TIER

Gender peer effects in university: Evidence from a randomized experiment

H. Oosterbeek and R. van Ewijk

TIER WORKING PAPER SERIES
TIER WP 10/24

Universiteit van Amsterdam | Maastricht University Leading in Learning! | rijksuniversiteit groningen

Gender peer effects in university: Evidence from a randomized experiment[1]

Hessel Oosterbeek[2] Reyn van Ewijk[3]

[1]This version: July 2010. We gratefully acknowledge valuable comments from seminar participants in Amsterdam.

[2]University of Amsterdam, Tinbergen Institute and TIER. h.oosterbeek@uva.nl

[3]VU University Amsterdam, 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 their dropout decision when surrounded by more girls, and there is also a modest reduction in early absenteeism. On the other hand, boys perform worse on courses with high math content when assigned to a group with many girls. Overall, however, we fail to find substantial gender peer effects on achievement. This in spite of the fact that students in groups with many girls help each other more often and study together more often.

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, 2011; Sacerdote, 2011). Using data from 3-6 graders in Texas, Hoxby (2000) finds that a 10 percentage point rise in the share of girls leads to an increase of around 4 percent of a standard deviation in reading scores, and an increase of up to 8 percent of a standard deviation in math scores.[1] Given the magnitude of the impact on math scores, Hoxby argues that peers' achievement is unlikely to be the only channel. Using data from Israeli primary, middle and high schools, Lavy and Schlosser (2007) also find large effects from the share of girls in the classroom, and they also argue that these effects are reflecting more than only an increase in peers' average test scores.[2] This paper also provides evidence indicating that the share of girls works through lowering classroom disruption and violence and through improving inter-student and student-teacher relationships. Finally, Whitmore (2005) finds that pupils and 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.[3]

These previous studies deal with gender peer effects in primary and secondary schools. At these education levels, the boys-girls ratio is fixed and virtually constant over time. Only the existence of single-sex schools changes the expected boys-girls ratio at the level of schools.[4] 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 considerably. This naturally leads to the question whether the increased participation of girls in higher education has had an impact of achievement. Extrapolating the findings for primary and secondary school pupils, one would expect that this development raised academic performance. Such extrapolation ignores, however, that the interaction between boys and girls as well as their relative achievement is very different at different

[1]To identify the causal impact of peer composition, Hoxby exploits variation in the share of girls in a grade in a school across adjacent cohorts.

[2]Lavy and Schlosser apply the same approach as Hoxby did.

[3]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.

[4]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.

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. The first year consists of fourteen compulsory courses. Students attend tutorial sessions for all these 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 the relevant peer group. For two consecutive years (2007/8 and 2008/9), we got 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 workgroup. 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, we find that students' perceptions of behavior in terms of helping each other and studying together are positively influenced by the share of girls. This does, however, not affect their outcomes.

The contribution of this paper is threefold. This is the first paper that looks at gender peer effects in university education.[5] This is also the first study reporting about an experiment that explicitly randomizes students into different gender peer groups. And, related to this, there is much more and wider variation in the share of girls in the data used in this study than in the data used in other studies. The ratio in the share of girls between the high and the low treatments is 3 ($= \frac{1/2}{1/6}$). Such variation in treatment intensity will not occur very frequently in studies that use within grade-school variation in the share of girls in adjacent cohorts. And if so, it is much more likely to occur in small schools.

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

[5]Hansen 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.

2 Context, experimental design and data

2.1 Context

The bachelor program in economics and business at the University of Amsterdam has a nominal duration of 3 years. In the first academic year, which runs from September until August, all students in economics and business follow exactly the same program. The first year program is divided into four terms of 7 weeks each in the years that the experiment was conducted. It is important to note that, since the program is fixed, students cannot substitute easy for difficult courses. Every 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. Only after the first term of their second academic year that students choose different packages of courses to specialize either in economics or in business. Students meeting the admission requirements are automatically accepted for the study without further selection.[6] 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/assignments. The instructor of a workgroup is in most cases a member of the permanent (junior) 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. There were 14 workgroups in 2007/8 and also 14 in 2008/9.

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 the teaching hours – slightly over 60% – takes place in workgroup meetings. We therefore assume that the other students assigned to the same workgroup are a relevant peer group.[7]

2.2 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,

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

[7] Defining 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.

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	1	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

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 receive 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 were 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 kept the average motivation/ability of boys and girls assigned to different groups more or less 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 extra separate groups where the share of girls was not manipulated. These groups are removed from all our analyses.

Figure 1 shows the actual distributions of shares of girls across workgroups by year.

Figure 1. Tutorial groups & their share of females

Note: Each dot represents one tutorial group.

Eight hundred students (exactly 400 per cohort) were randomly assigned to 22 workgroups (11 per year).[8] The share 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 disproportionally 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.

If we want to attribute any effects we may find of the share of girls in workgroups to gender peer effects, it is essential that our manipulation only affected the share of girls, and not the distribution of other important characteristics across groups. We take a set of student characteristics that are known to be correlated with study performance or that in other ways might induce peer effects themselves. We regress these students' characteristics at the individual level on the share of girls in their group, controlling for the students' own sex and a dummy for the group consisting of students with an advanced

[8]These numbers exclude students that applied very late and who were assigned to the groups for which the share of girls was not manipulated.

math background.[9]

As Table 2 shows, the share of girls in workgroups does not systematically co-vary with application date (as proxy for students' motivation), nor with performance or track choice in pre-university education, nor with age. We conclude that the experimental manipulation of the share of girls is not contaminated by covariates, so that any effects we may find of the share of girls in workgroups can be attributed to gender peer effects and not to other peer group characteristics.

2.3 Data

Our main data come from the student administration of the department of economics and business of the University of Amsterdam. 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.

Girls tend to apply a bit earlier than boys.[10] 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.

3 Results

We regress individual students' study outcomes on the share of girls in their workgroup, controlling for their own sex and a dummy indicating Math B 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

[9]Keeping own sex constant implies that we examine whether boys (or girls) are randomly assigned across groups with different shares of girls. Not conditioning on own sex biases the estimate of share of girls because that will then also partially reflect the effect of being a girl since there are more girls in groups with high shares of girls (and more boys in groups with low shares of girls). Not including own sex gives an estimate of share of girls on math grade in secondary school of 0.68 (s.e. 0.26) and on GPA in secondary school of 0.39 (s.e. 0.18).

[10]Since students who ended up in the extra workgroups that were created towards the end are excluded from the analysis, the mean of the percentile rank is lower than 50.

Table 2. Balancing

	Application date (pctile)	Math grade secondary school	No Math secondary school	GPA in secondary school	Age	Secondary school track	
						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	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.

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).[11]

3.1 Dropouts

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 students' own sex.

The academic year is divided into four periods. 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. About 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 requirements that students have to obtain at least 35 credits (out of 60) and pass the Mathematics 1 exam. This explains 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.[12]

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 being driven by males, although the interaction effect is estimated quite imprecisely. The effect on dropping out disappears toward 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 after all.

[11] In 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% percent 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.

[12] The respective *p*-values are 0.014, 0.007, 0.000 and 0.000.

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

	Dropout after period 1 (1)	Dropout after period 2 (2)	Dropout after period 3 (3)	Dropout at end of year (4)
	Only sex and group type controls			
Share girls	-0.093	-0.141*	-0.117	-0.034
	(0.084)	(0.079)	(0.102)	(0.150)
	Full set of controls			
Share girls	-0.095	-0.182**	-0.132	-0.013
	(0.082)	(0.077)	(0.091)	(0.112)
	Only sex and group type controls			
Student is girl × share girls	0.093	0.152	0.273	0.138
	(0.194)	(0.235)	(0.287)	(0.326)
Share girls	-0.121	-0.187	-0.199	-0.075
	(0.127)	(0.131)	(0.171)	(0.198)
	Full set of controls			
Student is girl × share girls	0.086	0.163	0.252	0.019
	(0.195)	(0.217)	(0.281)	(0.320)
Share girls	-0.121	-0.231*	-0.208	-0.019
	(0.123)	(0.124)	(0.163)	(0.163)
N	800	800	800	800
Mean	0.094	0.151	0.254	0.424
Mean - girls	0.056	0.100	0.157	0.273
Mean - boys	0.111	0.174	0.298	0.492

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. 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.

Table 4. Effect estimates of share of girls on absenteeism

	(1)	(2)	(3)	(4)
Share girls	-1.19	-1.00**	-0.64	-0.39
	(0.66)	(0.40)	(0.98)	(0.81)
Student is girl × share girls			-1.74	-1.97
			(1.51)	(1.69)
Full set of controls	no	yes	no	yes

Note: Coefficients refer to an increase of the female share from 0% to 100%. Critical values for testing (drawn from a t-distribution with 8 degrees of freedom) are 3.36, 2.31 and 1.86 for the 1 (***), 5(**) and 10 (*) percent significance level, respectively. The numbers of observations equals 243.

3.2 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 groups. This academic skills practical consisted of seven sessions, of which students were allowed to miss at most two. The average number of sessions missed was 0.74 (0.84 for girls and 0.78 for boys). As Table 4 shows, increasing the share of girls is associated with a sizable reduction in absenteeism.

3.3 Credits

As measures 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. Column (1) to (4) report effects on credits obtained per period. The four periods differ slightly in length and correspondingly in the number of attainable credits.

Early in the year (after period 1), a higher share of girls in workgroups is associated with more study success in terms of credits. This effect is the same for boys and girls. This effect disappears, however, already in the second period. By the end of the second period, a substantial share of the students have dropped out, and hence obtain zero credits in each remaining period. Potentially, effects on study performance at this point in time may be

[13] An 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 affects the estimates is hard to gauge. We estimated regressions similar to those reported in Table 5 with GPA as dependent variable; none of the estimates for share of girls differs significantly from zero.

masked by the previously reported lower dropout rate in groups with more girls. This may lead weaker students to stay on more frequently and hence suppresses the average number of credits obtained per student in these groups. By the end of the year, however, the effect on dropping out has disappeared, while effects on the number of credits students obtain are still absent: the initial positive effect on study performance dies out as the year proceeds. The effect on the total number of credits students obtain in their first year (column (5)) remains insignificant.

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 who postponed their dropout decision by a few months as a result of a higher share of girls, and in the mean time 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. In Table 6, we examine whether the students who are still taking exams in the next year have been affected by their peer group composition. If our assumption of no effect on dropouts has been violated, a spurious relation between a higher share of girls and poorer performance among the non-dropouts may exist. This would diminish the validity of the estimates presented here. Table 3 shows that this is unlikely to pose a problem.

Our previous finding of no effects on the total number of credits by the end of the first year is confirmed when we single out those students who continued into their second year. But our point estimate is now smaller, which may be due to postponed dropout decisions having pushed the coefficient in Table 5 upward. The estimates for non-math courses hardly change when we look at this sub-sample. The effects for math-related courses, however, become much 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. This negative effect on boys' math performance is so strong, that it is unlikely to be caused by any unaccounted for effects on dropping out.

Table 5. Effect estimates of share of girls on study performance

	Credits period 1 (1)	Credits period 2 (2)	Credits period 3 (3)	Credits period 4 (4)	Credits total (5)	Credits Math courses (6)	Credits Non-math courses (7)
			Only sex and group type controls				
Share girls	2.82*	-0.19	0.61	0.24	3.49	-0.81	4.30
	(1.61)	(2.19)	(1.49)	(1.86)	(6.14)	(1.60)	(4.97)
			Full set of controls				
Share girls	2.81**	0.69	0.08	0.53	4.12	-1.46	5.58
	(1.31)	(1.70)	(1.05)	(1.49)	(4.90)	(1.03)	(4.03)
			Only sex and group type controls				
Student is girl× share girls	-1.30	-1.14	-0.21	2.55	-0.10	0.45	-0.55
	(3.61)	(4.69)	(2.38)	(3.53)	(12.80)	(3.16)	(10.24)
Share girls	3.21	0.15	0.68	-0.52	3.52	-0.95	4.47
	(2.12)	(2.92)	(1.59)	(2.48)	(8.10)	(1.63)	(6.78)
			Full set of controls				
Student is girl× share girls	-0.33	-0.47	0.43	3.94	3.57	1.81	1.75
	(3.60)	(4.34)	(2.05)	(3.13)	(11.76)	(2.92)	(9.38)
Share girls	2.91	0.83	-0.05	-0.65	3.05	-2.01*	5.06
	(1.73)	(2.30)	(1.16)	(1.85)	(6.42)	(1.00)	(5.48)
N	800	800	800	800	800	800	800
Maximum / Scale	13	17	12.5	17.5	60	14	46
Mean	9.2	11.1	6.3	7.0	33.6	5.3	28.3
(SD)	(4.5)	(6.9)	(4.4)	(6.6)	(19.6)	(5.1)	(15.4)
Mean - girls	10.2	13.2	7.9	9.6	40.9	7.8	33.2
Mean - boys	8.8	10.1	5.6	5.8	30.3	4.3	26.1

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.

Table 6. Effect estimates for students who did not drop out

	Credits	Credits Math courses	Credits Non-math courses
	(1)	(2)	(3)
	Only sex and group type controls		
Share girls	0.47	-1.86	2.33
	(4.71)	(1.81)	(3.40)
	Full set of controls		
Share girls	0.82	-2.61***	3.43
	(3.46)	(0.91)	(2.87)
	Only sex and group type controls		
Student is girl× share girls	-2.39	1.95	-4.34
	(7.69)	(2.55)	(6.02)
Share girls	1.35	-2.58	3.93
	(6.56)	(2.18)	(4.79)
	Full set of controls		
Student is girl × share girls	2.11	3.77**	-1.66
	(5.74)	(1.67)	(5.38)
Share girls	0.04	-4.01***	4.04
	(5.00)	(1.18)	(4.18)
N	461	461	461
Maximum	60	14	46
Mean	47.2	8.6	38.6
(SD)	(10.2)	(4.1)	(7.4)
Mean - girls	49.7	10.0	39.7
Mean - boys	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. Students who did not drop out, are those who were still 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.

3.4 Heterogeneous effects by ability

Peer effects need not be homogeneous. Above we already considered the possibility of heterogeneous effects by allowing for the possibility that the share of girls in a workgroup affects boys and girls differently. We therefore interacted the share of girls with own sex. Another characteristic that is often included in interaction with peer composition is own ability.

Table 7 reports regressions that test whether gender peer effects vary with own ability level. 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.[14] 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 7, 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.

We also tested for non-linear effects by adding the square of the share of girls to the regression equation. We found no evidence for non-linearities: none of the quadratic effects gets close to significance.

3.5 Exploring the causes 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 filled out 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% of the respondents dropped out after period 3, compared to 25% of all students. 39% of the respondents are female, which is somewhat higher than the share of girls in the population of 31%. Due to the non-response, we have to interpret the results from the questionnaire with some caution.

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 Sleegers, 2010). Lavy and Schlosser (2007), for example, find that the presence of more boys is disruptive, leads to a deteriorated learning

[14] If 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 7. Heterogeneous effect estimates by ability

	Dropout end of year (1)	Absen-teeism (2)	Credits (3)	Credits Math (4)	Credits Other courses (5)
	Full set of controls				
Share girls * 25% highest ability	-0.18	1.61	-2.41	-3.09	0.68
	(0.26)	(1.47)	(9.61)	(2.11)	(8.09)
Share girls	0.02	-1.29**	4.70	-0.77	5.47
	(0.12)	(0.46)	(4.57)	(1.06)	(3.69)
	Full set of controls				
Share girls * 50% highest ability	0.08	0.32	-0.36	-0.73	0.36
	(0.25)	(0.90)	(10.59)	(1.97)	(9.02)
Share girls	-0.05	-1.13***	4.31	-1.11	5.42
	(0.18)	(0.33)	(6.37)	(1.51)	(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	28.29
(SD)	0.494	1.12	(19.6)	(5.10)	(15.43)
Mean - girls	0.273	0.83	40.9	7.75	33.19
Mean - boys	0.492	0.70	30.3	4.26	26.08

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. 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.

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 lead to different sorts of gender peer effects.

Columns (1) to (5) of Table 8 show effects on the ratings students gave of their peers and workgroup. Columns (6) to (11) refer to ratings of students' own behavior. Regression equations are the same as in previous models, with the addition of a control for parental education. 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 a 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 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.

Table 8. Responses to the questionnaire

	Rating of peers					Own behavior					
	Atmosphere[a]	Distracting[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	0.07	0.14	-0.66	1.16*	-0.56	0.18	0.51	-0.60*	0.88*	1.09**
	(0.80)	(0.57)	(0.53)	(0.54)	(0.56)	(0.40)	(0.52)	(0.55)	(0.34)	(0.45)	(0.46)
Student is girl × share girls	0.52	-0.75	-1.32	0.23	1.49**	0.53	1.79*	1.09	0.89**	-0.88	-2.41
	(1.15)	(1.25)	(1.17)	(1.11)	(0.61)	(0.58)	(0.80)	(0.71)	(0.36)	(0.70)	(1.48)
Share girls	-0.91	0.38	0.69	-0.76	0.54	-0.78	-0.55	0.06	-0.98**	1.25**	2.09*
	(0.97)	(0.86)	(0.79)	(0.75)	(0.73)	(0.48)	(0.64)	(0.65)	(0.35)	(0.44)	(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.

4 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 in this male-dominated study (69% of the students in our sample are male) 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.[15]

Our results stand in contrast to some of the results reported in previous studies. Hoxby (2000) and Lavy and Schlosser (2007) find that as the share of boys in classes increased, school performance declined. The effect sizes they find using very large samples are, however, modest in size. A key difference between this previous research and ours is 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 there are more girls present. They talk more in class about non-study related topics, and more often refrain from asking

[15] In the "invidious comparison" model outcomes are harmed by the presence of higher achieving peers (cf. Hoxby and Weingarth, 2005). According to Sacerdote (2011) this model is certainly possible from a theoretical point of view but may be less important from an empirical perspective.

questions because they fear it might make them look dumb. We cannot conclude whether this is what affects their performance in math-related courses. But if so, we would have been likely to find similar effects on other types of courses, which we did not. It therefore seems more probable that the finding of negative effects on math performance for boys and no effects for girls, is caused by something else. Boys often do better at mathematics than girls. Despite the fact that girls actually outperform boys in our sample, perhaps the (male) teachers of the workgroups dumbed down their teaching when there were more girls in their class, or they changed their teaching style in such a way that they expected it to be better suited for girls. The result, however, is that girls do not benefit, while boys' performances suffer. The specifics of the interactions between boys and girls at this age may also explain the postponement of dropout decisions that we found: perhaps some boys just stay on longer because of the girls, but if they are weak, they drop out in the end anyway.

Since girls have on average 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 probably 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 better students did 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.

Cameron, C., Gelbach, J., and Miller, D. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90:414–427.

Epple, D. and Romano, R. (2011). 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. Technical report, NBER.

Hoxby, C. M. (2000). Peer effects in the classroom: Learning from gender and race variation. Technical report, NBER.

Hoxby, C. M. and Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Technical report.

Lavy, V. and Schlosser, A. (2007). Mechanisms and impacts of gender peer effects at school. Technical report, NBER.

Lazear, E. P. (2001). Educational production. *Quarterly Journal of Economics*, 116:777–803.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, 116:681–704.

Sacerdote, B. (2011). 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*. Elsevier Science Publishers B.V.

Van Ewijk, R. and Sleegers, P. (2010). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement*, forthcoming.

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 number 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)
				Cluster & use t-2 distribution				
Share of girls	-0.095	-0.182**	-0.132	-0.013	-1.00**	4.12	-1.46	5.58
	(0.082)	(0.077)	(0.091)	(0.112)	(0.40)	(4.90)	(1.03)	(4.03)
				Bias Reduced Linearization SE's				
Share of girls	-0.095	-0.182**	-0.132	-0.013	-1.00**	4.12	-1.46	5.58
	(0.088)	(0.083)	(0.101)	(0.124)	(0.47)	(5.50)	(1.14)	(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.

TIER WORKING PAPER SERIES
TIER WP 10/24

ISBN 978-94-003-0023-1