Gender Peer Effects Around the World

José María Cabrera*

July 28th, 2015

Abstract
What is the impact on human capital formation of gender peer effects? Do boys (and girls) increase their test scores when they are surrounded by more female peers? To motivate the analysis, I show that in a field experiment in Kenyan primary schools, where students were randomly allocated to classrooms, the experimentally induced variation in the proportion of female classmates increases the performance of boys in the long run. Given the limited external validity of this experimental setting, I then provide evidence that this result also holds for 15 year old students in many countries. In several education systems there seems to be no selection into schools according to the gender of the student. I identify empirically those countries where the allocation of students into schools looks random. I provide evidence that a greater proportion of females increases the performance of both girls and boys. This result also holds for schools that don't face competition in their area (so there should be less selection). A possible channel is a better relation with teachers.

* jmcabrera@um.edu.uy

I thank Samuel Berlinski, Matías Busso, Alejandro Cid, Juan Dubra, Pascaline Dupas, André Portela Souza and Dinand Webbink for their comments at different stages of the project.

1 Introduction

The composition of a group of peers (colleagues, classmates) has an impact on one's performance. If a person is surrounded by *high achieving* peers, he may benefit from their positive influence. Does he also benefit from the *gender composition* of his group of peers? I will study the influence that a group with a greater proportion of females has on human capital formation. Do boys (and girls) increase their test scores when they have more female peers?

Suppose that a student is placed in a group with a higher proportion of females. This may have a positive or negative impact on his academic performance. On the one hand, if girls score lower than boys (e.g. in math), a group with a greater proportion of females will be negative for him: his group of peers will be under-performing relative to a predominantly male group. But on the other hand, if girls behave better in the classroom, a teacher can give the lesson with less noise and disorder, which is positive for learning. These two opposite effects are present in my data. I find that being placed in a group with more female peers is positive for achievement.

Previous researchers have analyzed this question for different countries and age groups. In an influential article, Hoxby (2000) found that both boys and girls perform better in classrooms that have more percentage of females. She uses data for elementary public schools in Texas. Lavy and Schlosser (2011) worked with data for elementary, middle and high school students in Israel. They also find that an increase in the proportion of girls improves boys' and girls' cognitive outcomes.

The identification strategy of both papers rests on variations in gender composition across adjacent cohorts within the same grade and school. Since there can be unobservable shocks, not related to demographic variations, that may cause non-random changes in the proportion of females that attend a given school between two consecutive years, a better strategy can be to rely on truly random assignments of students to classrooms. But this research design is extremely scarce. A notable exception is Whitmore (2005), who uses experimental variation across classrooms from Tennessee's Project STAR experiment, in kindergarten and the first years of primary school.

In the first piece of empirical evidence that I present, I also use a research design which randomly assigned primary students into classrooms. The experimental data for the first part of this paper was gathered by Duflo, Dupas and Kremer (2011). They had done a field experiment with first grade primary school students in Kenya. Authors looked at the influence of high achieving peers, and I will look at the influence of being randomly placed in a group with a higher proportion of female classmates. I find that having more female peers has a positive impact for boys, driven mainly by math scores, and only in the long run.

After presenting this evidence of peer effects in a well-designed randomized control trial with high internal validity, my next step will be to try to answer the question for a much broader context (many countries), and for a different age group (15 year old students), using PISA 2009 data. There is a tradeoff between internal and external validity. In a small setting, the selection process of students into classes can be better controlled, but the answers we get may be extremely local[1]. Are the results obtained in Israel useful for students in Peru, for example? Are the results from primary school

[1] There are other well designed experiments that also have local validity: university students in Essex and Amsterdam. Booth, Cardona-Sosa and Nolen (2013) randomly assign students to all-female, all-male, and coed classes at University of Essex. Treatment was one hour a week lecture with a teacher assistant. They look mainly at the effect of single-sex education, and find that it benefits women. Oosterbeek and van Ewijk (2014) manipulate the shares of females in 14 workgroups of 39 students each for first year students at University of Amsterdam. They find that males tend to postpone, but not abandon their dropout decision when surrounded by more females and perform worse on courses with high math content. Overall, they don't find substantial gender peer effects on achievement.

students in Texas in the 90´usefull for policy design in Germany in 2015? In the case of this paper: Do 9 year old Kenyan students offer advice on gender peer effects for other developing or developed countries?[2] So in the second part of this paper I will use data from many countries around the world. This exercise will have a higher external validity that the small experiment, but the identification strategy will be more challenging.

The research design is based on the fact that in some countries there is no selection into schools based on the *gender* of the student. Parents may send their children (boys and girls) to the same school. And school principals are not allowed to discriminate an incoming student because of her sex. Obviously, there can be selection into schools by parent's wealth, for example; but maybe boys and girls with the same income go to the same school. In those countries where the allocation of students into schools doesn't depend on their sex, two schools may have a different proportion of female students just by random chance. Take for example 30.000 students (half of them are girls) and 1000 groups (or schools) of 30 students each. Also suppose that parents and schools don't discriminate according to gender (a father will send his son or daughter to the same school). In this case, where there is no selection of schools based on the sex of the children, a school may end with 60% of girls, while another similar school will randomly have 40% of girls and 60% of boys. This "statistical variation" will be the one used for identification (Figure 1). This means that even if two schools are identical, a student in one or the other will face a very different gender composition of peers in the classroom, and this may have an impact on learning. Students may benefit for randomly being in a school with a greater proportion of female classmates.

With this idea in mind, I will argue that in some countries the allocation of 15 year old students into schools seems random: it resembles Figure 1. The procedure to select countries that display a random allocation -*by gender*- of students into schools is explained in detail in section 3.1. From the original 74 countries in PISA sample, I keep 46 of them.

A second strategy will be to use only public schools which face no competition in their catchment area, so there is plausibly less selection from parents and schools, and variation in the proportion of females in the schools should be associated with variations in the demographic composition at the area level.

Using this large database, I find evidence that gender peer effects are present around the world, in different cultures and educational systems. More girls have a positive influence on boys academic achievement, and also on girls themselves. A possible channel is that when there is a greater proportion of 15 year old girls in a school, students report a better climate and relations with the teachers (e.g. less disruptions during lessons).

The paper makes two main contributions to the literature. The first one is to estimate gender peer effects with a RCT in a developing country, extending the work of Whitmore (2005) for USA. The second contribution is to estimate gender peer effects in a much broader context. I propose a methodology to select education systems where the allocation of students by gender into schools seems random. Using these countries I can provide an answer with higher external validity (rather than small settings like one single university, some schools that choose to participate in an experiment, or schools from a city or a state).

[2] Pritchett and Sandefur (2014) claim that context really matters for the estimated effects, and that in some cases the "true" causal effect relevant for policy may be closer to the one obtained by a simple OLS regression in the country of interest, rather than from a randomized control trial in another location.

The rest of the paper is organized as follows. In section 2 I present the Kenyan experiment, and in section 3 the analysis for many countries (I describe PISA data, the empirical strategy and present the main results). In section 3.3 I provide additional results: non-linear effects, channels and heterogeneous effects. Finally, I conclude in section 4.

2 Kenyan experiment

2.1 Data and empirical strategy

In a field experiment in Kenya, Duflo, Dupas and Kremer (2011) study the effect of tracking (placing high achieving students in the same group). They offer an extra teacher to 121 schools so they could split a first-grade class into two sections. In 60 of them, they placed students into the two groups based on initial achievement (these schools won't be part of my analysis). And in the remaining 61 (not-tracking) schools, a total of 3,409 students were randomly assigned to classrooms. This procedure generated variation in the *proportion of females* in the classrooms: some students were placed in a "male" group with only a 31% of female students, while others were placed in a group that, by random chance, had 76% female classmates (Figure 2). I will look if this randomly generated variation in the gender composition of peers has an effect on a student's academic achievement[3].

The program lasted for 18 months, starting in the second term of the 2005 academic year (May) and continued throughout 2006. Short term results are from a test implemented at the end of the program. Longer term results are from a test in November 2007, one year after the program had finished. Since there were a few reassignments of students between groups after the randomization was performed, the treatment indicator will be the one assigned in the lottery.

In Table 1 we can see (descriptive statistics to be provided).

Given the experimental design that randomly assigned children to groups, I can consistently estimate by OLS the impact of the percentage of female peers with the following regression:

$$y_{igs} = \alpha + \beta females_{-igs} + \boldsymbol{X}\gamma + \mu_s + \varepsilon_{igs} \quad [\mathbf{1}]$$

where y_{igs} is the test score of student i in group g in school s (for math, literature and the total score). $Females_{-igs}$ is the proportion of females in the group of student i, excluding himself from the average (leave one out procedure). The vector of control variables $\boldsymbol{X}$ includes student gender[4]. It also includes average baseline grades of peers in order to isolate the effect of peers academic quality. If females perform better on average than boys, the impact of peers *gender* will be confounded with the impact of peers academic *quality*[5]. By including school fixed effects (v_s) the variation used in the estimation is

[3] This is a very nice design to study gender peer effects. The authors look at the effect of high achieving peers, but leave the question of gender peer effects unanswered. There are very few studies that have generated random allocation of students into classes (an exception is Whitmore, 2005, who uses the variation in the proportion of female peers generated by project STAR in Tennessee, US).
Data is available from: http://hdl.handle.net/1902.1/16787 UNF:5:LIcjw/R9R2mWHATBJhAiow== Jameel Poverty Action Lab [Distributor] V2 [Version]

[4] In principle, there should be no need to include controls in a randomized control experiment. To see why the omission of own-baseline controls may lead to biased estimations in this setting, see Angrist (2014). The reason is that there exists a mechanical relation between individual data and leave-one-out means.

[5] Other controls are: age of the student, being assigned to the new teacher and dummy variables for imputed baseline scores.

the difference in the female proportion between the two groups of the same school[6]. The estimation of the parameter β will be the causal effect of peer gender on academic performance (both directly through peer-to-peer interactions and through adjustment in teacher behavior, as pointed by Duflo, Dupas and Kremer, 2011). Standard errors are clustered at the classroom since the treatment is delivered at that level, and the variation used for the estimation is within school.

2.2 Results

In Table 2 we can see the causal effect of having a greater proportion of female peers. There are positive effects in the long run and they are concentrated in the math score. If the effects were linear, an increase from 0 to 1 in the proportion of females in the classroom would increase a student's test score by 0.79 of a standard deviation in the long run (column 3, first panel). In column 4 we see that results are driven by the gender of a student's classmates *(%females)*, and not by their quality (the regression controls for baseline test scores). The coefficient of *girls* shows that, regardless of the percentage of females in the classroom, girls score higher than boys, and the difference comes from their results in literature. Finally, boys are the ones that benefit the most from being in a classroom with more female peers[7].

3 Students from countries around the world

3.1 Data and empirical strategy

For this exercise, I will use PISA database for 2009. Raw data consists of information for 515.958 students, in 18.575 schools from 74 countries.

The identification strategy is based on the fact that in some countries there is no selection into schools based on the gender of the student. Parents may send their children (boys and girls) to the same school. And school principals are not allowed to discriminate an incoming student because of her sex. In other countries there may be a stronger selection process. Some parent may prefer to give a better education to their sons and send them to a different school than their daughters (or the other way round). Or there can be schools that are more "female oriented" because of the type of curriculum they offer. I will look at the distribution of students by gender into schools to separate those countries where the allocation of students into schools is random, from those where there is some kind of selection process.

In those countries where the allocation of students into schools doesn't depend on their sex, two schools will have a different proportion of female students by random chance. This "statistical variation" will be the one used for identification. The process that I have followed to construct my sample is explained bellow.

a. There is a variation in the proportion of females in schools when we look at the raw PISA data (Figure 3).
b. As a first step, I drop single sex schools (that account for a 10% of the observations)[8]. Parents that send their kids to those schools may be different in some unobservables from

[6] The proportion of females can be different between schools (a school may be more "female oriented"). The use of school fixed effects control for this difference, and also for other (fixed) differences across schools (like principals, school environment, etc). The proportion of females was randomly assigned inside each school.

[7] Results not shown, available upon request.

[8] I keep schools where the proportion of females belongs to [10%; 90%] to avoid coding errors (e.g. a school only for boys where one student was coded by mistake as a girl in the data, and so the school doesn`t have exactly 0% male students).

those who choose a coeducational school. So, I won't answer the question of whether single sex schools are better or worse for academic achievement than coeducational education (for a nice comparison of the effect of single sex schools, refer to the paper of Lee et al, 2014). Figure 4 shows the variation in the proportion of female peers in all the schools in the sample.

c. The variation shown in Figure 4 includes countries with no sex discrimination of students (explicit or implicit), and others where there is a selection process. To separate one from the other I will look if the allocation of male and female students in each individual country is random (for illustration purposes, I will show an example with the first country in the list: Albany).

At a first glance, the distribution of students in Albany may be random (Figure 5). It is centered in 50%, and seems well behaved. So the idea is to check if the empirical distribution that is observed in the data may be generated by a random allocation of students into schools. The procedure begins by *simulating* a random allocation of boys and girls *for each school*. I calculate the proportion of females in a country, and then drop the "sex" identifier for each student in the database. Then assign a random gender to each student, respecting the proportion at the country level. So each school will have a (fake) proportion of female students generated by random chance[9]. Then, I look if the empirical density of the proportion of females from PISA data is "similar" to the random one (Figure 6). To test if the two densities are equal (if the two samples have been drawn from the same population), I use two tests: Kolmogrov-Smirnov and runtest[10].

d. The simulated distribution of students may differ from the empirical one (PISA) by random chance. So the next step is to repeat the procedure 2000 times: generate random allocation of students and compare them with the original density. We can look (and test) if the empirical density belongs to the "family" of random allocations. In Figure 7 I've plotted the first 100 random allocations for Albany. The density from PISA seems to be one of the randomly generated ones, so I will keep Albany in my sample.

---

[9] The idea can be expressed in a different way. We can get all the students out of the schools, mix them and place them back again in the schools randomly.

[10] Kolmogrov-Smirnov test is well known. The Run Test is presented (in Stata Manual) as a "nonparametric test of the hypothesis that the observations occur in a random order by counting how many runs there are above and below a threshold. It can be used to test the null hypothesis that two samples are drawn from the same underlying distribution. The run test is sensitive to differences in the shapes, as well as the locations, of the empirical distributions."

I've found that both tests don't perform well with the randomly generated data (they under reject Ho, saying that the two densities are the same, when they aren't). Then I've checked that the two tests perform properly if all the densities are normal (i.e. they give the right proportion of significance at the 10%, 5% and 1% level). So with a bootstrap procedure with the fake random densities I have generated the new test values, based not in the tables but in the data. Recall that the random (fake) densities maintain the structure of the schools (number of schools in the country, number of students in each single school, etc), and then assign students randomly to those slots, and calculates the randomly generated proportion of females in each school. The density of the percentage of females in the schools randomly generated in this way may not be a normal density. The idea is to generate many different random densities, and then compare two of them and get the t-values of the equality of the densities (with the two tests). Then perform the comparison many times, and save the t-values. Those t-values (obtained by comparing two randomly generated densities) will have a distribution. I will get the value(s) that separates the 5% of the observations. The Kolmogrov Smirnov test is a one tail test, so I get the t-value that leaves the 95% of the other values at his left. In the case of the Run Test, it has two tails, so I get the value that leaves the 2.5% of the observations at his left, and also the t-value that leaves the 97.5% of the observations at his left. Those will be the test values that I will use to compare the PISA distribution with each of the randomly generated ones.

I repeat this bootstrap procedure for every country in the sample, and detect those countries where it is reasonable to conclude that the distribution of boys and girls into schools is random. For illustration purposes, I will show the figure for Italy, a country where the allocation of students by gender into schools is not generated by a random process. In Figure 8 we see that in Italy there are many schools with a greater proportion of girls (or boys) than the one that would have been observed by randomly allocating students into schools. There is a (strong) selection process taking place where boys go to some type of schools and girls to others. Italy won't be part of the analysis.

e. Each country is classified in one of three categories (A, B and C) depending on how random is the allocation by gender of students into schools[11]. In an online appendix I present the graphs for every country, which gives a nice visual illustration of why some countries are incorporated in the sample[12]. In Table 3 I list the classification of countries into the three categories[13]. There are 74 countries in the database, and I have classified 31 of them as A and 15 of them as B.

f. Finally, I perform the analysis with the countries where the allocation of students by gender appears to be generated by a random process.

To estimate the impact of the gender composition on test scores, I run the following regression:

$$y_{isc} = \alpha + \beta females_{-is} + \boldsymbol{X}\gamma + \mu_c + \varepsilon_{isc} \qquad \textbf{[2]}$$

where y_{it} is the test score of student i from school s in country c. The parameter of interest β measures the impact on academic performance of the proportion of 15 year old students in school s[14]. Country dummies are represented by μ_c . The vector $\boldsymbol{X}$ controls for pre-determined students characteristics in order to reduce (part of) the remaining selection (grade retention of each student in primary and secondary education, family wealth and mothers highest schooling). It also includes school controls (a quadratic function of school enrollment and dummy variables for school community: village, small town, big city, etc). Standard errors are clustered at the country level[15].

In the previous section, the experiment in Kenya was based in the variation between classes (groups) in the same school. Now I will look in the variation *between* schools. The database is not organized at the group level. But even if that was the case, and information on each group was available, there could be selection into one or other group in the same schools. The experiment from

[11] Countries in category A pass the Kolmogrov Smirnov and the Run Test. So there are no statistical differences between the randomly generated distribution of students into schools, and the one from PISA data. But both tests in some cases may give different answers. So countries in category B are the ones that pass one test, but not the other. And countries in category C are those who don't pass any tests (the PISA density is very different from a randomly generated one).

[12] http://www2.um.edu.uy/jmcabrera/Research/country%20figures.pdf

[13] From the original PISA sample I have dropped three countries because they had less than 20 coeducational schools (not big enough to perform the analysis). These countries are Jordan, Liechtenstein and Malta. The average country that remains in the sample has 229 schools (with a minimum of 34 and a maximum of 1.482).

[14] The proportion of females in the school $female_{-is}$ was calculated leaving student i out of the mean. Test scores have not been adjusted for PISA plausible values.

[15] They allow for correlation between students in the same country and suppose that students from different countries are independent. If one were to consider that students are linked in a school, but are independent in different schools in a country, then standard errors could be clustered at the school level (and estimated parameters would have smaller standard errors than clustering at the country level).

Kenya had randomization at the classroom level, but that is not what happens in non-experimental settings. Hoxby (2000) and Lavy and Schlosser (2011) neither use variation between classes; they compare between adjacent cohorts within the same grade and school (since selection into classes or groups may be endogenous). In my case, treatment is to be in a school with a higher proportion of 15 year old students. This can help in the case of different retention rates. Ciccone and García-Fontes (2014) have a model that proposes to use gender composition variation across birth cohorts within a school, instead of using the variation at a given grade level in different years. They argue that their approach is better when there is grade retention.

Some aditional clarifications are in order. I am not arguing that there is no selection of students into schools, but that, in some countries, there is no selection *by gender* of the students. Obiously, there is selection acording to other observable characteristics of students: wealth, mothers education, etc. For example, there are schools where the sons of wealthy parents go in a greater proportion, but a wealthy parent may send his sons and also his daughters to that school, so there is selection by wealth but not by sex of the student. To ilustrate this point, take the example of Spain, where the alocation of students into schools *by gender* is random in the data (Figure 9). I have plotted in gray how the random allocation of students into schools would have looked regarding parents wealth, mothers education and grade retention of students. We can see that although there is not a selection process by gender, there is clearly selection of students into schools by wealth of the parents, eductation of the mother, or grade retention of the students. If I were to compare the impact of wealth on academic achievement, the results would contain a strong selection bias, since the allocation of students into schools is not random. But not for the geneder of students.

Second Strategy using PISA data: School competition.

If there is only one school located in a region, there should be less selection from parents ("I don't send my daughter to that school because there are many boys"). So a second strategy using PISA data is to look at the impact of the proportion of females on test scores only for schools that don't face competition from other schools in the location[16]. With this strategy, I can use 71 countries in the database since it doesn't matter how students sort into different schools at the country level, if in the *locality* where the family lives there is only one school. The percentage of females should be fairly exogenous in this case, and may be related to demographic variations in the proportion of females in the catchment area of each school. As a further refinement, I will keep only public schools, since private schools may exert monopsony power if they don't face competition, and may choose for example to receive a higher proportion of females (by rejecting some boy applicants)[17].

The external validity may be smaller (or different) than strategy #1, since families in areas where there is just one school may be different from the rest of the families: e.g. some wealthy parents may choose to move to other neighborhood if there is only one school where they live.

The proportion of females in schools that face no competition (in the whole sample) is very similar to the proportion of females in those schools that face competition (Figure 10).

[16] The exact wording in the PISA questionnaire to schools principals is: "We are interested in the options parents have when choosing a school for their children. Which of the following statements best describes the schooling available to students in your location? (Please tick only one box)

1. There are two or more other schools in this area that compete for our students
2. There is one other school in this area that competes for our students
3. There are no other schools in this area that compete for our students".

[17] In locations where the school faces no competition, only 5% of the students attend private schools.

3.2 Results

First Strategy: Country level variations.

In Table 4 we can see the impact that a greater proportion of females has on students' academic performance. Results, using different sample of countries, consistently show that students (boys and girls) benefit from being in a school with more females students. Conversely, students from schools with a greater proportion of boys perform significantly worse. The positive results are in the math and reading tests. These results are consistent with Hoxby (2000) and Lavy and Schlosser (2011) who, in general terms, show that an increase in the proportion of females in the classroom increases both boys and girls cognitive outcomes. But now I obtain this same pattern with a much more general data.

Regarding the magnitude of the effects, the impact on boys math scores is a 4.8% increase of a standard deviation, for each 10pp increase in the proportion of females. In the nonlinear case (not shown), the estimated impact depends on the point of the distribution. A 10pp increase in the proportion of females, from 50% to 60% in a school, causes an increase in math test scores of 4.8% of a sd for boys (just as in the linear case). But a 10pp increase from 60% to 70% of females, has a 1.1% impact on cognitive skills. Results for girls are positive and significant, but smaller in magnitude.

Second Strategy: School that face no competition

Now I will look at another sample of schools, where there should be little selection (from parents and/or school principals). These schools are located in areas where they report that they face no competition from other schools for their students. Furthermore, I will keep only public schools, as explained in the empirical strategy section. In Table 5 we can see that the effects are consistent with the first strategy: both boys and girls improve math and reading test scores when the percentage of females increases. Results are smaller in magnitude for boys and bigger for girls than the ones estimated with the first strategy.

3.3 Other results

In this section I will (a) present a placebo test, and then study (b) if the effects are non-linear (higher in some parts of the distribution of the percentage of females in a school; (c) heterogeneous effects (i.e. by socioeconomic status or for migrants); (d) and possible channels that can drive the results (i.e. if a school with a greater proportion of females has a better educational climate). For the sake of brevity I will only show peer effect results using the first strategy with PISA data[18].

Placebo test. The placebo exercise consists in randomly allocating students into schools, so the percentage of females constructed with this simulated data shouldn't have an impact on a student's test score, since there are no real social forces operating. I get all the students out of the schools, and then randomly place them back into the schools. Each student has his real characteristics (test scores, family background, etc), but the peers he now faces are faked, in particular, the proportion of females in his "new" school is not real. In this setting, we expect not to find evidence of peer effects, even with the high number of observations in the database that had allowed me to estimate peer effects with small confidence intervals in the previous section. In Table 6 we see that in fact there is no evidence of peer effect when the simulated school data is used.

[18] Another question, that I haven't answered at this stage of the project, is whether the impact differs at different points of the conditional distribution of ability. Do higher ability boys benefit more or less with a greater proportion of females?

Angrist (2014) shows that there may be a mechanical relationship between own and peer characteristics that leads to an estimation of peer effects, even if there are no underlying real social relations. In this exercise I have constructed the proportion of females using the placebo data, and found that in this setting with no real social interactions, there are no peer effects, as one would have expected.

Non-linear effects. Figure 11 shows the estimated impact in the linear (benchmark) model for Math (results from columns 1 and 2 in the first panel of Table 4). We see that (a) boys have better results in math than girls; (b) both boys and girls benefit when the proportion of females in the generation increases; (c) the impact is bigger for boys than for girls (the slope is higher for boys, increasing the gender gap in math). It also shows the estimated magnitudes. The estimated increase in test scores for girls goes from 424 points when there are 0% girls in the school generation, to 460 points when the proportion of females is 100%. In the case of men the estimated impact of 50.9 points, runs from 432 to 483 score in math.

It is interesting to compare the results for math with the ones for reading. In both cases, students benefit when they are placed in a school with more proportion of females. The positive impact is larger for boys. In the case of reading, where girls perform significantly better than boys, the gap would be closed in the hypothetical case that that student attended an all-girls school (if the effects were linear, but that is not the case, as the next figure shows).

In Figure 12 we observe non linear effects, from a model that includes a third order polynomial on the percentage of females. Gender peer effect have a different impact, depending on the proportion of females. The maximum score for a student (girl or boy, in math or reading) is when the proportion of females is close to 70%.

Channels. Is there a better classroom climate when there are more females? And, on the other hand, is it more difficult to work when there are many male students? Students are asked about the climate in their classes, in a four point scale, from "never or hardly ever" to "in all lessons"[19]. The regressions are similar to equation [2], except for the fact that I don't separate between boys and girls (a gender dummy variable is added to the controls). I find that a greater proportion of females is associated with a significant better climate in the classroom (Table 7). If there are more girls, students listen more to what the teacher says, there is less noise and disruption, students work better, etc.

Heterogeneous effects. Do students from lower socioeconomic background or immigrants benefit more? Although a higher proportion of females could be more beneficial for disadvantaged groups (Lavy and Schlosser, 2011), no such evidence is found in the data (Table 8). The main coefficient of interest is the interaction term in the third row of each panel. It is not significant, so the proportion of females doesn't have a differential impact according to the socioeconomic background of students. The

[19] The wording of the question is: "How often do these things happen in your <test language lessons>?

a) Students don't listen to what the teacher says
b) There is noise and disorder
c) The teacher has to wait a long time for the students to <quieten down>
d) Students cannot work well
e) Students don't start working for a long time after the lesson begins

Classroom climate variables are constructed as dummy variables. They are equal to one (bad climate) if the student answers "in most lessons" or "in all lessons", and is set to zero if the student answers "never or hardly never" or "in some lessons".

other result is that, as expected, students from a more disadvantaged context perform significantly worse at school.

4 Conclusion

I have studied the impact of gender peer effects in different settings. In a RCT with first grade children in Kenya, I have found that there is a positive impact of the proportion of female classmates for boys. But results are located in math scores, and only in the long run.

Since those results are very local by construction (they have small external validity) I propose two procedures to credibly estimate gender peer effects in a much general context. I construct different samples of countries and schools with PISA data where the allocation of students to schools by gender seems random. Results show a positive impact on test scores from a higher proportion of females in a school. Both boys and girls benefit, in math and in literature. Gains seem larger for boys. A possible channel is a better relation with teachers. These results also hold for public schools that don't face competition in their area, so there should be less selection in their intake of students. The maximum score for a student (girl or boy, in math or reading) is when the proportion of females is close to 70%. I am confident that a group with more females has a positive influence on its members. My next step will be to study policy implications.

References

Booth, Alison L., Lina Cardona-Sosa and Patrick Nolen (2013). “Do Single-Sex Classes Affect Exam Scores? An Experiment in a Coeducational University.” WP.

Ciccone, Antonio and Walter Garcia-Fontes (2014). “Gender Peer Effects in School, a Birth Cohort Approach”. CEPR Discussion Paper No. DP10042.

Duflo, Esther, Pascaline Dupas and Michael Kremer (2011). “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya.” *American Economic Review, 101(5): 1739-74.*

Hoxby, Caroline (2000). “Peer Effects in the Classroom: Learning from Gender and Race Variation.” NBER WP 7867.

Lavy, Victor and Analía Schlosser (2011). “Mechanisms and impacts of gender peer effects at school.” *American Economic Journal: Applied Economics, 3(2): 1-33.*

Lee, Soohyung, Lesley J. Turner, Seokjin Woo and Kyunghee Kim (2014). “All or Nothing?. The Impact of School and Classroom Gender Composition on Effort and Academic Achievement”. NBER WP 20722.

Oosterbeek, Hessel and Reyn van Ewijk (2014). “Gender Peer Effects in University: Evidence from a Randomized Experiment.” *Economics of Education Review, 38: 51-63.*

Pritchett, Lant and Justin Sandefur (2014). “Context Matters for Size: Why External Validity Claims and Development Practice Don’t Mix”. *Journal of Globalization and Development. Volume 4, Issue 2.*

Whitmore, Diane. (2005). “Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment.” *American Economic Review P&P, 95(2): 199-203.*

Appendix

Figure 1

Simulated random distribution of students into schools

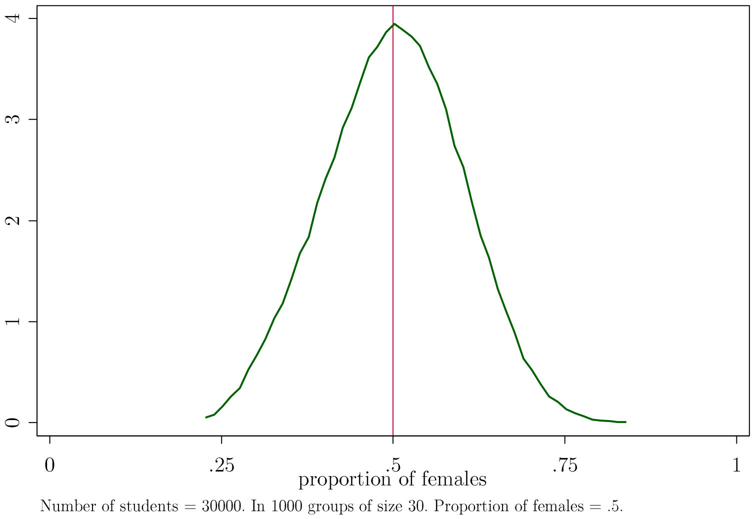


Figure 2

Proportion of females in the classroom

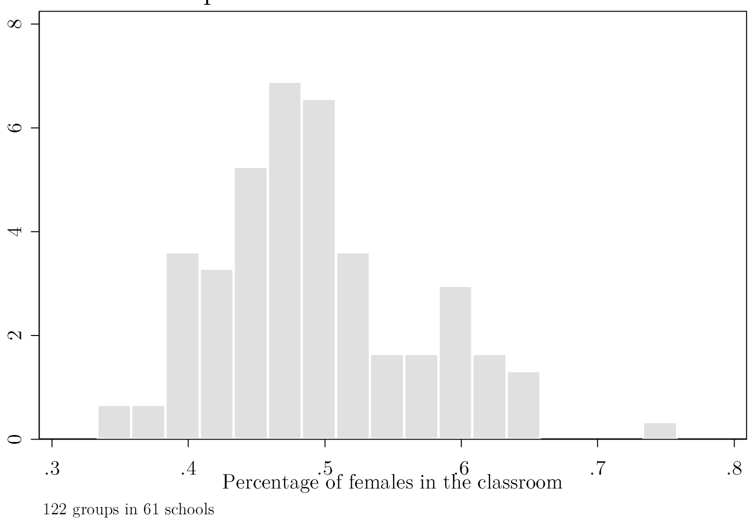


Table 2

Gender Peer Effects in Kenya

	(1)	(2)	(3)	(4)
	Short-run effects		Longer-run effects	
Total Score				
% Females	-0.061	-0.106	0.787	0.74
	(0.578)	(0.422)	(0.464)*	(0.354)**
Girl	0.079	0.046	0.161	0.112
	(0.041)*	(0.035)	(0.042)***	(0.037)***
Other Controls	NO	YES	NO	YES
Observations	2,814	2,812	2,661	2,660
R-squared	0.18	0.36	0.15	0.32
Math Score				
% Females	0.346	0.353	0.841	0.832
	(0.582)	(0.501)	(0.412)**	(0.383)**
Girl	0.018	0.001	0.053	0.024
	(0.038)	(0.036)	(0.041)	(0.037)
Other Controls	NO	YES	NO	YES
Observations	2,814	2,812	2,661	2,660
R-squared	0.13	0.31	0.10	0.26
Literature Score				
% Females	-0.39	-0.469	0.59	0.53
	(0.541)	(0.383)	(0.479)	(0.347)
Girl	0.115	0.074	0.212	0.155
	(0.044)***	(0.037)**	(0.043)***	(0.038)***
Other Controls	NO	YES	NO	YES
Observations	2,815	2,813	2,663	2,662
R-squared	0.21	0.33	0.18	0.31

Standard errors, reported in parentheses, are clustered at the classroom level.

*** p<0.01, ** p<0.05, * p<0.1.

There are 61 schools (and 122 groups) used in the analysis. Students were randomly assigned to one of the two groups in each school.

All regressions control for school fixed effects and gender. *Controls* in columns 2 and 4 include age, being assigned to the contract teacher, average baseline score of classmates, own baseline score, and dummy variables for imputed baseline scores.

Figure 3

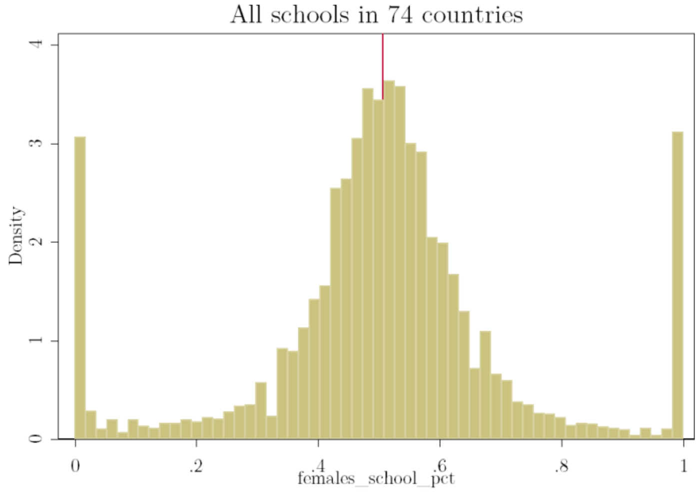


Figure 4

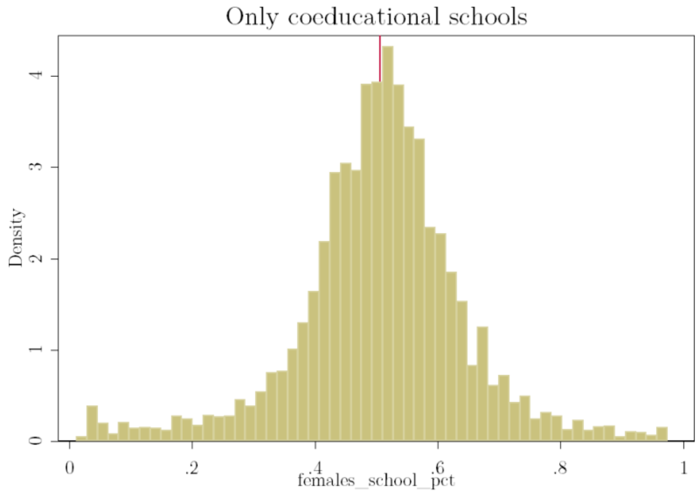


Figure 5

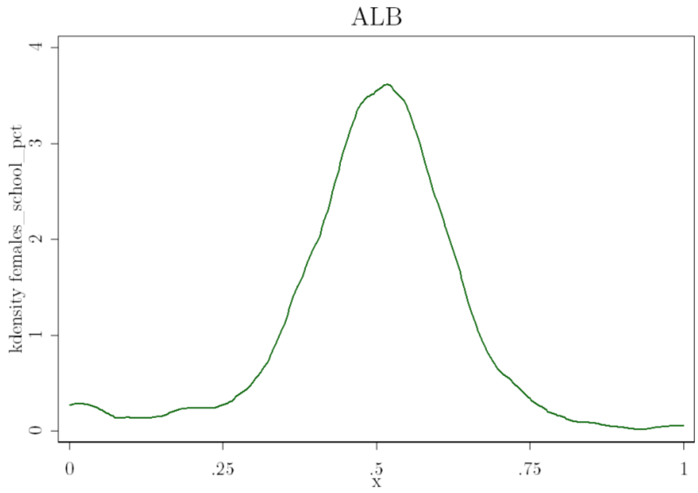


Figure 6

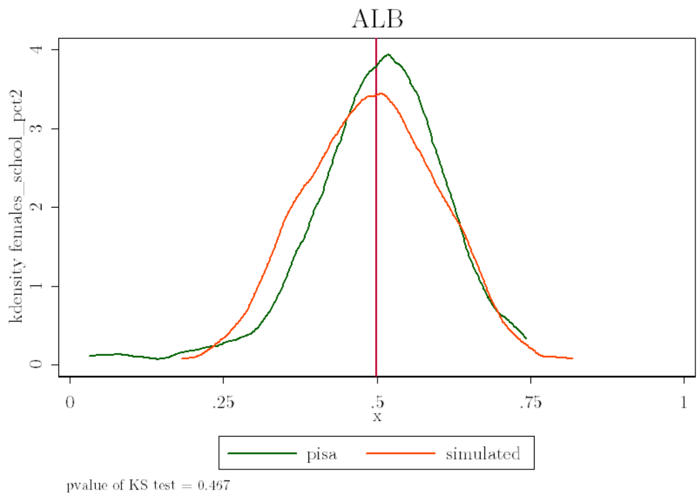
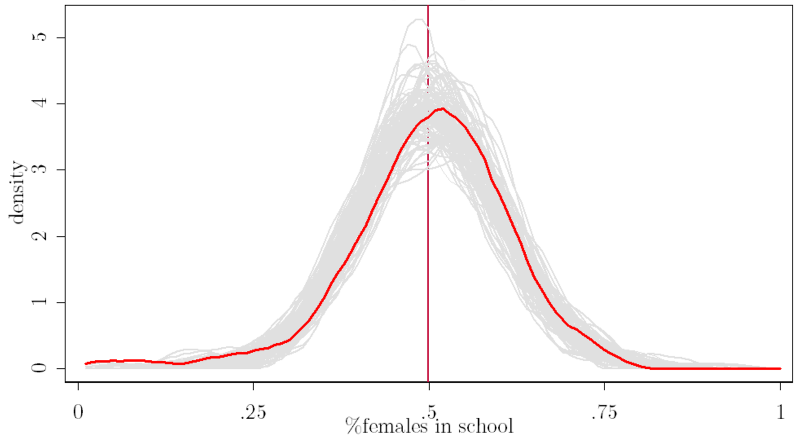


Figure 7

ALB

Country Category: A.

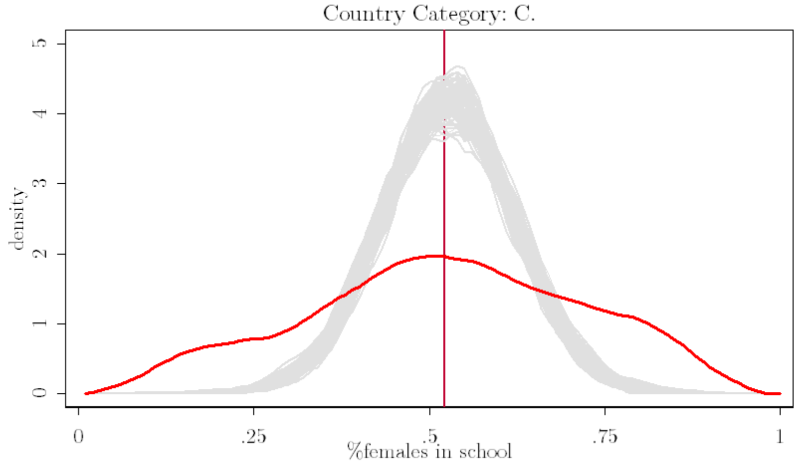


Simulated densities ploted: 100.
pval RunTest ok: YES. // pval KSmirnov ok: YES.
5% significance level for tests

Figure 8

ITA

Country Category: C.



Simulated densities ploted: 100.
pval RunTest ok: NO. // pval KSmirnov ok: NO.
5% significance level for tests

Table 3

Country classification

Country	Kolmogrov-Smirnov OK	RunTest OK	Category	Country	Kolmogrov-Smirnov OK	RunTest OK	Category
1 Albania	1	1	A	38 Liechtenstein	.	.	.
2 United Arab Emirates	1	1	A	39 Lithuania	1	1	A
3 Argentina	0	0	C	40 Luxembourg	0	0	C
4 Australia	1	0	B	41 Latvia	1	1	A
5 Austria	0	0	C	42 Macao-China	1	0	B
6 Azerbaijan	1	1	A	43 Republic of Moldova	1	1	A
7 Belgium	0	0	C	44 Mexico	1	0	B
8 Bulgaria	0	0	C	45 Malta	.	.	.
9 Brazil	0	1	B	46 Montenegro	0	0	C
10 Canada	0	0	C	47 Mauritius	1	1	A
11 Switzerland	0	0	C	48 Malaysia	0	1	B
12 Chile	1	1	A	49 Netherlands	0	0	C
13 Colombia	1	1	A	50 Norway	0	0	C
14 Costa Rica	0	1	B	51 New Zealand	0	0	C
15 Czech Republic	0	0	C	52 Panama	1	1	A
16 Germany	1	1	A	53 Peru	1	1	A
17 Denmark	1	1	A	54 Poland	0	0	C
18 Spain	1	1	A	55 Portugal	1	1	A
19 Estonia	0	1	B	56 Qatar	1	1	A
20 Finland	0	0	C	57 Shanghai-China	1	1	A
21 France	1	1	A	58 Himachal Pradesh-Ind	1	1	A
22 United Kingdom	0	0	C	59 Tamil Nadu-India	1	0	B
23 Georgia	1	1	A	60 Miranda-Venezuela	1	1	A
24 Greece	1	0	B	61 Romania	0	0	C
25 Hong Kong-China	1	1	A	62 Russian Federation	1	1	A
26 Croatia	0	0	C	63 Singapore	0	1	B
27 Hungary	0	0	C	64 Serbia	0	0	C
28 Indonesia	0	0	C	65 Slovak Republic	0	0	C
29 Ireland	1	1	A	66 Slovenia	0	0	C
30 Iceland	1	1	A	67 Sweden	0	0	C
31 Israel	1	1	A	68 Chinese Taipei	1	0	B
32 Italy	0	0	C	69 Thailand	0	0	C
33 Jordan	.	.	.	70 Trinidad and Tobago	1	1	A
34 Japan	1	0	B	71 Tunisia	1	0	B
35 Kazakhstan	1	1	A	72 Turkey	0	0	C
36 Kyrgyzstan	0	1	B	73 Uruguay	0	0	C
37 Korea	1	0	B	74 United States	0	0	C

Countries in category A are the ones were the distribution of students by gender into schools isn´t different from a random allocation, acording to Kolmogrov-Smirnov and RunTest, at the 95% level, using 2000 simulated random allocations. Countries in category B pass one of the two tests. If both KS and RT reject the null hypothesis of equality of the PISA distribution of students to the randomlly generated one, a country is classified as C. The detailed procedure is explained in the text.

Figure 9

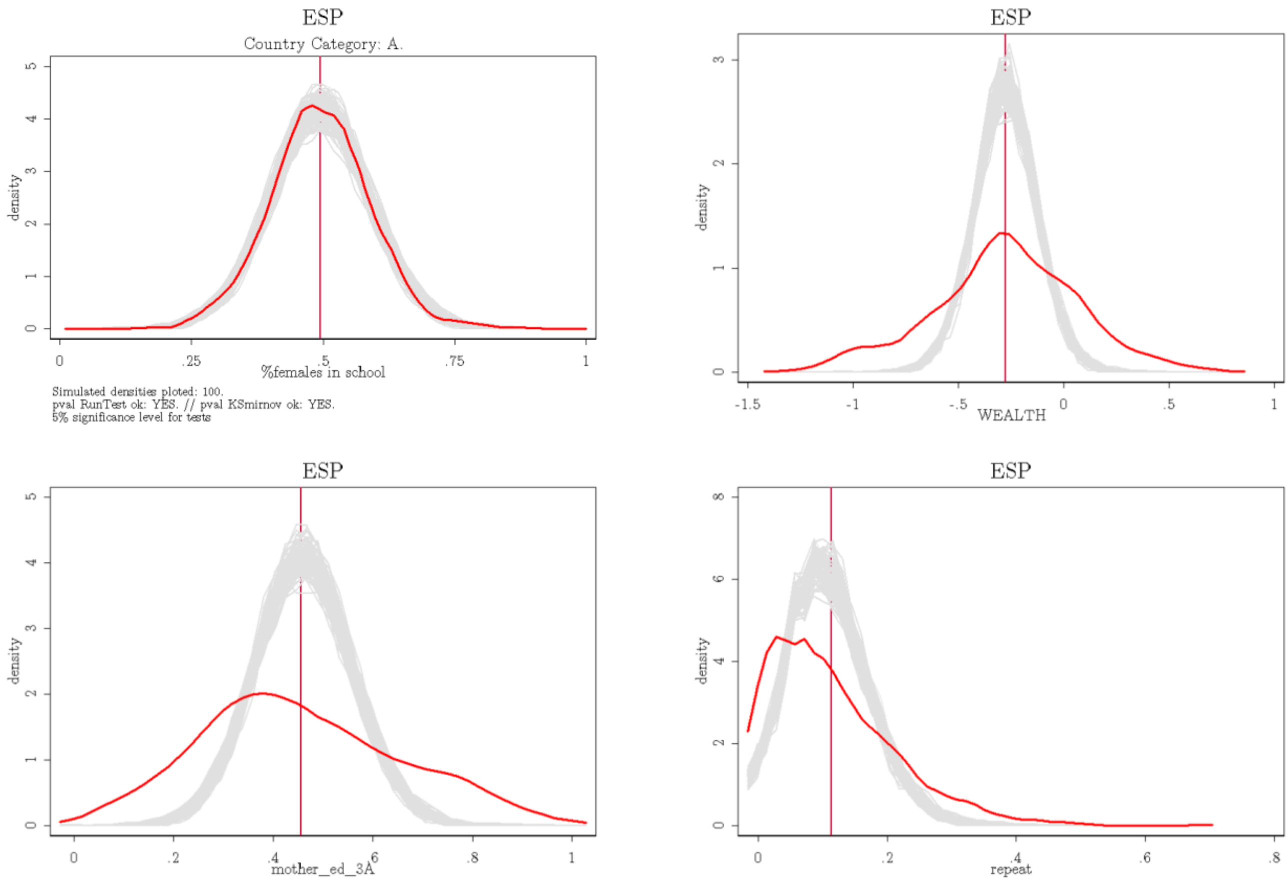
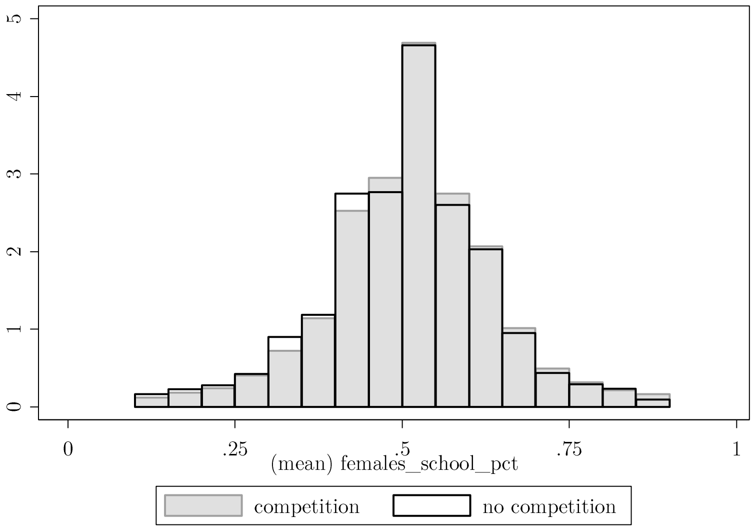
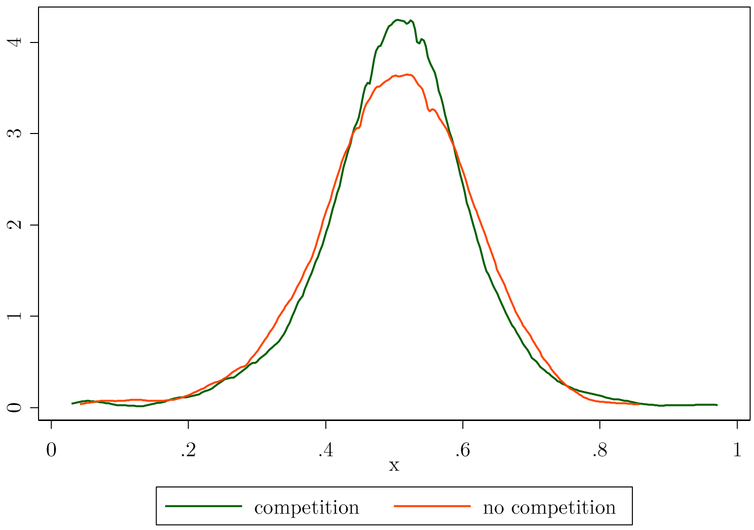

ESP
Country Category: A.
density
%females in school
Simulated densities ploted: 100.
pval RunTest ok: YES. // pval KSmirnov ok: YES.
5% significance level for tests
ESP
density
WEALTH
ESP
density
mother_ed_3A
ESP
density
repeat

Figure 10



All countries



only A and B countries

Table 4

Peer effects around the world

(First Strategy with PISA data)

	(1)	(2)	(3)	(4)
	Math Results		Reading Results	
	Boys	Girls	Boys	Girls
Countries category "A" and "B"				
% Females	50.90***	35.86***	54.88***	39.28***
	(12.25)	(9.2)	(11.25)	(8.36)
Observations	119,260	123,202	119,260	123,202
Countries category "A"				
% Females	58.91***	37.36**	56.62***	37.17***
	(18.3)	(13.71)	(15.75)	(11.54)
Observations	62,528	62,802	62,528	62,802

*** p<0.01, ** p<0.05, * p<0.1

Standard Errors in parenthesis clustered at the country level. Controls include: country dummies, family wealth, school size and it square, and full set of dummy variables for students grade retention in primary education, in secondary education, mothers highest schooling, and school community.

There are 46 countries in category A+B and 31 countries in category A.

Table 5

Second strategy with PISA data (Public Schools that face no competition)				
	(1)	(2)	(3)	(4)
	Math Results		Reading Results	
	Boys	Girls	Boys	Girls
All countries				
% Females	53.81***	41.12***	70.34***	54.13***
	(10.66)	(8.62)	(9.69)	(6.54)
Observations	43,379	44,425	43,379	44,425
Countries category "A" and "B"				
% Females	45.98**	50.74***	50.13***	48.71***
	(17.82)	(15.02)	(16.33)	(13.23)
Observations	19,919	20,503	19,919	20,503

*** p<0.01, ** p<0.05, * p<0.1

Standard Errors in parenthesis clustered at the country level. Controls include: country dummies, family wealth, school size and it square, and full set of dummy variables for students grade retention in primary education, in secondary education, mothers highest schooling, and school community.

Estimations using public schools in 71 countries, located in areas were they face no competition for students. There are 46 countries in category A or B.

Table 6

	(1)	(2)	(3)	(4)
Placebo Test				
(First Strategy with PISA data)				
	Math Results		Reading Results	
	Boys	Girls	Boys	Girls
Countries category "A" and "B"				
% Females	-4.52	0.85	-3.89	0.72
	(2.94)	(2.27)	(2.72)	(1.99)
Observations	119,203	123,232	119,203	123,232
Countries category "A"				
% Females	-5.26	1.87	-2.99	0.4
	(4.39)	(3.83)	(4)	(3.13)
Observations	62,529	62,794	62,529	62,794

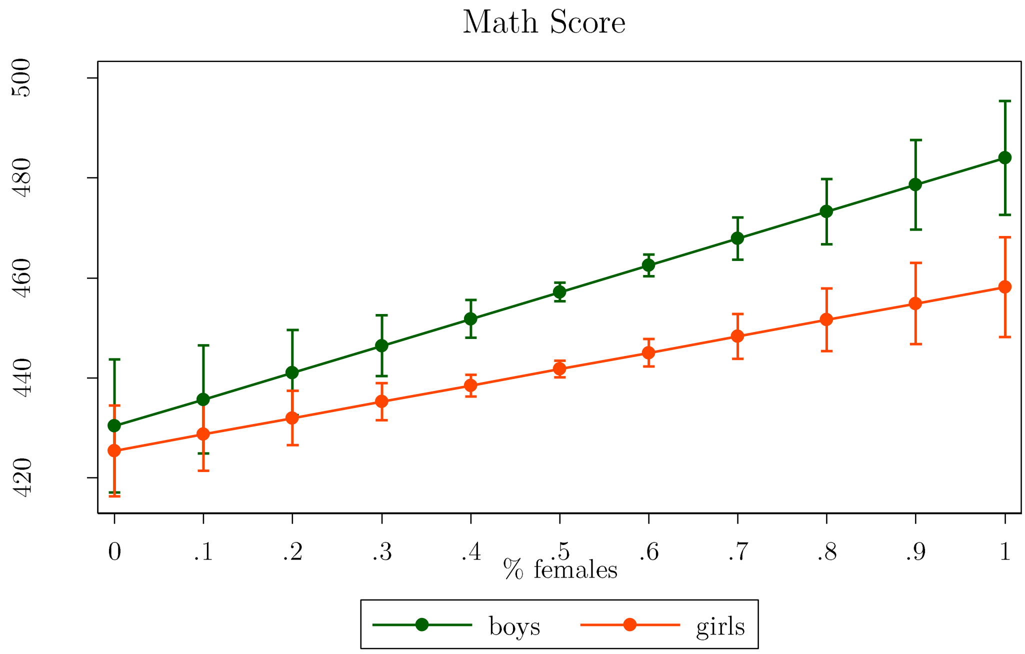
*** p<0.01, ** p<0.05, * p<0.1

The placebo exercise consists in randomly allocating students into schools, so the percentage of females constructed with this simulated data shouldn´t have an impact on a student's test score, since there are no real social forces operating.

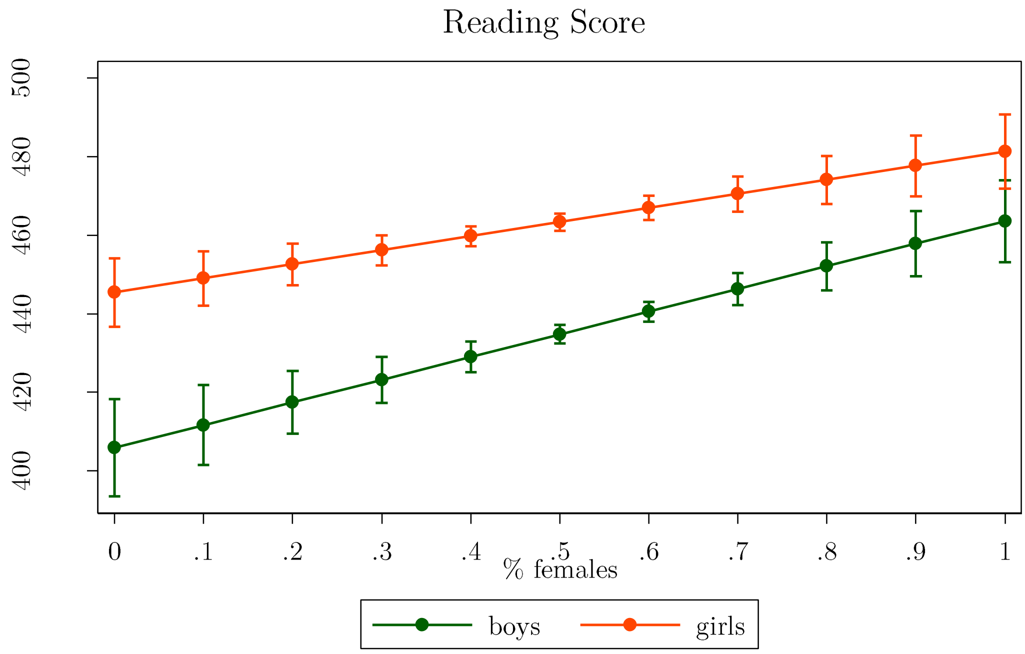
Standard Errors in parenthesis clustered at the country level. Controls include: country dummies, family wealth, school size and it square, and full set of dummy variables for students grade retention in primary education, in secondary education, mothers highest schooling, and school community.

There are 46 countries in category A+B and 31 countries in category A.

Figure 11



PISA data for 2009, A and B countries
SE are clustered at the country level. Model includes controls.
Dash lines are 95% confidence intervals

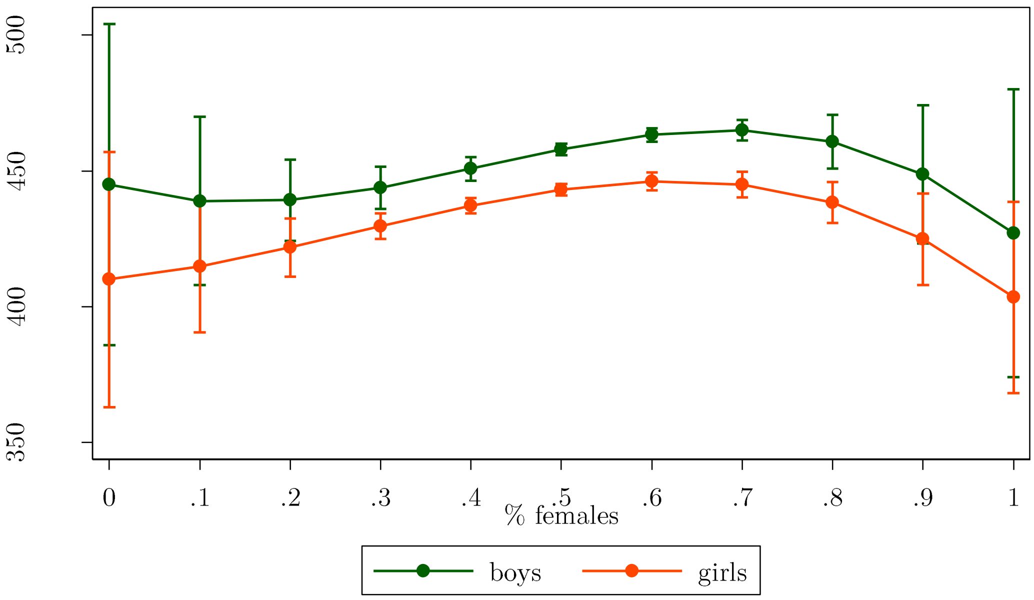


PISA data for 2009, A and B countries
SE are clustered at the country level. Model includes controls.
Dash lines are 95% confidence intervals

Figure 12

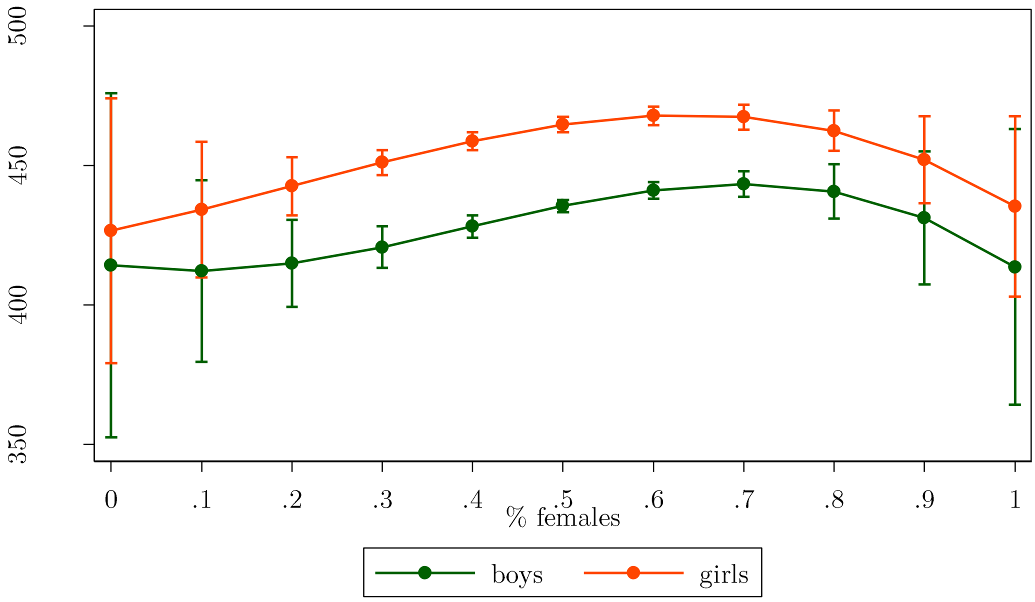
Impact of the percentage of females on test scores

Math Score



PISA data for 2009, A and B countries
SE are clustered at the country level. Model includes third order polynomial of percentage of females and controls.
Dash lines are 95% confidence intervals

Reading Score



PISA data for 2009, A and B countries
SE are clustered at the country level. Model includes third order polynomial of percentage of females and controls.
Dash lines are 95% confidence intervals

Table 7

Channels

	(1)	(2)	(3)	(4)	(5)
	Students don´t listen	Noise and disorder	Wait to quiet down	Sudents cannot work well	Sudents don´t start working
Countries category "A" and "B"					
% Females	-0.09***	-0.13***	-0.13***	-0.10***	-0.09***
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Observations	237,957	237,675	237,151	237,149	237,402
Countries category "A"					
% Females	-0.10***	-0.12***	-0.13***	-0.10***	-0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	122,703	122,528	122,264	122,232	122,356

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Each cell is a separte regression of the climate at the students class on the percentage of females. Boys and girls are included in the sample.

Standard Errors in parenthesis clustered at the country level. Controls include: country dummies, student gender, family wealth, school size and it square, and full set of dummy variables for students grade retention in primary education, in secondary education, mothers highest schooling, and school community.

There are 46 countries in category A+B and 31 countries in category A.

Table 8

Heterogeneous effects

(With First Strategy)

	(1)	(2)	(3)	(4)
	Math Results		Reading Results	
	Boys	Girls	Boys	Girls
Mother with low education				
% Females	48.80***	30.90***	54.86***	37.37***
	(12.65)	(10.35)	(12.24)	(9.295)
low educ mother	-22.64***	-31.37***	-19.74***	-26.32***
	(5.83)	(5.086)	(5.724)	(4.843)
*% Females*low educ*	5.87	14.02	0.491	4.882
	(9.66)	(9.843)	(9.643)	(9.39)
Observations	115,329	120,032	115,329	120,032
Parents with blue collar education				
% Females	47.45***	31.65***	50.41***	36.86***
	(13.06)	(10.8)	(11.75)	(9.627)
blue collar	-20.78***	-23.16***	-23.90***	-21.67***
	(4.475)	(4.043)	(4.423)	(4.134)
*% Females*blue collar*	3.216	9.147	4.613	4.285
	(8.061)	(8.267)	(7.786)	(8.476)
Observations	112,009	117,388	112,009	117,388
Inmigrant status				
% Females	48.64***	32.99***	52.87***	36.90***
	(11.84)	(8.593)	(10.9)	(7.87)
Inmigrant	-21.55*	-28.94**	-16.02	-25.29**
	(11.93)	(12.99)	(11.48)	(11.36)
*% Females*inmigrant*	25.72	31.75	15.99	25.08
	(23.89)	(28.62)	(22.69)	(25.79)
Observations	116,597	121,300	116,597	121,300

*** p<0.01, ** p<0.05, * p<0.1

Standard Errors in parenthesis clustered at the country level. Included countries: category "A" and "B". Controls: country dummies, family wealth, school size and it´s square, and full set of dummy variables for students grade retention in primary education, in secondary education, and school community. The third panel (Inmigrant status) also controls for mothers highest schooling dummies.