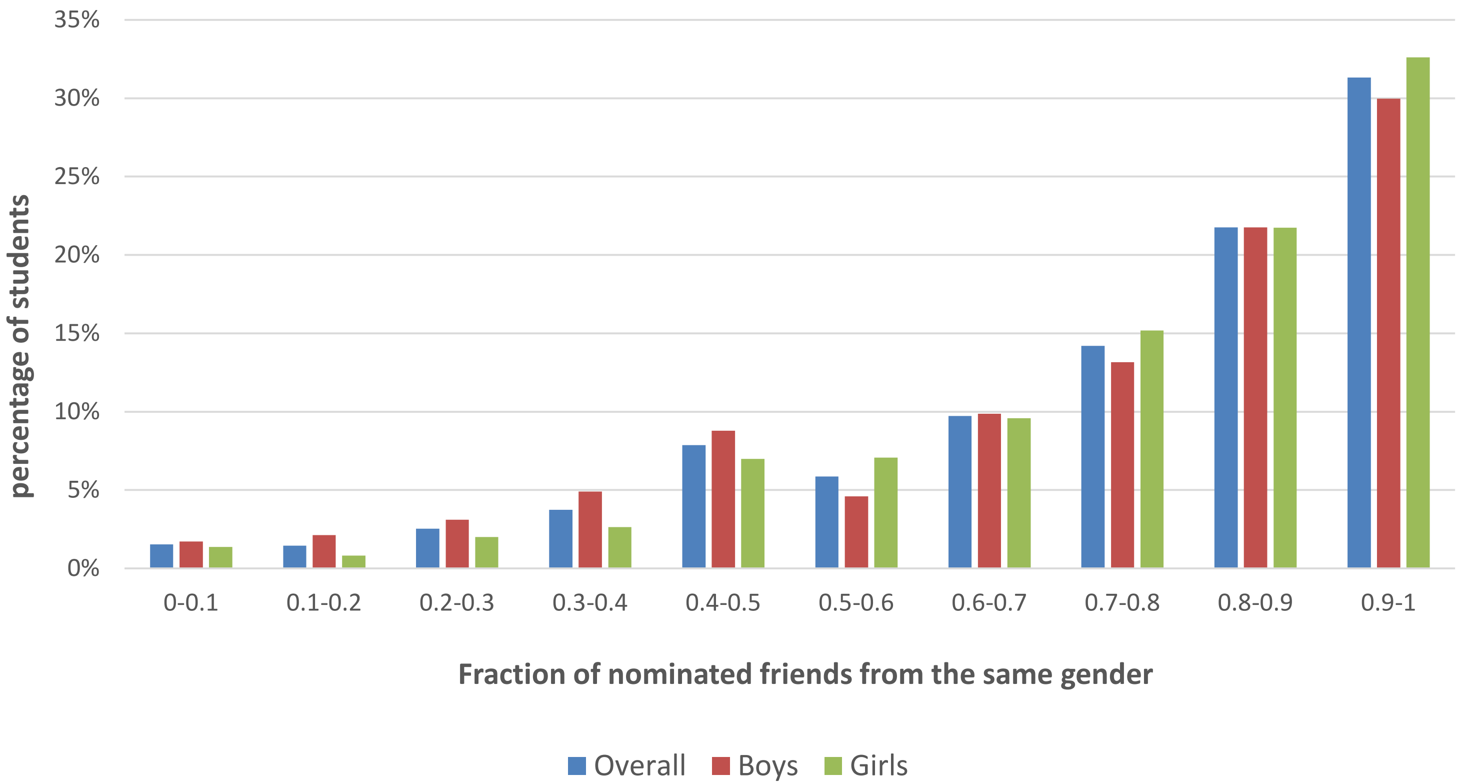


Figure 4: Distribution of students by friendship relationships in a study-group

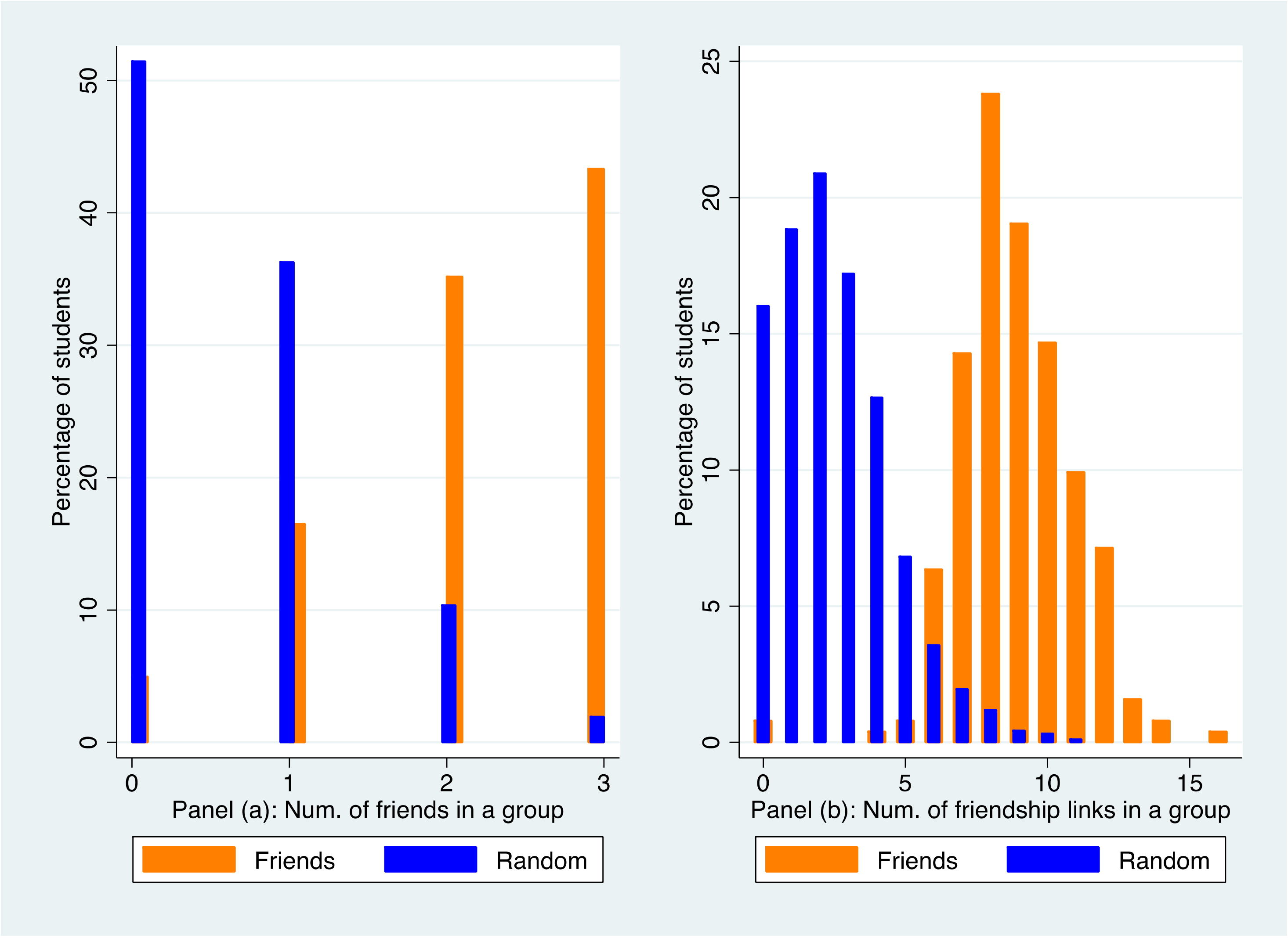
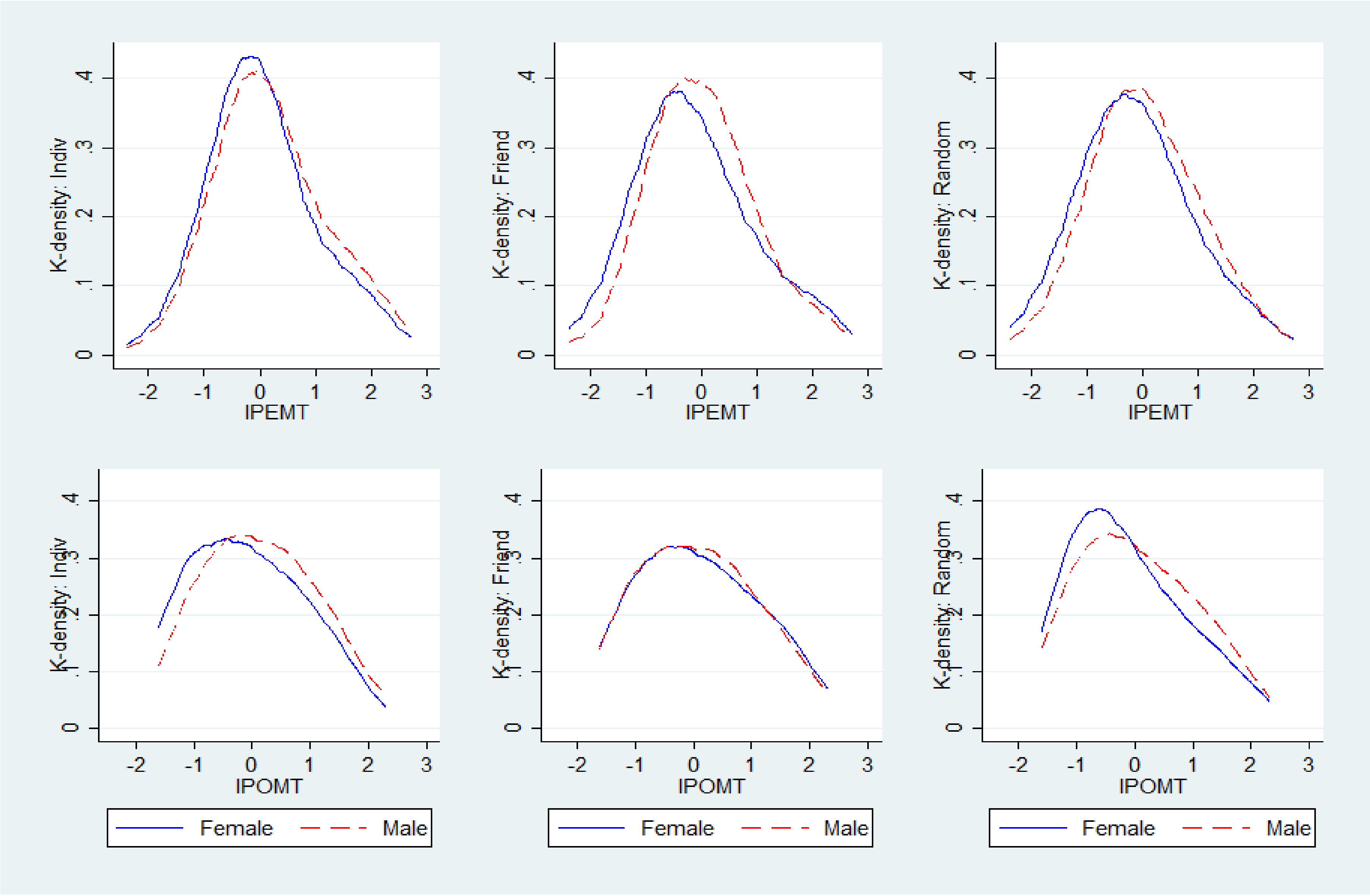


Figure 5: Gender gap before and after the experiment by group type



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

Figure 6: Non linear effects of groupings

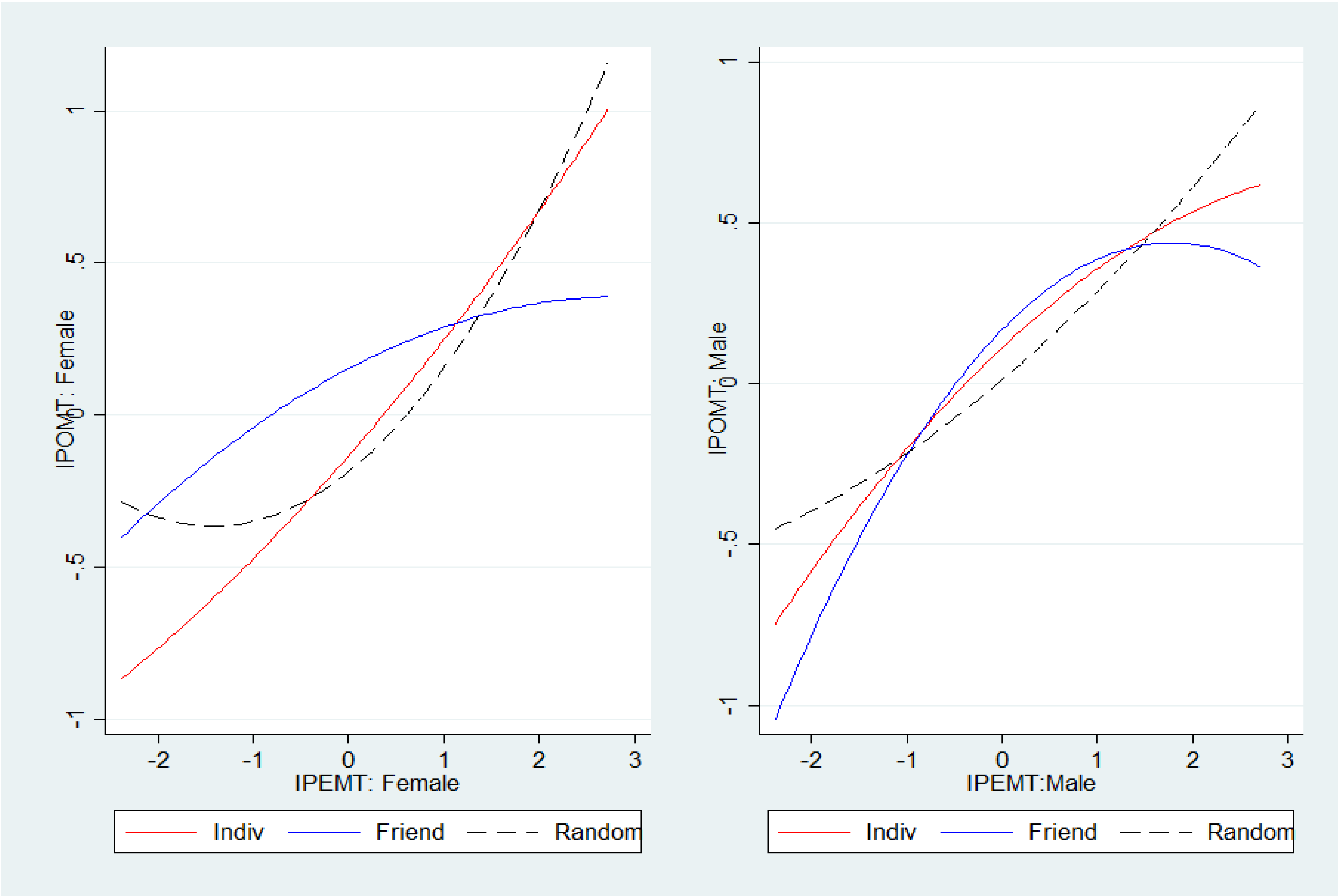


Figure 7: Effect on Group general knowledge test and group math test

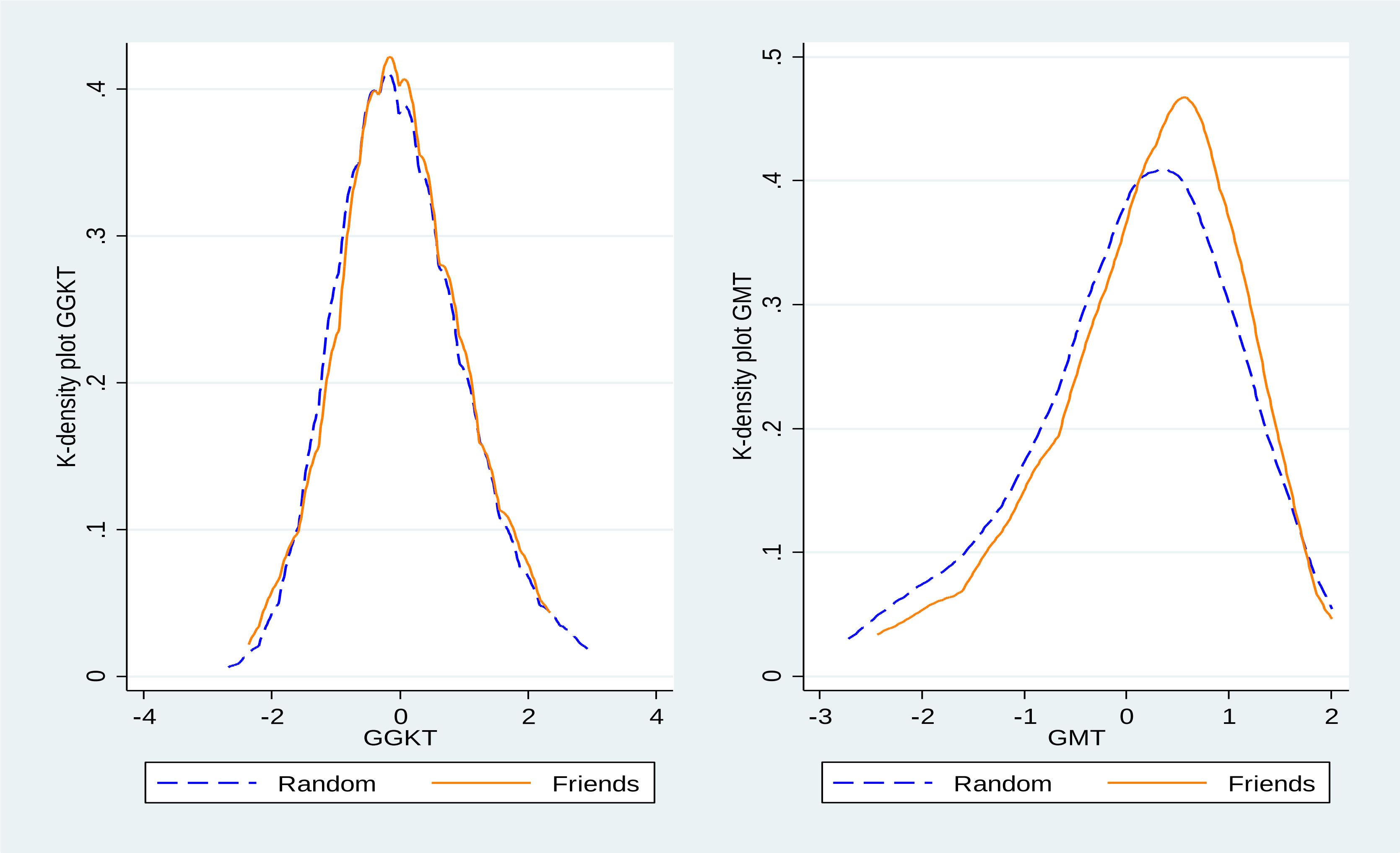


Table 1: Friendship nomination by ability, parental education, and household income

	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: Friends (friendship group)								
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: Random (random-peer group)								
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)	(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

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

	Random	Friendship	Individual	*p*-value of the difference: Friendship vs. random	*p*-value of the difference: Individual vs. friendship	*p*-value of the difference: 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)			
Observations	3,671	1,005	1,660			

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

Table 4: Do students perform better when studying with peers than studying alone?
Dependent variable is individual post-experiment math test (IPOMT)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Entire sample			**Low**			**High**		
Friends	0.076	0.126	0.112	0.228	0.248	0.249	0.026	0.026	0.006
	(0.642)	(0.424)	(0.466)	(0.239)	(0.194)	(0.183)	(0.879)	(0.880)	(0.971)
Random	-0.060	-0.013	-0.024	0.041	0.062	0.076	-0.075	-0.070	-0.094
	(0.661)	(0.918)	(0.849)	(0.776)	(0.666)	(0.584)	(0.598)	(0.619)	(0.493)
IPEMT		0.267***	0.239***		0.164**	0.182***		0.259***	0.213***
		(0.000)	(0.000)		(0.013)	(0.006)		(0.000)	(0.000)
Observations	6336	6336	6336	2894	2894	2894	3442	3442	3442
Controls	No	No	Yes	No	No	Yes	No	No	Yes

Note: The control variables 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: Do students perform better when studying with peers than studying alone?
Dependent variable is individual post-experiment math test (IPOMT)

	(1)	(2)	(3)	(4)	(5)	(6)
		Low			**High**	
			Panel A: Females			
Friends	0.430*	0.453**	0.451**	0.037	0.010	-0.019
	(0.223)	(0.218)	(0.217)	(0.195)	(0.196)	(0.192)
Random	0.100	0.123	0.136	-0.078	-0.081	-0.099
	(0.165)	(0.162)	(0.158)	(0.148)	(0.143)	(0.141)
IPEMT		0.139**	0.159**		0.349***	0.312***
		(0.070)	(0.071)		(0.060)	(0.060)
Observations	1584	1584	1584	1651	1651	1651
			Panel B: Males			
Friends	-0.035	-0.024	-0.023	0.023	0.037	0.016
	(0.174)	(0.173)	(0.166)	(0.169)	(0.175)	(0.164)
Random	-0.032	-0.017	0.009	-0.074	-0.066	-0.090
	(0.144)	(0.142)	(0.135)	(0.149)	(0.150)	(0.146)
IPEMT		0.192**	0.224***		0.168**	0.120*
		(0.082)	(0.079)		(0.066)	(0.064)
Observations	1310	1310	1310	1791	1791	1791
Controls	No	No	Yes	No	No	Yes

Note: The control variables 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 6: Robustness checks.
Dependent variable is individual post-experiment math test (IPOMT)

	(1)	(2)	(3)	(4)	(5)	(6)
		Low			**High**	
Panel A: Females						
Friends	0.402**	0.405**	0.396**	0.104	0.070	0.056
	(0.198)	(0.196)	(0.197)	(0.187)	(0.185)	(0.177)
Fraction female peers	-0.212*	-0.221*	-0.228*	0.035	0.067	0.066
	(0.124)	(0.125)	(0.123)	(0.116)	(0.114)	(0.107)
IPEMT		0.115	0.132		0.353***	0.311***
		(0.077)	(0.080)		(0.075)	(0.072)
Observations	1219	1219	1219	1175	1175	1175
Panel B: Males						
Friends	-0.038	-0.042	-0.062	0.097	0.107	0.118
	(0.166)	(0.167)	(0.159)	(0.160)	(0.164)	(0.155)
Fraction female peers	-0.116	-0.119	-0.101	0.001	0.010	0.034
	(0.141)	(0.143)	(0.137)	(0.120)	(0.120)	(0.116)
IPEMT		0.174*	0.197**		0.187***	0.150**
		(0.094)	(0.089)		(0.070)	(0.066)
Observations	1008	1008	1008	1274	1274	1274
Controls	No	No	Yes	No	No	Yes

Note: The control variables 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 7: Robustness checks (cont.)
Dependent variable is individual post-experiment math test (IPOMT)

	(1)	(2)	(3)	(4)
	Low		**High**	
	Panel A: Females			
Friends	0.474**	0.465**	0.072	0.138
	(0.192)	(0.194)	(0.176)	(0.172)
Average IPEMT of peers	0.262***		-0.074	
	(0.098)		(0.080)	
Fraction female peers	-0.199	-0.216*	0.054	0.067
	(0.124)	(0.122)	(0.107)	(0.106)
Std. dev. IPEMT of peers		0.391*		0.373**
		(0.223)		(0.170)
Observations	1219	1219	1175	1175
	Panel B: Males			
Friends	-0.022	0.020	0.131	0.148
	(0.164)	(0.172)	(0.158)	(0.163)
Average IPEMT of peers	0.157*		-0.052	
	(0.082)		(0.076)	
Fraction female peers	-0.076	-0.104	0.027	0.034
	(0.138)	(0.133)	(0.112)	(0.115)
Std. dev. IPEMT of peers		0.365*		0.107
		(0.204)		(0.177)
Observations	1008	1008	1274	1274

Note: The control variables 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 8: Placebo tests
Dependent variable is individual post-experiment math test (IPOMT)

	(1)	(2)	(3)	(4)
	Low	**High**	**Low**	**High**
	Female		**Male**	
Panel A: Random groups with the same group composition on fraction of female students as friendship groups				
Placebo Friends	0.022	0.084	-0.062	-0.112
	(0.066)	(0.065)	(0.068)	(0.076)
Panel B: Random groups with the same group average IPEMT as friendship groups				
Placebo Friends	0.013	-0.009	0.029	-0.025
	(0.021)	(0.030)	(0.030)	(0.025)
Panel C: Random groups with the same group composition on parental education as friendship groups				
Placebo Friends	0.033	0.027	-0.032	0.018
	(0.025)	(0.029)	(0.027)	(0.028)
Panel D: Random groups with the same group composition on household income as friendship groups				
Placebo Friends	0.015	0.045	-0.014	0.010
	(0.027)	(0.028)	(0.027)	(0.029)

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. We control for individual characteristics as well as fraction of female peers in the group. The other control variables 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 9: Potential channels of influence in friendship grouping

	(1)	(2)	(3)	(4)	(5)	(6)
		Low			**High**	
	Num Met	Team Hrs	HW Hrs	Num Met	Team Hrs	HW Hrs
			Panel A: Females			
Friends	-0.002	0.308	0.045	0.142	0.448*	0.007
	(0.247)	(0.229)	(0.271)	(0.224)	(0.227)	(0.312)
Observations	1219	1219	1217	1175	1173	1175
			Panel B: Males			
Friends	0.176	0.163	0.171	0.211	0.222	0.394
	(0.223)	(0.236)	(0.297)	(0.187)	(0.240)	(0.262)
Observations	1007	1006	1007	1274	1273	1274

Note: The dependent variable for col. (1) and (4) indicates number of times met as a team (Num met); (2) and (5) indicates how many hours the group met as a team (Team Hrs); (3) and (6) how many hours a student spent in total doing the Group math test (HW Hrs); We control for individual characteristics as well as fraction of female peers in the group. The control variables 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 10: Alternative definitions of friendship relationships
Dependent variable is individual post-experiment math test (IPOMT)

	(1)	(2)	(3)	(4)
	Low		**High**	
	Panel A: Females			
Num. of friends in a group	0.134**		-0.029	
	(0.055)		(0.041)	
Number of group links		0.053***		0.003
		(0.019)		(0.019)
Observations	1219	1219	1175	1175
	Panel B: Males			
Num. of friends in a group	-0.004		0.024	
	(0.050)		(0.049)	
Number of group links		0.000		0.015
		(0.018)		(0.017)
Observations	1008	1008	1274	1274

Note: We control for individual characteristics as well as fraction of female peers in the group. The other control variables 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.