

Gender peer effects in university: Evidence from a randomized experiment¹

Hessel Oosterbeek

Reyn van Ewijk

¹This version: July 2011. We gratefully acknowledge valuable comments from Eric Bettinger, Adam Booij and Victor Lavy and from seminar participants in Amsterdam, Mannheim and Zürich. Oosterbeek is affiliated with the University of Amsterdam and the institute for evidence-based education research TIER. Van Ewijk is affiliated with VU University Amsterdam, Netspar and Tinbergen Institute. Emails: h.oosterbeek@uva.nl; rewijk@feweb.vu.nl.

Abstract

Recent studies for primary and secondary education find positive effects of the share of girls in the classroom on achievement of boys and girls. This study examines whether these results can be extrapolated to post-secondary education. We conduct an experiment in which the shares of girls in workgroups for first year students in economics and business are manipulated and students are randomly assigned to these groups. Boys tend to postpone, but not abandon, their dropout decision when surrounded by more girls and perform worse on courses with high math content. There is also a modest reduction in absenteeism early in the year. Overall, however, we fail to find substantial gender peer effects on achievement. This in spite of the fact that according to students' perceptions, both their own, and their peers' behavior are influenced by the share of girls.

JEL-codes: I22; I28; D83

Keywords: Field experiment; Peer effects; University students

1 Introduction

While in general the literature on peer effects in education shows mixed results, the few studies dealing with gender peer effects tend to agree that more girls in the classroom improve test scores (Epple and Romano, 2010; Sacerdote, 2010). Using data for 3-6 graders in Texas, Hoxby (2000) finds that a 10 percentage point rise in the share of girls to increases of around 4 and 8 percent of a standard deviation in reading and math scores, respectively. To identify these effects, she exploits variation in the share of girls in a grade in a school across adjacent cohorts. Although girls in general outperform boys, Hoxby argues that the impact on math scores is so large, that the effect is unlikely to be driven solely by peers' achievement. Applying a similar approach to data from Israeli primary, middle and high schools, Lavy and Schlosser (2011) also find gender peer effects, which they, too, argue are due to more than only peers' higher achievement. Lavy and Schlosser provide evidence that effects of the share of girls work through lowering classroom disruption and violence and through improving inter-student and student-teacher relationships. Black et al. (2010) apply the same method, now to 9th graders in Norway and report positive point estimates of the impact of the share of girls on math scores of boys and girls. These estimates are, however, not significantly different from zero, probably due to the small variation in the share of girls across cohorts within schools. Whitmore (2005) finds that pupils in kindergarten and second grade do better if there is a majority of girls in the class. First graders are not affected by gender peer effects, and third graders seem to be harmed when they are in a class with more than 50 percent girls.¹ Finally, using that an upper-secondary school in Switzerland forms girls-only classes to avoid that the few boys in this school are spread out too thinly, Eisenkopf et al. (2010) find a positive effect of single-sex education on math grades of girls.

These previous studies deal with gender peer effects in primary and secondary schools. At these education levels, the boys-girls ratio is virtually constant over time. Only the existence of single-sex schools changes the expected boys-girls ratio at the level of schools.² This is different at the post-secondary level. The boys-girls ratio varies across fields of study, and in recent decades the fraction of female students in higher education increased

¹ Whitmore exploits the variation in share of girls across classrooms induced by the random assignment of pupils to large and small classes in the STAR-experiment. In her estimations, she controls for students' own gender, free-lunch status, race, class size assignment and school fixed effects. As explanations for the deviating result for third graders, Whitmore suggests a developmental change in children around this age related to peer pressure, or girls-dominated classes progressing too rapidly for boys.

² One important question is then whether girls and boys that attend single-sex schools perform better than girls and boys that attend mixed-sex schools. Moreover, since most single-sex schools are girls-schools, the existence of such schools also affects the boys-girls ratio for other schools that draw from the same population. It is then legitimate to ask whether the boys and girls that are left behind, suffer from the presence of girls-schools.

considerably. This naturally leads to the question whether the increased participation of girls in higher education has an impact on achievement. Extrapolating the findings for primary and secondary school pupils, one would expect that this development raises academic performance. Such extrapolation ignores, however, that the interaction between boys and girls as well as their relative achievement are different at different ages.

This paper reports about a randomized experiment that we conducted amongst first year undergraduate students in economics and business at the University of Amsterdam. Students attend tutorial sessions for all compulsory courses in groups of around 40 students – so-called workgroups. During the entire first year, students are assigned to the same workgroup. This implies that they share around 9 of their 15 weekly teaching hours in this same group of students. We assume that this is a relevant peer group.³ For two consecutive years (2007/8 and 2008/9), we obtained permission to differentiate the share of girls across workgroups and to assign boys and girls randomly to these groups. While the share of girls in the total population is around 0.3, the share of girls in workgroups varies between 0.14 and 0.51. We examine the impact of the share of girls on dropout, absenteeism, and number of credits collected during various stages of the first year and for different types of courses.

By and large our results show very little evidence of gender peer effects of any relevant size. There appears to be a tendency amongst male students to postpone their dropout decision when they are surrounded by many girls in their workgroups. There is also a modest reduction in absenteeism early in the year. On the other hand, boys in workgroups with a high share of girls, do worse in courses with a high math content (mathematics and statistics). In search for mechanisms for potential peer effects, we find that students' perceptions of their own behavior and that of their peers is influenced by the share of girls. Students in groups with many girls find their groupmates more helpful, but are also more afraid to ask questions, and talk more often during class. This does, however, not affect their outcomes.

This paper sets – as Sacerdote (2010) phrases it – a next logical step in the literature on peer effects in education by “the designing of experiments to see whether the peer effects measured in observational data can be exploited by university administrators to maximize GPA or some other objective.” It is the first paper that looks at gender peer effects in university education.⁴ It is also the first study reporting about an experiment that explicitly randomizes students into different gender peer groups. Related to this, there is

³Defining the relevant peer group is not obvious. Some studies explore this issue by defining peer groups at different levels. For example, Sacerdote (2001) examines peer effects of roommates as well as of dormmates.

⁴Hansen et al. (2006) look at the effect of group diversity in terms of gender, age and race in an undergraduate management class. For a part of the class requirements students are assigned to four or five member teams. The study focuses mainly on the impact of group composition on group performance.

more variation in the share of girls in the data used in this study than in the data used in other studies. Commensurate with the difference in student populations between schools and university-level economics studies, our manipulation concentrates on situations in which girls generally form a minority, and our share of girls never exceeds 51 percent.

The paper continues as follows. The next section gives more details on the context and the data. Section 3 describes the design of the experiment. Section 4 presents and discusses the empirical results, and Section 5 summarizes and discusses the implications of our findings.

2 Context and data

2.1 Context

The bachelor program in economics and business at the University of Amsterdam has a nominal duration of 3 years. Students meeting the admission requirements are automatically accepted for the study without further selection.⁵ In the first academic year, which runs from September until August, all students in economics and business follow exactly the same program. In the years that the experiment was conducted, the first year program was divided into four terms of 7 weeks each. Since the first year program is fixed, students cannot substitute easy for difficult courses.⁶ Each term ended with exams shortly after the courses finished and the re-take exams are organized later in the year. The first academic year thus consists of 28 study weeks, which are allotted to different courses in the form of 60 credit points. The total number of first year students in the economics and business program is 593 in 2007/8 and 606 in 2008/9.

Teaching during the first year takes place in the form of central lectures for all first year students together and in workgroup meetings for groups of at most 40 students. In workgroup meetings students typically receive in depth explanation of the material, ask questions, and practice and discuss exercises and assignments. The instructor of a workgroup is in most cases a member of the permanent academic staff. Students are assigned to a specific workgroup before the start of the year and are supposed to stay in the same group for the entire first year. In each year, there were 14 workgroups.

The curriculum in the first year consists of fourteen compulsory courses. Table 1 lists the courses together with their scheduling in the year and their study load in terms of total teaching hours, workgroup hours and credit points. This table shows that most of total teaching hours – slightly over 60% – takes place in workgroup meetings, while the other

⁵The main requirement is that students graduated from the highest (pre-university) track in secondary education.

⁶Only after the first term of their second academic year students choose different packages of courses to specialize either in economics or in business.

Table 1. Overview of the first year courses in the economics and business program

Course	Term	Total teaching hours	Workgroup hours	Credit points
Financial accounting	1	28	14	5
Organization	1	12	12	5
Orientation fiscal economics	1	6	0	2
Mathematics 1	1/2	56	28	5
Academic skills 1	1/2	28	28	2
Management accounting	2	28	14	4
Microeconomics	2	42	28	7
Organization and management	3	28	14	6
Statistics	3	42	14	5
Mathematics 2	3/4	56	28	4
Academic skills 2	3/4	28	28	3
Finance	4	21	21	5
Macroeconomics	4	42	28	7
Total		417	257	60

40% take place in central lectures attended by several hundred students. We therefore assume that the other students assigned to the same workgroup are a relevant peer group.

2.2 Data

Our main data come from the student administration of the department of economics and business of the University of Amsterdam.⁷ They contain students' background characteristics, their workgroup assignment and their study performance and study status during the first year. The bottom part of Table 2 shows descriptive statistics of students' background characteristics, in the last two lines for boys and girls separately.

Students enrolling in the economics and business program of the University of Amsterdam score GPA's (overall and for math) somewhat below 7 (on a scale from 1-10). Girls outperform boys with regards to the GPA's in secondary school. The average age at the moment of enrollment is 19 years and three months. Students who enroll without any delay, would on average enter at the age of 18 years and six months, indicating that a substantial share of the students enters with a delay of one year or more. Three quarters of the enrolling students graduated from the economics/society specialization from secondary school, and another 18% followed the science/health specialization. With respect to age and secondary school specialization, we observe no large differences between boys and girls.

⁷We also collected additional data through a survey amongst students. We describe (and report about) this data source in Subsection 4.6.

Table 2. Balancing

	Application	Math grade	No Math	GPA in	Age	Secondary school track	
	date (pctile)	secondary school	secondary school	secondary school	(5)	Economics	Science/Health
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share girls	8.9	0.29	-0.09	0.13	0.10	0.05	0.05
	(10.7)	(0.26)	(0.20)	(0.16)	(0.32)	(0.29)	(0.14)
N	800	682	800	678	725	800	800
Scale	1-100	1-10	0-1	1-10	yrs	0-1	0-1
Mean	45.5	6.79	0.16	6.68	19.24	0.67	0.18
SD	30.4	0.94	0.36	0.46	1.31	0.47	0.38
Mean boys	46.8	6.67	0.13	6.59	19.32	0.69	0.19
Mean girls	42.7	7.10	0.22	6.88	19.06	0.62	0.15

Note: Coefficients refer to the effect of an increase of the female share from 0% to 100%. Table shows results from regressions of individual students' characteristics on share of girls, controlling for students' own sex and dummy for advanced math group. Application date is calculated as a percentile score per cohort. Students without math in secondary school either followed pre-university education but did not choose to take math, or came in via another route than pre-university education, or had missing background characteristics.

3 Design

For the experiment we manipulated the share of girls in first year workgroups and assigned boys and girls randomly to these groups. The 2007/8-cohort consists of 33 percent girls, the 2008/9-cohort of 30 percent. Beforehand we decided that one third of the workgroups would contain few girls (around $\frac{1}{6}$), one third of the groups would have an average share of girls (around $\frac{1}{3}$), and one third of the groups would have many girls (around $\frac{1}{2}$).

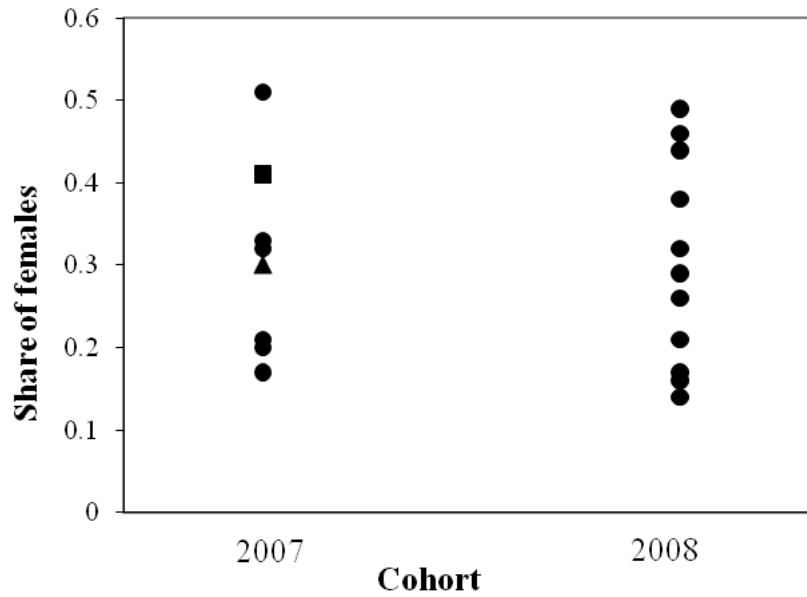
Students can apply in the period between May and September for the academic year that starts in September. Shortly after they apply and are accepted, students are sent an acceptance letter which, among other things, informs them about the workgroup to which they are assigned. Because of this system of rolling acceptance, new applicants are assigned to groups every week or every two weeks. In doing so, we kept the shares of girls as close as possible to the predetermined shares. This implies that boys and girls are stratified by week of application. As the week of application is probably related to students' motivation and is a strong predictor of their subsequent performance, this should keep average motivation/ability of boys and girls assigned to different groups the same.

Students who took the more advanced (instead of the standard) mathematics track in secondary education are assigned to separate workgroups. In each year there were only two of such workgroups. The shares of girls in these groups were manipulated as well and were 0.30 and 0.41 in 2007/8 and 0.14 and 0.38 in 2008/9. Dropping these four groups from our analyses does not change our results. Very late appliers were placed into separate groups where the share of girls was not manipulated. These groups are removed from all our analyses.⁸

To examine whether girls and boys that are assigned to groups with different shares of girls are indeed not systematically different, we regressed a set of observable student characteristics on the share of girls in their groups. We thereby control for the students' own sex and a dummy for the groups consisting of students with an advanced math background. As the top part of Table 2 shows, the share of girls in workgroups does not systematically co-vary with application date (as a proxy for students' motivation), nor with performance or track choice in pre-university education, nor with age. We conclude that our randomization scheme worked properly.

Students assigned to groups with many girls thus do not differ in characteristics from other same-sex students assigned to groups with few girls. However, boys and girls in general differ in average background characteristics (e.g. girls on average had higher GPAs in high school), so that groups with high and low shares of girls differ in more

⁸This explains why the mean of the percentile rank in Table 2 is lower than 50.



Note: Each dot represents one tutorial group. A triangle represents two groups and a square represents three groups.

Figure 1. Tutorial groups and their share of females

dimensions than only in the share of girls. This is unavoidable, unless we would stratify on a large range of characteristics. This implies that the gender peer effects that we estimate are gross effects; they also include the effect of girls being different from boys in other characteristics than just in their gender. This issue is not specific to the randomized design used in this study, it arises in all other gender peer studies as well. To see to what extent a larger share of girls comes together with other peer characteristics, we aggregate our data by group and calculate correlations between shares of girls and group averages of the characteristics listed in the top row of Table 2, thereby conditioning on being placed in an advanced math group. Two correlations are substantial and significantly different from zero. The conditional correlation between the share of girls and group-average GPA equals 0.39 (s.e. 0.18) and the conditional correlation between the share of girls and group-average math grade in secondary school equals 0.68 (s.e. 0.26). This implies that part of the gender peer effects that we report in this paper may operate through (an interaction with) ability peer effects.

Figure 1 shows the actual distributions of shares of girls across workgroups by year. Eight hundred students (exactly 400 per cohort) were randomly assigned to 22 workgroups (11 per year). The shares of girls ranged from 0.14 to 0.51. Deviations from the intended ratios occurred because some students who signed up for the study never showed up, because of slight deviations in the actual shares of girls from the expected share of one third, and because of the lower share of girls with an advanced math preparation.

Compliance to workgroup assignment is high: according to the attendance lists of the first course of the 2008/9 cohort, all students attended their assigned group. Of the students in this cohort who had not dropped out by February, only four percent had formally changed group by the second semester. We do know that in some courses informal group changes occurred, but these were infrequent. Dropouts (who are disproportionately often males) change the gender composition slightly. The correlation between the share of girls in the group that started the year and the share of girls in the group that still studies, however, never drops below 0.90.

4 Results

We regress individual students' study outcomes on the share of girls in their workgroup, controlling for their own sex and a dummy indicating advanced math groups. To gain precision, we control for cohort, application order, age at the start of the academic year, specialization in secondary school, GPA in secondary school and math grade and level in secondary school. GPA, math grades and age are set to zero when missing and dummies are included for missing data. To accommodate the clustered nature of the data, we calculate cluster-robust standard errors. As the number of clusters is fairly small, we test using critical values drawn from a t -distribution where degrees of freedom equals the number of groups minus two (Cameron et al., 2008).⁹

4.1 Dropouts

The academic year is divided into four periods. The bottom part of Table 3 shows that 9.4% of the students dropped out after the exams of period 1. If a student deregisters from the study before February, the tuition fee is reimbursed; 15.1% of students decide to drop out before this date, which coincides with the end of the second period. By the end of period 3, the dropout rate has risen to 25.4%. Like most university faculties in The Netherlands, the University of Amsterdam's economics and business department does not select students before enrolling (as long as the student completed pre-university education, or obtained an equivalent qualification). Instead, selection takes place mainly throughout the first year of the study, by setting the strict requirement that students have to obtain at least 35 credits (out of 60) and pass the Mathematics 1 exam. This explains

⁹In most analyses, we have 22 groups and thus test at critical values of 2.85, 2.09 and 1.73 for the 1%, 5% and 10% significance level, respectively. To ensure that this method is sufficiently conservative, we compare the standard errors with Bias Reduced Linearization (BRL) standard errors (Bell and McCaffrey, 2002). Appendix Table A1 shows some of our main results next to those obtained using BRL. BRL standard errors are somewhat smaller, but we test at higher critical t -values, so that both methods yield equivalent results.

the high dropout rates: 42.4% of students do not start in the second academic year. These dropout rates after each period are significantly higher for boys than for girls.¹⁰

The top part of Table 3 shows the effects of the share of girls in workgroups on the probability of dropping out. The first row shows results from regressions in which only own sex and group type are included as controls. The regressions presented in the second row are based on a specification with the full set of controls. The lower panel shows results from similar regressions, in which the share of girls is now also interacted with a dummy which is equal to one if the student is a girl.¹¹ By February of the first study year, the share of dropouts is substantially lower in workgroups with a higher proportion girls. Increasing the share of girls by 10 percentage points decreases the dropout rate by 1.8 percentage point. This effect seems mainly to be driven by males, although the interaction effect is estimated quite imprecisely. The effect on dropping out disappears towards the end of the year. It seems that a higher share of girls leads (male) students to postpone their dropout decision, but does not prevent them from dropping out in the end.

We conducted the same analysis using data of the cohorts 2004, 2005 and 2006 when the share of girls in workgroups was not manipulated. Results for end of year dropout as dependent variable show a significantly positive coefficient in the specification with controls but without interaction: 0.209 (s.e. 0.107)¹². Using non-experimental data does in this case thus lead to erroneous inferences. Presumably this is because the share of girls is now endogenous, as both boys and less motivated students tend to sign up for the study later, thus ending up in groups with a lower share of girls.

4.2 *Absenteeism*

Workgroup attendance is not centrally registered. We collected absenteeism data for the first practical in 2008/9. In this particular course (and only here), workgroups were randomly split into two halves, in which the original share of females was kept as constant as possible. Absenteeism data were available for 13 of the 22 subgroups. These 13 subgroups come from 10 workgroups. The academic skills practical consisted of seven sessions, of which students were allowed to miss at most two. As the final column in Table 3 shows, increasing the share of girls is associated with a sizable reduction in absenteeism. This impact tends to be larger for girls than for boys, but the difference is not significant.

¹⁰The respective *p*-values are 0.014, 0.007, 0.000 and 0.000.

¹¹In this specification, the coefficient of “share girls” gives the impact on boys while the sum of the coefficients of “share girls” and “student is girl × share girls” is the impact on girls.

¹²*p*=0.06 when testing using a t-distribution with 28 degrees of freedom (we have 30 groups).

Table 3. Effect estimates of share of girls on dropping out and absenteeism

	Dropout after period 1	Dropout after period 2	Dropout after period 3	Dropout at end of year	Absenteeism
	(1)	(2)	(3)	(4)	(5)
Share girls	-0.093 (0.084)	<i>Only sex and group type controls</i> -0.141* (0.079)	-0.117 (0.102)	-0.034 (0.150)	-1.19 (0.66)
Share girls	-0.095 (0.082)	<i>Full set of controls</i> -0.182** (0.077)	-0.132 (0.091)	-0.013 (0.112)	-1.00** (0.40)
Student is girl \times share girls	0.093 (0.194)	<i>Only sex and group type controls</i> 0.152 (0.235)	0.273 (0.287)	0.138 (0.326)	-1.74 (1.51)
Share girls	-0.121 (0.127)	-0.187 (0.131)	-0.199 (0.171)	-0.075 (0.198)	-0.64 (0.98)
Student is girl \times share girls	0.086 (0.195)	<i>Full set of controls</i> 0.163 (0.217)	0.252 (0.281)	0.019 (0.320)	-1.97 (1.69)
Share girls	-0.121 (0.123)	-0.231* (0.124)	-0.208 (0.163)	-0.019 (0.163)	-0.39 (0.81)
N	800	800	800	800	243
Mean	0.094	0.151	0.254	0.424	0.74
Mean - girls	0.056	0.100	0.157	0.273	0.84
Mean - boys	0.111	0.174	0.298	0.492	0.70

Note: Coefficients refer to an increase of the female share from 0% to 100%. Critical values for testing are 2.85, 2.09 and 1.73 for the 1 (**), 5 (***) and 10 (*) percent significance level, respectively (for absenteeism: 3.36, 2.31 and 1.86). Dropout at end of the year is measured by whether the student still took exams after the end of the first academic year. For both the 2007 and 2008 cohort, exam taking has been registered till the end of November 2009, which means three months into the 2009 academic year.

4.3 Credits

As measure of students' performance, we look at the number of credits that students collect during various stages of the first year.¹³ During the first study year, students can obtain a maximum of 60 credits. On average students obtained slightly more than half of this. Columns (1) to (4) of Table 4 report effects on credits obtained per period. The four periods differ slightly in length and correspondingly in the number of attainable credits. Table 4 has the same structure as Table 3.

Early in the year (after period 1), a higher share of girls in workgroups has a positive impact on the number credits. This effect is the same for boys and girls. This effect disappears, however, already in the second period: for the second, third and fourth period, we find no significant impact of the share of girls on the number of credits students obtain. Also for the total number of credits at the end of the first year, we see no impact (see column (5)). For comparison, we again exploited the non-experimental variation in the share of girls in workgroups from cohorts 2004-2006 (results not reported). Although not significantly different from zero, the results tend to suggest that a larger share of girls in a workgroup harms achievement. Again, results from non-experimental data point in a different direction than the experimental results.

Note that this finding of no effect is not a consequence of smaller sample sizes and hence lower power in comparison with studies like Hoxby (2000) and Lavy and Schlosser (2011). The upper bound of the 90% confidence interval of a 10 percentage point increase of the share of girls on the total number of credits equals 0.062 of a standard deviation (based on the specification with controls and without interactions).¹⁴ An effect of 0.08 of a standard deviation that Hoxby reports for math is thus outside the 90% confidence interval of our estimates.

In columns (6) and (7), courses are split up into courses with a high math content (mathematics and statistics courses) and all other courses which contain fewer math components. We find that boys (but not girls) obtain a lower number of math-related credits when the share of girls in their workgroup is higher. This effect is not present for non-math-related courses. Perhaps, as the share of girls increases, instruction in math-related courses becomes geared more to girls. If so, this does not seem to benefit girls, but only to harm boys.

Our estimates on the numbers of credits may have been pushed upwards by students

¹³An alternative measure of students' performance is their GPA in the first year of their study. This measure is difficult to interpret though, since it is based only on exams that students actually took. Many students skip certain exams, and dropouts do not take exams that fell later in the year. This distorts GPA in a direction that depends on the difficulty of the exams that were not taken. How this would exactly affect the estimates is hard to gauge. We estimated regressions similar to those reported in Table 4 with GPA as dependent variable; none of the estimates for share of girls differs significantly from zero.

¹⁴ $(0.412 + 1.645 \times 0.490) / 19.6 = 0.062$

Table 4. Effect estimates of share of girls on number of credits per period and type of course

	Period 1 (1)	Period 2 (2)	Period 3 (3)	Period 4 (4)	Total (5)	Math (6)	Non-math (7)	Total (8)	Math (9)	Non-math (10)
Share girls	2.82* (1.61)	-0.19 (2.19)	0.61 (1.49)	0.24 (1.86)	3.49 (6.14)	-0.81 (1.60)	4.30 (4.97)	0.47 (4.71)	-1.86 (1.81)	2.33 (3.40)
	<i>Full set of controls</i>									
Share girls	2.81** (1.31)	0.69 (1.70)	0.08 (1.05)	0.53 (1.49)	4.12 (4.90)	-1.46 (1.03)	5.58 (4.03)	0.82 (3.46)	-2.61*** (0.91)	3.43 (2.87)
	<i>Only sex and group type controls</i>									
Student is girl × share girls	-1.30 (3.61)	-1.14 (4.69)	-0.21 (2.38)	2.55 (3.53)	-0.10 (12.80)	0.45 (3.16)	-0.55 (10.24)	-2.39 (7.69)	1.95 (2.55)	-4.34 (6.02)
Share girls	3.21 (2.12)	0.15 (2.92)	0.68 (1.59)	-0.52 (2.48)	3.52 (8.10)	-0.95 (1.63)	4.47 (6.78)	1.35 (6.56)	-2.58 (2.18)	3.93 (4.79)
	<i>Full set of controls</i>									
Student is girl × share girls	-0.33 (3.60)	-0.47 (4.34)	0.43 (2.05)	3.94 (3.13)	3.57 (11.76)	1.81 (2.92)	1.75 (9.38)	2.11 (5.74)	3.77** (1.67)	-1.66 (5.38)
Share girls	2.91 (1.73)	0.83 (2.30)	-0.05 (1.16)	-0.65 (1.85)	3.05 (6.42)	-2.01* (1.00)	5.06 (5.48)	0.04 (5.00)	-4.01*** (1.18)	4.04 (4.18)
N	800	800	800	800	800	800	800	461	461	461
Sample	all	all	all	all	all	all	all	ndo	ndo	ndo
Maximum / Scale	13	17	12.5	17.5	60	14	46	60	14	46
Mean	9.2	11.1	6.3	7.0	33.6	5.3	28.3	47.2	8.6	38.6
SD	4.5	6.9	4.4	6.6	19.6	5.1	15.4	10.2	4.1	7.4
Mean - girls	10.2	13.2	7.9	9.6	40.9	7.8	33.2	49.7	10.0	39.7
Mean - boys	8.8	10.1	5.6	5.8	30.3	4.3	26.1	45.6	7.7	38.0

Note: Coefficients refer to an increase of the female share from 0% to 100%. Critical values for testing are 2.85, 2.09 and 1.73 for the 1 (**), 5 (***) and 10 (*) percent significance level, respectively. “ndo” indicates that the sample is restricted to non-dropouts.

who postponed their dropout decision by a few months due to a higher share of girls, and in the meantime managed to obtain a few more credit points. Given that the estimated effect on dropping out after a year is practically zero, one may be more interested in the peer effect on those people who continued studying, as dropouts would have dropped out anyway, irrespective of the share of girls in their workgroup, and effects on the number of credits they obtained before dropping out seem less relevant. Columns (8)-(10) are based on the restricted sample of students who are still taking exams in the next year. Results remain very similar to the results based on all observations, but the effects for math-related courses become more pronounced: boys obtained fewer math credits if they were assigned to workgroups where the share of girls was higher, but for girls, the share of girls made no difference.

4.4 *Heterogeneous effects by ability*

The absence of an impact for the full sample may conceal effects for specific subgroups of students. Particularly, effects may be heterogeneous by ability.¹⁵ We therefore allow gender peer effects to vary with students' own ability levels (measured before treatment). Table 5 reports regressions results. In the upper panel, interactions are added between the share of girls and a dummy indicating whether the student is in the top-25% highest ability. The ability measure is based on GPA in secondary education.¹⁶ We find that effects are neither consistently stronger, nor weaker on the most able students than on other students. For the bottom panel of Table 5, students are split into two halves based on their ability. The interaction effects again do not point to any systematic differences in gender peer effects between students from different parts of the ability distribution. We conclude that the gender peer effect is not heterogeneous across students' own ability level.

4.5 *Nonlinear effects*

Lavy and Schlosser (2011) find that the gender peer effect is largest when the share of girls exceeds 58.7% and Hoxby (2000) finds the largest peer effects in cohorts where the share of girls is above 66%. Such high shares are unusual for university level economics studies. But Lavy and Schlosser also find significant peer effects when the share of girls is below 50% and Hoxby finds effects at lower shares of girls as well, until the share of girls drops

¹⁵Leuven et al. (2010) for example find that an overall zero effect of financial incentives on student performance, is the result of a positive effect for high ability students and a negative effect for low ability students.

¹⁶If this GPA was unknown (virtually always because the student did not take pre-university education, but entered the study via another route, such as via a lower-level bachelor), s/he was placed into the lowest half of the ability distribution.

Table 5. Heterogeneous effect estimates by ability

	Dropout end of year (1)	Absen- teeism (2)	Credits (3)	Credits Math (4)	Credits Other courses (5)
	<i>Full set of controls</i>				
Share girls × 25% highest ability	-0.18 (0.26)	1.61 (1.47)	-2.41 (9.61)	-3.09 (2.11)	0.68 (8.09)
Share girls	0.02 (0.12)	-1.29** (0.46)	4.70 (4.57)	-0.77 (1.06)	5.47 (3.69)
	<i>Full set of controls</i>				
Share girls × 50% highest ability	0.08 (0.25)	0.32 (0.90)	-0.36 (10.59)	-0.73 (1.97)	0.36 (9.02)
Share girls	-0.05 (0.18)	-1.13*** (0.33)	4.31 (6.37)	-1.11 (1.51)	5.42 (5.09)
N	800	242	800	800	800
Scale	0-1	0-1	60	0-14	0-46
Mean	0.424	0.74	33.6	5.35	28.3
SD	0.494	1.12	19.6	5.10	15.4
Mean - girls	0.273	0.83	40.9	7.75	33.2
Mean - boys	0.492	0.70	30.3	4.26	26.1
Mean - 25% highest ability	0.24	0.65	42.1	7.56	34.6
Mean - 75% lowest ability	0.48	0.76	30.9	4.62	26.2
Mean - 50% highest ability	0.33	0.67	37.8	6.23	31.7
Mean - 50% lowest ability	0.52	0.78	29.5	4.49	25.0

Note: Coefficients refer to an increase of the female share from 0% to 100%. Critical values for testing are 2.85, 2.09 and 1.73 for the 1 (***) , 5 (**) and 10 (*) percent significance level, respectively (for absenteeism: 3.36, 2.31 and 1.86). The upper (lower) panel reports regressions that include the full set of controls from our previous analyses, but additionally add the interaction between share of girls in tutorial group and a dummy indicating whether students belonged to the highest 24.8 (50.6) percent in the university preparatory education GPA-ranking.

Table 6. Nonlinear peer effects

	Dropout end of year (1)	Absen- teeism (2)	Credits (3)	Credits Math (4)	Credits Other courses (5)
<i>Only sex and group type controls</i>					
Share girls < 33%	0.027 (0.290)	-0.73 (1.66)	2.41 (13.07)	0.37 (3.46)	2.04 (10.07)
Share girls ≥ 33%	-0.111 (0.356)	-1.73 (1.29)	4.85 (13.57)	-2.31 (3.47)	7.15 (10.89)
<i>Full set of controls</i>					
Share girls < 33%	0.072 (0.309)	-0.53 (1.37)	5.95 (15.02)	-0.97 (2.97)	6.93 (12.30)
Share girls ≥ 33%	-0.116 (0.373)	-1.52 (1.28)	1.91 (15.46)	-2.06 (3.10)	3.96 (12.49)
N	800	243	800	800	800

Note: Coefficients refer to an increase of the female share from 0% to 100%. Regression includes a spline for share of girls with a break at 33%. Critical values for testing are 2.85, 2.09 and 1.73 for the 1 (***) , 5 (**) and 10 (*) percent significance level, respectively (for absenteeism: 3.36, 2.31 and 1.86). The upper (lower) panel reports regressions that include the full set of controls from our previous analyses.

below 33%. We test for non-linear effects using the same cut-off as Hoxby, by estimating regressions in which we include the share of girls as a spline with a break at 33%. Table 6 shows that there are neither peer effects when the share of girls is below 33%, nor when the share of girls is higher. We also find no significant differences between the coefficients for the high- and low-share situations. We also tested for non-linear effects by adding the square of the share of girls to the regression equation. We again find no evidence for non-linearities: none of the quadratic effects gets close to significance.

4.6 Exploring the black box of gender peer effects

At the end of both academic years, we carried out a survey among the students in our experiment. The purpose was to gain further insight into how gender peer effects may arise as a result of peer group related changes in students' behavior. In total, 307 students completed the questionnaire; response rates were 36 percent in 2007/8 and 41 in 2008/9. Due to the timing of the survey, dropouts are underrepresented in the response: 6 percent of the respondents dropped out after period 3, compared to 25 percent of all students. Note that given the high dropout rates, response rates among those students who were still active were much higher than the rates reported above. 39 percent of the respondents are female, which is somewhat higher than the share of girls in the population of 31 percent, but response does not depend on the share of girls in a workgroup: a regression of a response indicator on the share of girls in a workgroup returns a coefficient of -0.11 (s.e.

0.15). Due to the non-response, we have to interpret the results from the questionnaire with some caution.¹⁷

Traditionally, explanations for peer effects have mainly been sought in classroom disruptions, adaptations of teaching styles to the students in the class, social comparisons, and excellent students providing good examples for others (Lazear, 2001; Hoxby and Weingarth, 2005; Van Ewijk and Slegers, 2010). Lavy and Schlosser (2011), for example, find that the presence of more boys is disruptive, leads to a deteriorated learning environment, and induces teachers to adapt their teaching styles. Most research on peer effects, however, and especially research on gender peer effects, focused on school children. In-class behavior in university is likely to be different, and boys may not be as disruptive in workgroups in university as they are in class at younger ages. The different interaction between boys and girls at this age may give rise to different sorts of gender peer effects.

Columns (1) to (5) of Table 7 show effects on the ratings students gave of their peers and workgroup. Columns (6) to (11) refer to ratings students gave of their own behavior. Regression equations are the same as in previous models, with the addition of a control for parental education.¹⁸ Most questions are answered on 5-point scales, with answer categories ranging from not/never (1) to very often (5). Atmosphere in the workgroup is measured on a 10-point scale.

We find that students are neither distracted less often, nor do they pay more attention or rate the atmosphere as better as the share of girls increases. Students in groups with more girls also do not report their peers to pay less attention (column (3)). We conclude that the traditional explanation for gender peer effects of disruptions by boys does not apply to university students.

Other potential channels for gender peer effects are competitiveness and helpfulness of students. Students do not report themselves to be more inclined to help others when the share of girls in their workgroup increases, but do report an increase in average helpfulness among their peers. This is probably not a peer effect, but simply a consequence of girls on average being more helpful than boys. Students neither report changes in their own, nor in their peers' competitiveness, when the share of girls changes (columns (4) and (7)).

More than in primary and secondary school, studying together outside of class can be an aspect of studying at university. About half of all students do sometimes study

¹⁷We also ran the peer effects analyses described above separately for the group of students who completed the questionnaire at the end of the year. The results are similar to those for the students who were still taking exams in the next year.

¹⁸Information on parental education was collected in the questionnaire. Leaving out the parental education control does not alter results.

together with fellow economics students. This share decreases for boys when the share of girls increases, even though in each workgroup, there were enough other males they could study with (if they would only want to study with another boy): the share of girls is at most 0.51. We also find that as the share of girls increases, boys (but again not girls) talk more in class about non-study related topics, and more often refrain from asking questions in class because they are afraid to look dumb. Perhaps the boys want to make a good impression on the girls, and talk more with the girls, or more likely (since the girls themselves do not report talking more often in class), talk about the girls.

5 Summary and discussion

In this paper we examine gender peer effects by means of an experiment in which the share of girls in workgroups attended by first year university students in economics and business was manipulated and students were randomly assigned to these groups. We find little evidence of gender peer effects of a relevant size.

Some students postpone their dropout decision when the share of girls among their peers is higher. But this does not lead them to more study success, as at the end of the year, no effect on dropout ratios remains. Early in the year, we find that absenteeism is reduced for students in workgroups with many girls. It is also at the beginning of the year that gender peer effects induce students to pass more courses. This effect disappears during the year. The effect on the total number of credit points is insignificant and small, especially when taking into account that our original estimate on the number of credits obtained was pushed upward by students who postponed their dropout decision and managed to pass some additional courses before finally dropping out.

We find one exception for our general conclusion that gender peer effects are at most small in size. Boys, but not girls, perform poorer in courses with a high math component if the share of girls in their workgroup increases. Past researchers have tried to separate gender peer effects from ability peer effects (Hoxby, 2000). Girls in our sample outperform boys in math, but at the same time give rise to peer effects that suppress boys' performance. As the ability peer effect of having more able peers is unlikely to be negative, we can interpret this as a true gender peer effect on math performance.¹⁹

Our results stand in contrast to some of the results reported in previous studies. Hoxby (2000) and Lavy and Schlosser (2011) find that as the share of boys in classes increases, school performance declines. The effect sizes they find using very large samples are, however, modest in size. A key difference between this previous research and ours is

¹⁹In the "invidious comparison" model, outcomes are harmed by the presence of higher achieving peers (cf. Hoxby and Weingarth, 2005). According to Sacerdote (2010) this model is certainly possible from a theoretical point of view but may be less important from an empirical perspective.

Table 7. Responses to the questionnaire

	Rating of peers					Own behavior					
	Atmosphere ^a	Distraction ^b	Paying attention ^b	Competitive ^b	Helping each other ^b	Paying attention ^a	Competitive ^b	Helping others ^b	Studying together ^a	Talking during class ^b	Afraid to ask questions ^b
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Share girls	-0.69 (0.80)	0.07 (0.57)	0.14 (0.53)	-0.66 (0.54)	1.16* (0.56)	-0.56 (0.40)	0.18 (0.52)	0.51 (0.55)	-0.60* (0.34)	0.88* (0.45)	1.09** (0.46)
Student is girl × share girls	0.52 (1.15)	-0.75 (1.25)	-1.32 (1.17)	0.23 (1.11)	1.49** (0.61)	0.53 (0.58)	1.79* (0.80)	1.09 (0.71)	0.89** (0.36)	-0.88 (0.70)	-2.41 (1.48)
Share girls	-0.91 (0.97)	0.38 (0.86)	0.69 (0.79)	-0.76 (0.75)	0.54 (0.73)	-0.78 (0.48)	-0.55 (0.64)	0.06 (0.65)	-0.98** (0.35)	1.25** (0.44)	2.09* (0.81)
N	296	149	150	147	148	299	147	150	294	150	148
Scale	1-10	1-5	1-5	1-5	1-5	1-5	1-5	1-5	0-1	1-5	1-5
Mean	6.99	2.82	2.95	2.38	3.27	3.69	2.22	3.07	0.51	3.45	2.30
SD	1.31	0.76	0.79	0.79	0.75	0.75	1.08	0.73	0.50	0.88	0.94
Mean - girls	6.98	2.84	2.98	2.29	3.31	3.69	1.98	3.12	0.53	3.36	2.11
Mean - boys	6.99	2.80	2.93	2.44	3.25	3.69	2.37	3.04	0.50	3.51	2.59

Note: Coefficients refer to an increase of the female share in tutorial groups from 0% to 100%. Regressions control for the full set of covariates described earlier. Additionally, the highest of paternal and maternal education, and a dummy for missing parental education are included as covariates. ^a Question was in both 2007 and 2008 questionnaires; critical values for testing are 2.85, 2.09 and 1.73 for the 1 (***), 5 (***) and 10 (*) percent significance level, respectively. ^b Question was in 2008 questionnaire only; critical values for testing are 3.25, 2.26 and 1.83 for the 1, 5 and 10 percent significance level, respectively.

that we focus on university students, whereas their main focus is on younger children. Interactions between boys and girls are different for different age groups.

Using a survey, we took a closer look at what happens in class as the share of girls changes. Unlike at younger ages, among university students, the presence of more boys does not work disruptively in a traditional sense, and does not lead students to pay less attention. Some peer effects, however, may result from girls being more helpful. But perhaps more importantly, boys change some behaviors when more girls are present. They talk more in class about non-study related topics, and more often refrain from asking questions because they fear it might make them look dumb. The specifics of the interactions between boys and girls at this age may also explain the postponement of dropout decisions that we find: perhaps some boys just stay on longer because of the girls, but if they are weak, they drop out in the end anyway.

Hoxby (2000) and Lavy and Schlosser (2011) in their samples with younger students found that gender peer effects are largest when the proportion of girls exceeds about two thirds. In typical economics curricula, even given the current rise in participation of girls in higher education, the share of girls does not reach this level. Hoxby and Lavy and Schlosser, however, also find gender peer effects for lower shares, notably for situations in which the share of girls is lower than 50%. We find no effects of increasing the female ratio within this range. We argue that this difference is caused by differences in how boys and girls interact at various ages.

Since girls on average have a higher ability (as measured by their GPA in secondary school) and better study performance (as measured by their credits) than boys, the rising share of girls amongst first year students in the economics and business program at the University of Amsterdam and elsewhere has boosted the average quality of the student inflow. Our results suggest that – contrary to what one would expect on the basis of findings of gender peer effects in primary and secondary education – this inflow of (more able) female students does not spill over to the performance of other students.

References

- Bell, R. and McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28:169–179.
- Black, S. E., Devereux, P. J., and Salvanes, K. G. (2010). Under pressure? The effect of peers on outcomes of young adults. Working Paper.
- Cameron, C., Gelbach, J., and Miller, D. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90:414–427.

- Eisenkopf, G., Hessami, Z., Fischbacher, U., and Ursprung, H. (2010). Cognitive skills and single-sex schooling. Unpublished working paper.
- Epple, D. and Romano, R. (2010). Peer effects in education: A survey of the theory and evidence. In *Handbook of Social Economics*. Elsevier Science Publishers B.V.
- Hansen, Z., Owan, H., and Pan, J. (2006). The impact of group diversity on performance and knowledge spillovers – An experiment in a college classroom. NBER Working Paper.
- Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper.
- Hoxby, C. and Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Unpublished.
- Lavy, V. and Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3:1–33.
- Lazear, E. P. (2001). Educational production. *Quarterly Journal of Economics*, 116:777–803.
- Leuven, E., Oosterbeek, H., and van der Klaauw, B. (2010). The effect of financial rewards on students achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, 8:1243–1265.
- Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, 116:681–704.
- Sacerdote, B. (2010). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of Economics of Education 3*. Elsevier Science Publishers B.V.
- Van Ewijk, R. and Slegers, P. (2010). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement*, 21(3):237–265.
- Whitmore, D. (2005). Resource and peer impacts on girls academic achievement: Evidence from a randomized experiment. *American Economic Review*, 95:199–203.

Appendix

Table A1. Dealing with clustering in limited numbers of groups

	Dropout period 1 (1)	Dropout February (2)	Dropout period 3 (3)	Dropout end of year (4)	Absen- teeism (5)	Credits year total (6)	Credits Math courses (7)	Credits Other courses (8)
Share of girls	-0.095 (0.082)	-0.182** (0.077)	-0.132 (0.091)	-0.013 (0.112)	-1.00** (0.40)	4.12 (4.90)	-1.46 (1.03)	5.58 (4.03)
Share of girls	-0.095 (0.088)	-0.182** (0.083)	-0.132 (0.101)	-0.013 (0.124)	-1.00** (0.47)	4.12 (5.50)	-1.46 (1.14)	5.58 (4.52)
N	800	800	800	800	242	800	800	800

Note: The upper panel shows the same results as reported in the main text. Critical values for testing are drawn from a t-distribution with [number of groups minus two] degrees of freedom. Critical values for testing are 2.85, 2.09 and 1.73 (and only for absenteeism 3.36, 2.31 and 1.86) for the 1 (***) , 5 (**) and 10 (*) percent significance level, respectively. The bottom panel shows the same results, in which standard errors have been calculated using bias reduced linearization (Bell and McCaffrey, 2002). Corresponding critical values for testing are 2.58, 1.96 and 1.65.