Economics of Education Review 38 (2014) 51–63

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Gender peer effects in university: Evidence from a randomized experiment☆

Hessel Oosterbeek[a], Reyn van Ewijk[b,c,d,*]

[a] University of Amsterdam and the Institute for Evidence-Based Education Research TIER, The Netherlands
[b] University Medical Centre Mainz, Germany
[c] University of Mainz, Germany
[d] VU University Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 2 November 2012
Received in revised form 31 October 2013
Accepted 10 November 2013

JEL classification:
I23
J16
D83

Keywords:
Field experiment
Peer effects
Gender
University students

ABSTRACT

Recent studies for primary and secondary education find positive effects of the share of females in the classroom on achievement of males and females. This study examines whether these results can be extrapolated to higher education. We conduct an experiment in which the shares of females in workgroups for first year students in economics and business are manipulated and students are randomly assigned to these groups. Males tend to postpone, but not abandon, their dropout decision when surrounded by more females and perform worse on courses with high math content. There is also a modest reduction in absenteeism early in the year. Overall, however, we find no 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 females.

1. Introduction

This paper reports about a randomized experiment designed to analyze gender peer effects among university students. The experiment was conducted among first year undergraduate students in economics and business at the University of Amsterdam. These students are placed in workgroups that have a fixed composition for the entire first year and more than 60% of all teaching hours take place in these groups. Per cohort there are 14 of such groups with an average size of 39 students. In the academic years 2007/8 and 2008/9, we were granted permission to manipulate the gender composition of the workgroups, and to randomly assign incoming students to groups. While the share of females in the total population is around 0.3, we let the share of females in workgroups vary between 0.14 and 0.51. The manipulation ensures that we have the necessary large variation in gender composition across groups. The random assignment of students to groups – conditional on their gender – ensures that students who are of the same gender but are assigned to different groups, are comparable. We examine the impact of the share of females in a group on student performance.

Previous studies dealing with gender peer effects in education either looked at effects of female shares within coeducational classes or compared single-sex with coeducational education. The former type of studies are most comparable to our work. These previous studies are almost exclusively based on data from primary and secondary schools and tend to agree that more females in the

☆ We gratefully acknowledge valuable comments from Eric Bettinger, Adam Booij and Victor Lavy and from seminar participants in Amsterdam, Mannheim and Zürich.

* Corresponding author at: University Medical Centre Mainz, Germany. Tel.: +49 6131 17 3252.
E-mail addresses: h.oosterbeek@uva.nl (H. Oosterbeek), vanewijk@uni-mainz.de (R. van Ewijk).

0272-7757/$ – see front matter
http://dx.doi.org/10.1016/j.econedurev.2013.11.002

classroom improve test scores (Epple & Romano, 2011; Sacerdote, 2010).[1] This naturally leads to the question whether these findings can be extrapolated to higher education where the interactions between students of opposite sexes may be very different. The answer to this question is important for various reasons. First, there is substantial variation in the male–female ratio across fields of study and there has been a considerable increase in the overall share of female students in the past decades. Knowledge about the pattern of gender peer effects may help explain differences in student performance across fields and changes in performance over time. Second, positive effects of the share of females on student performance potentially have implications for education policy. As possible policy implications, Lavy and Schlosser (2011) mention: (1) resource allocation – classes/groups with many males may need to be compensated; (2) teacher assignment – weaker teachers can be assigned to classes/groups with more females; and (3) the placement of low-achievers. To the extent that the findings for primary and secondary education carry over to higher education, these policy implications may also be relevant for higher education institutes.

Concerning effects of single-sex education, Halpern et al. (2011) in their review argue that "there is no well-designed research showing that single-sex education improves students' academic performance". Two recent studies form exceptions. Using that an upper-secondary school in Switzerland forms females-only classes to avoid that the few males in this school are spread out too thinly, Eisenkopf, Hessami, Fischbacher, and Ursprung (2011) find a positive effect of single-sex education on math grades of females. Booth, Cardona-Sosa, and Nolen (2013) conduct an experiment in university in which students are randomly assigned to single-sex or coeducational classes. They find that single-sex education (for one out of twelve teaching hours per week) benefits females. While we consider single-sex education an interesting topic, we have designed our study to address the effects of less extreme forms of gender variation. Commensurate with the difference in student populations between schools and university-level economics studies, our manipulation concentrates on situations in which females generally form a minority, and our share of females never exceeds 51%.[2]

The research in primary and secondary school settings often attributes gender peer effects to males being more disruptive in class. Increased female shares then lead to less disruption, better learning environments, better teacher-student relationships and better performance (Lavy & Schlosser, 2011). Alternatively, higher female fractions might increase performance through ability peer effects if females on average outperform males (Hoxby, 2000). Although the latter still applies in the university setting, the former might play less of a role as university students are older.

Additionally, psychological factors may contribute to gender peer effects. According to stereotype threat theory (Steele & Aronson, 1995), confronting people with the stereotype that people with their background characteristics perform poorly, may make them start acting in a way that confirms this stereotype. If a stereotype that females perform poor in math is more likely to be activated when the female share is low, then females', but not males' performance will expectedly rise in the female share.[3] Teachers might also adapt their teaching style with the female share (Van Ewijk & Sleegers, 2010). Stereotypes about females performing poorly in math and males performing poorer in other topics (on average, females outperform males in our faculty) might cause teachers to "dumb down" teaching levels if more students who are expected to perform poorly attend a class. Furthermore, among university students trying to appeal to members of the opposite sex might be another cause of peer effects. In our experiment, females are never in a clear majority, so we would mainly expect the behavior of males to change as the share of potentially interesting members of the opposite sex increases. This might affect males' behavior in several dimensions, such as paying less attention in class, being more disruptive, refraining from asking questions out of fear of looking dumb, which might lead to poorer performance. On the other hand, if a more balanced male–female ratio improves the classroom atmosphere or if females on average are more inclined to help their peers if they e.g. get stuck with exercises, the effect would be in the opposite direction. In short, most theories would lead us to expect a higher female share to

[1] Using variation in the share of females in a grade in a school across adjacent cohorts, Hoxby (2000) finds for 3rd to 6th graders in the US that a 10%-point rise in the share of females increases reading and math scores by 4% and 8% of a SD, respectively. Applying a similar approach to data from Israeli primary, middle and high schools, Lavy and Schlosser (2011) also find positive effects of more females. They provide evidence that this operates through less disruption and violence and improved inter-student and student-teacher relationships. Again using the same method, Black, Devereux, and Salvanes (2013) find that a higher proportion of girls in 9th grade in Norway is generally good for the longer run outcomes of girls but bad for boys. Whitmore (2005) finds that pupils in kindergarten and 2nd grade (but not in 1st and 3rd grade) do better if there is a majority of females in the class. Graham, Imbens, and Ridder (2009) argue that non-linear effects should be a main point of focus of peer effects studies, as they enable maximization of average achievement. Lavy and Schlosser, and Hoxby find larger peer effects when shares of females are high, but also demonstrate gender peer effects when shares of females are lower.

[2] There are some other related studies. Hansen, Owan, and Pan (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. Carrell, Page, and West (2010) show that having a female professor increases women's performance in math and sciences in an undergraduate institution, but does not affect men's performance. Han and Li (2009) study effects of roommates' ability and show that this affects females', but not males' outcomes. Similarly, Ost (2010) finds that ability peer effects in university more strongly affect females than males. Ficano (2012) finds that higher academic abilities among male (but not among female) peers increases males' GPAs, and finds no such effects on females. Although not a focus point of her study, her regressions also adjust for proportion of males in the course section. She finds no effects of gender peer group over and above the effects of peer ability.

[3] It is not clear-cut, however, that this stereotype threat is weaker with more females in the group. In their review on single-sex education Halpern et al. (2011) discuss evidence that sex segregation increases gender stereotyping.

boost performance, but a few arguments for effects in the opposite direction can be thought of as well.

By and large our results show very little evidence of gender peer effects on achievement of any relevant size. There appears to be a tendency among male students to postpone their dropout decisions when they are surrounded by many females in their workgroups. There is also a modest reduction in absenteeism early in the year. On the other hand, males in workgroups with a high share of females do worse in courses with a high math content (mathematics and statistics). We find no evidence for heterogeneous peer effects by student ability or for non-linear effects within the relevant range for economics curricula.[4] We also analyzed non-experimental data from earlier cohorts and find a positive association between the share of females on achievement. The experimental results show that this association is due to selection bias (with late applicants who happen to be less motivated male students, being placed together) and cannot be given a causal interpretation.

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 females. Students in groups with many females find their groupmates more helpful, but male students are also more afraid to ask questions, and talk more often during class. This does, however, not affect their outcomes. The fact that the results from studies for primary and secondary education cannot be extrapolated to the higher education setting in our study also means that the policy implications that previous studies have related to their findings, do not apply to our setting.

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. 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.[5] 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. If a student passes a course, (s)he will obtain the respective number of credits, irrespective of the grade that was attained. Students failing a course receive no credits.

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 the highest (pre-university track) in secondary 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 the 60) and pass the Mathematics 1 exam. Over 40% of students do not start in the second academic year. Grades play no role in this and grade point averages in general only play a minor role at this stage of the study: master programs do not admit students based on GPAs but solely on courses that were passed and there are no internships later during the study where acceptance is influenced by GPAs. Hence, credits are the most important measure of student achievement.[6]

The total number of first year students in the economics and business program is 503 in 2007/8 and 505 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. Instructors change with the subject so that e.g. no macroeconomist would tutor a microeconomics course. Instructors are assigned to workgroups by the faculty by persons who (like the tutors themselves) were unaware of our experiment. There is one instructor per workgroup, but some tutors might teach more than one workgroup. 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 thirteen 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 slightly more than 60% of the

[4] Lavy and Schlosser (2011) and Hoxby (2000) found the largest effects when the share of females exceeded about two-thirds. Typical economics curricula are unlikely to reach such a high proportion of females.

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

[6] There might be some concern that the number of credit points is a "rough" measure of performance only able to discriminate between very weak students and above-average students, thereby making it hard to find an impact. This is, however, not true. Less than 20% of the first year students manages to collect all first year credits point during the first year. The mean is 34 with a standard deviation of 20. In a study based on previous cohorts Leuven, Oosterbeek, and van der Klaauw (2010) find substantial positive effects of financial incentives on credit points for students with above mean ability. In a study on subsequent cohorts Booij, Leuven, and Oosterbeek (2013) find substantial and statistically significant ability peer effects on the number of credit points. This proves that our outcome measure is "fine" enough to detect effects. It should also be noted that the number of credit points is the relevant outcome measure in the view of the board of the economics department of the University of Amsterdam, as the internal funding scheme of the university is partially based on it (and not on GPA).

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

teaching hours takes place in workgroup meetings, while the other 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 males and females separately.

Students enrolling in the economics and business program of the University of Amsterdam score an average GPA in secondary school equal to 6.68 (on a scale from 1 to 10). This is almost identical to the average GPA of 6.70 on the final exam for all students in the pre-university track. Females outperform males with regards to the GPAs in secondary school. Note that secondary school exams in The Netherlands are standardized, so that grades are comparable between students who attended different schools. 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. Dutch secondary schools allow students to choose between four nation-wide standardized specializations. 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 males and females.

[7] We also collected additional data through a survey among students. We describe (and report about) this data source in Section 4.7.

3. Design

For the experiment we manipulated the shares of females in first year workgroups and assigned males and females randomly to these groups. The 2007/8-cohort consists of 33% females, the 2008/9-cohort of 30%. Beforehand we decided that one third of the workgroups would contain few females (around 1/6), one third of the groups would have an average share of females (around 1/3), and one third of the groups would have many females (around 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 one or two weeks. In doing so, we kept the shares of females as close as possible to the predetermined shares. This implies that males and females 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 males and females assigned to different groups the same.[8]

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 females 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

[8] Students usually apply for a study during their last year in secondary school. Arguably, the more certain they are about their study choice and the more motivated they are, the earlier they apply. Recall that students meeting the admission requirements are unconditionally accepted without further selection. The odd highly motivated student may apply late because of first being rejected by institutions abroad, but this plays only a minor role.

Table 2
Checks on the randomization procedure.

	Application date (pctile)	Math grade secondary school	No math secondary school	GPA in secondary school	Age	Secondary school specialization	
						Economics	Science/Health
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share females	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 males	46.8	6.67	0.13	6.59	19.32	0.69	0.19
Mean females	42.7	7.10	0.22	6.88	19.06	0.62	0.15

Note: Each column shows a regression of individual students' characteristics on share of females, controlling for students' own sex and a dummy for advanced math group. Coefficients refer to how much the respective dependent variable changes when the proportion of females our sample if raised from 0 to 1. Standard errors are between brackets. Significant effects mean that groups with different shares of males and females are systematically different on the respective variable. 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.

the share of females was not manipulated. These groups are removed from all our analyses.[9]

To examine whether females and males that are assigned to groups with different shares of females are indeed not systematically different, we ran a number of regressions in which each time an observable characteristic of individual students was regressed on the share of females in their workgroups. We thereby controlled 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 females 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 and that students assigned to groups with many females do not differ in characteristics from other same-sex students assigned to groups with few females.[10]

However, males and females do in general differ in average background characteristics (e.g. females on average had higher GPAs in high school), so that groups with high and low shares of females differ in more dimensions than only in the share of females. 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 females being different from males 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 females comes together with other peer characteristics, we aggregate our data by group and calculate correlations between shares of females and group averages of the characteristics mentioned in 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 females and group-average secondary school grade equals 0.39 (s.e. 0.18) and the conditional correlation between the share of females 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.

Fig. 1 shows the actual distributions of shares of females across workgroups by year. Eight hundred students (exactly 400 per cohort) were randomly assigned to 22 workgroups (11 per year). The shares of females ranged from 0.14 to 0.51. Deviations from the intended ratios occurred because some students who signed up for the

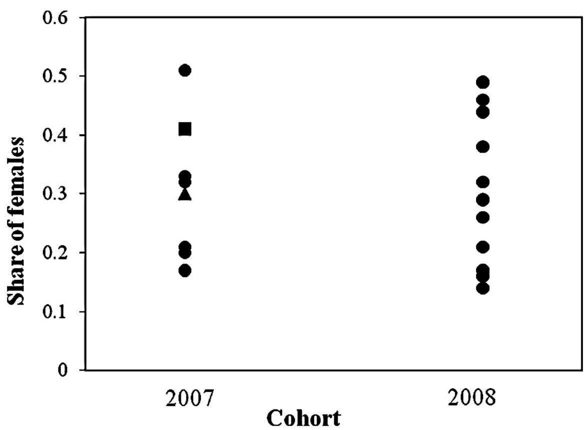

Fig. 1. Tutorial groups and their share of females. *Note*: Each dot represents one tutorial group. A triangle represents two groups and a square represents three groups.

[9] This explains why the numbers of observations in Table 2 falls short of the total number of students enrolled, and why the mean of the percentile rank in Table 2 is lower than 50.

[10] As e.g. Eisenkopf et al. (2011) show, there may be interaction effects of teacher gender with student gender. This would bias our results if male or female teachers are systematically assigned to groups with a high share of females. A similar argument could be made about teacher quality. In our study, instructors change with the subject, while neither instructors nor the course coordinators assigning instructors to groups were aware of the experiment. Although we do not avail of background information on individual teachers, we therefore expect teacher gender and quality to balance out over the working groups.

study never showed up, because of slight deviations in the actual shares of females from the expected share of one third, and because of the lower share of females 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 4% 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 females in the group that started the year and the share of females in the group that still studies, however, at no point in time drops below 0.90.

4. Results

We regress individual students' study outcomes on the share of females 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, Gelbach, & Miller, 2008).[11]

4.1. Dropouts

Due to the strict selection of students at the end of the first academic year (discussed above), dropout rates are high: 42.4% of students do not start in the second academic year. These dropout rates after each period are significantly higher for males than for females.[12] 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%.

The top part of Table 3 shows the effects of the share of females 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 females is now also interacted with a dummy which is equal to one if the student is a female.[13] By February of the first study year, the share of dropouts is substantially lower in workgroups with a higher proportion females. Coefficients refer to an increase in the proportion of females 0–1. As such an increase is only hypothetical, when interpreting results, we will sometimes discuss effects of increases by 10 percentage points, which means that coefficients are divided by a factor 10. For example, increasing the share of females by 10 percentage points decreases the dropout rate by 1.8 percentage point (dropping out is measured as a binary variable). This effect seems mainly to be 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 females 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 females 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).[14] Using non-experimental data does in this case thus lead to erroneous inferences. Presumably this is because the share of females is now endogenous, as both males and less motivated students tend to sign up for the study later, thus ending up in groups with a lower share of females.

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 females is associated with a sizable reduction in absenteeism. This impact tends to be larger for females than for males, but the difference is not significant.

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. During the first study year, students

[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% significance level, respectively. To ensure that this method is sufficiently conservative, we compare the standard errors with Bias Reduced Linearization (BRL) standard errors (Bell & McCaffrey, 2002). Table A2 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.

[13] In this specification, the coefficient of "share females" gives the impact on males while the sum of the coefficients of "share females" and "student is female × share females" is the impact on females.

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

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

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

Note: Each column shows results from four separate regressions. Coefficients refer to changes in the probability of dropping out/absenteeism (both are measured as binary variables) when the proportion of females in a group is increased from 0 to 1. Standard errors are between brackets. Critical values for testing are 2.85, 2.09 and 1.73 for the 1% (***), 5% (**) and 10% (*) 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.

can obtain a maximum of 60 credits. On average students obtained slightly more than half of this. There is no policy of adjusting grades in a certain way to achieve certain targets, nor is there grading on a curve. Columns (1)–(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 females in workgroups has a positive impact on the number credits. This effect is the same for males and females. 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 females 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 females in workgroups from cohorts 2004 to 2006 (results not reported). Although not significantly different from zero, the results tend to suggest that a larger share of females in a workgroup harms achievement. Again, results from non-experimental data (where share of females is probably endogenous) 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 95% confidence interval of a 10 percentage point increase of the share of females on the total number of credits equals 0.07 of a standard deviation (based on the specification with controls and without interactions).[15] 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 males (but not females) obtain a lower number of math-related credits when the share of females in their workgroup is higher. This effect is not present for non-math-related courses. Perhaps, as the share of females increases, instruction in math-related courses becomes geared more to females. If so, this does not seem to benefit females, but only to harm males.

Our estimates on the numbers of credits may have been pushed upwards by students who postponed their dropout decision by a few months due to a higher share of females, 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 females 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

[15] $(0.412 + 1.96 \times 0.490)/19.6 = 0.07$.

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

	Period 1	Period 2	Period 3	Period 4	Total	Math	Non-math	Total	Math	Non-math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Only sex and group type controls										
Share females	2.82*	−0.19	0.61	0.24	3.49	−0.81	4.30	0.47	−1.86	2.33
	(1.61)	(2.19)	(1.49)	(1.86)	(6.14)	(1.60)	(4.97)	(4.71)	(1.81)	(3.40)
Full set of controls										
Share females	2.81**	0.69	0.08	0.53	4.12	−1.46	5.58	0.82	−2.61***	3.43
	(1.31)	(1.70)	(1.05)	(1.49)	(4.90)	(1.03)	(4.03)	(3.46)	(0.91)	(2.87)
Only sex and group type controls										
Student is female× share females	−1.30	−1.14	−0.21	2.55	−0.10	0.45	−0.55	−2.39	1.95	−4.34
	(3.61)	(4.69)	(2.38)	(3.53)	(12.80)	(3.16)	(10.24)	(7.69)	(2.55)	(6.02)
Share females	3.21	0.15	0.68	−0.52	3.52	−0.95	4.47	1.35	−2.58	3.93
	(2.12)	(2.92)	(1.59)	(2.48)	(8.10)	(1.63)	(6.78)	(6.56)	(2.18)	(4.79)
Full set of controls										
Student is female× share females	−0.33	−0.47	0.43	3.94	3.57	1.81	1.75	2.11	3.77**	−1.66
	(3.60)	(4.34)	(2.05)	(3.13)	(11.76)	(2.92)	(9.38)	(5.74)	(1.67)	(5.38)
Share females	2.91	0.83	−0.05	−0.65	3.05	−2.01*	5.06	0.04	−4.01***	4.04
	(1.73)	(2.30)	(1.16)	(1.85)	(6.42)	(1.00)	(5.48)	(5.00)	(1.18)	(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 – females	10.2	13.2	7.9	9.6	40.9	7.8	33.2	49.7	10.0	39.7
Mean – males	8.8	10.1	5.6	5.8	30.3	4.3	26.1	45.6	7.7	38.0

Note: Each column shows results from four separate regressions. Coefficients refer to increases in the proportion of females in a group from 0 to 1. Standard errors are between brackets. Critical values for testing are 2.85, 2.09 and 1.73 for the 1% (***), 5% (**) and 10% (*) significance level, respectively. "ndo" indicates that the sample is restricted to non-dropouts.

courses become more pronounced: males obtained fewer math credits if they were assigned to workgroups where the share of females was higher, but for females, the share of females made no difference.

4.4. GPA

An alternative measure of students' performance is their GPA in the first year of their study. As discussed, GPA is a less relevant measure in this case than number of credits. Furthermore, effects on GPA are difficult to interpret, since GPA 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 and find no effects (Table 5).

4.5. 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.[16] We therefore allow gender peer effects to vary with students' own ability levels (measured before treatment). Table 6 reports regressions results. In the upper panel, interactions are added between the share of females and a dummy indicating whether the student is in the top-25% highest ability. The ability measure is based on GPA in secondary education.[17] 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 6, 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.

4.6. Nonlinear effects

Lavy and Schlosser (2011) find that the gender peer effect is largest when the share of females exceeds 58.7% and Hoxby (2000) finds the largest peer effects in cohorts where the share of females is above 66%. Such high shares are unusual for university level economics studies. But

[16] Leuven 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.

[17] 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 5
Effect estimates of share of females on GPA.

	Period 1 (1)	Period 2 (2)	Period 3 (3)	Period 4 (4)	Total (5)	Math (6)	Non-math (7)
Only sex and group type controls							
Share females	0.39	−0.42	−0.08	−0.18	0.02	−0.63	0.24
	(0.51)	(0.62)	(0.70)	(0.58)	(0.43)	(0.80)	(0.41)
Full set of controls							
Share females	0.25	−0.27	−0.54*	0.11	−0.04	−0.80	0.24
	(0.43)	(0.49)	(0.27)	(0.42)	(0.34)	(0.53)	(0.34)
N	778	721	633	597	790	687	788
Mean	6.11	5.98	5.74	5.42	5.64	687	788
SD	1.24	1.65	1.44	1.54	1.28	5.09	5.83
Mean – females	6.43	6.57	6.04	5.92	6.11	2.08	1.18
Mean – males	5.97	5.70	5.59	5.15	5.43	5.88	6.22

Note: Each column shows results from four separate regressions. Grades run from 1 to 10; students pass a course if they obtain grade 6 or higher. Coefficients refer to increases in the proportion of females in a group from 0 to 1. Standard errors are between brackets. Critical values for testing are 2.85, 2.09 and 1.73 for the 1% (***), 5% (**) and 10% (*) significance level, respectively.

Table 6
Heterogeneous effect estimates by ability.

	Dropout end of year (1)	Absenteeism (2)	Credits (3)	Credits math (4)	Credits other courses (5)
Full set of controls					
Share females × 25% highest ability	−0.18	1.61	−2.41	−3.09	0.68
	(0.26)	(1.47)	(9.61)	(2.11)	(8.09)
Share females	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 females × 50% highest ability	0.08	0.32	−0.36	−0.73	0.36
	(0.25)	(0.90)	(10.59)	(1.97)	(9.02)
Share females	−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	28.3
SD	0.494	1.12	19.6	5.10	15.4
Mean – females	0.273	0.83	40.9	7.75	33.2
Mean – males	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: Each column shows results from two separate regressions. Coefficients refer to increases in the proportion of females in a group from 0 to 1. Standard errors are between brackets.
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% (*) 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 females in tutorial group and a dummy indicating whether students belonged to the highest 24.8% (50.6%) in the university preparatory education GPA-ranking.

Lavy and Schlosser also find significant peer effects when the share of females is below 50% and Hoxby finds effects at lower shares of females as well, until the share of females drops 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 females as a spline with a break at 33%. Table 7 shows that there are neither peer effects when the share of females is below 33%, nor when the share of females 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 females to the regression equation. We again find no evidence for non-linearities: none of the quadratic effects gets close to significance.

4.7. 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% 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

Table 7
Nonlinear peer effects.

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

Note: Each column shows results from two separate regressions. Coefficients refer to increases in the proportion of females in a group from 0 to 1. Standard errors are between brackets. Regression includes a spline for share of females with a break at 33%. Critical values for testing are 2.85, 2.09 and 1.73 for the 1% (***), 5% (**) and 10% (*) 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.

Table 8
Responses to the questionnaire.

	Rating of peers					Own behavior					
	Atmosphere[a]	Distracting[b]	Paying attention[b]	Competitive[b]	Helping each others[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 females	−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 female × share females	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 females	−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 – females	6.98	2.84	2.98	2.29	3.31	3.69	1.98	3.12	0.53	3.36	2.11
Mean – males	6.99	2.80	2.93	2.44	3.25	3.69	2.37	3.04	0.50	3.51	2.59

Note: Each column shows results from two separate regressions. Coefficients refer to increases in the proportion of females in a group from 0 to 1. Standard errors are between brackets. 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% (*) 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% significance level, respectively.

dropped out after period 3, compared to 25% 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% of the respondents are female, which is somewhat higher than the share of females in the population of 31%, but response does not depend on the share of females in a workgroup: a regression of a response indicator on the share of females 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.[18]

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 (Hoxby & Weingarth, 2005; Lazear, 2001; Van Ewijk & Sleegers, 2010). Lavy and Schlosser (2011), for example, find that the presence of more males 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 males may not be as disruptive in workgroups in university as they are in class at younger ages. The different interaction between males and females at this age may give rise to different sorts of gender peer effects.

[18] We 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.

Columns (1)–(5) of Table 8 show effects on the ratings students gave of their peers and workgroups. Columns (6)–(11) refer to ratings students gave of their own behavior. Appendix A.3 contains a translation of the included questions. Regression equations are the same as in previous models, with the addition of a control for parental education.[19] 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 females increases. Students in groups with more females 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 males 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 females 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 females on average being more helpful than males. Students neither report changes in their own, nor in their peers' competitiveness, when the share of females 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 males when the share of females increases, even though in each workgroup, there were enough other males they could study with (if they would only want to study with another male): the share of females is at most 0.51. We also find that as the share of females increases, males (but again not females) 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. Maybe the males want to make a good impression on the females, and talk more with the females, or perhaps more likely (since the females themselves do not report talking more often in class), talk about the females. Not finding an impact of the share of females on females being afraid to ask questions is at odds with the stereotype threat theory which would predict that females feel more comfortable asking questions when surrounded by more other females.

5. Summary and discussion

In this paper we examine gender peer effects by means of an experiment in which the share of females 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 decisions when the share of females 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 females. 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 decisions and managed to pass some additional courses before finally dropping out. As our outcomes are limited to those in the first year of study, we cannot rule out longer-term effects.

We find one exception for our general conclusion that gender peer effects are at most small in size. Males, but not females, perform poorer in courses with a high math component if the share of females in their workgroup increases. Past researchers have tried to separate gender peer effects from ability peer effects (Hoxby, 2000). Females in our sample outperform males in math, but at the same time give rise to peer effects that suppress males' 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.[20]

The experiment was conducted in the setting of first year university students in economics and business at the University of Amsterdam. Several features of the experiment are specific to this setting: (i) around one third of all students were females; (ii) credits are the main performance measure; (iii) the curriculum is fixed and has an economics/business content. To what extent our findings generalize to other settings is an open question. It should also be mentioned that we have been unable to randomize or systematically assess the interaction between the gender composition of workgroups and the gender of the instructor. Results from other studies suggest that this may be an interesting extension.

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 males 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 that we focus on university students, whereas their main focus is on younger children.

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

[20] In the "invidious comparison" model, outcomes are harmed by the presence of higher achieving peers (cf. Hoxby & 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.

Interactions between males and females are different for different age groups. Our failure to find substantial gender peer effects also means that possible implications for education policy such as resource funding, teacher assignment and placement of low achievers, do not carry over to our setting.

Using a survey, we took a closer look at what happens in class as the share of females changes. Unlike at younger ages, among university students, the presence of more males does not work disruptively in a traditional sense, and does not lead students to pay less attention. Some peer effects, however, may result from females being more helpful. But perhaps more importantly, males change some behaviors when more females 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 males and females at this age may also explain the postponement of dropout decisions that we find: perhaps some males just stay on longer because of the females, 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 females exceeds about two thirds. For typical economics curricula such high shares are irrelevant as, even given the current rise in participation of females in higher education, the share of females are unlikely to 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 females 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 males and females interact at various ages. Note also that if we would have based our inferences on the non-experimental data from earlier cohorts, we erroneously would have concluded that also in the range that we study, the share of females has a positive impact on achievement.

Since females on average have a higher ability (as measured by their GPA in secondary school) and better study performance (as measured by their credits) than males, the rising share of females among 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.

Appendices

See Tables A.1 and A.2.

Table A.1
Descriptives per workgroup.

Share of females	Cohort	Number of females/males	Females: Dropout rate	Females: Absenteeism	Females: Total credits	Females: Credits math-related courses	Females: Credits non-math-courses	Males: Dropout rate	Males: Absenteeism	Males: Total credits	Males: Credits math-related courses	Males: Credits non-math-courses
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
14%	2008	5/32	20%	n.a.	49	11.0	38	25%	n.a.	37	6.1	31
16%	2008	6/32	50%	1.7	31	3.2	27	56%	1.2	27	3.3	24
17%	2007	5/25	0%	n.a.	53	10.4	42	52%	n.a.	32	5.0	27
17%	2008	7/34	29%	1.6	37	8.0	29	56%	0.4	31	4.3	27
20%	2007	7/28	0%	n.a.	48	10.0	38	61%	n.a.	22	3.3	19
21%	2007	8/31	50%	n.a.	28	4.8	24	42%	n.a.	31	5.0	26
21%	2008	8/31	38%	0.8	40	7.0	33	68%	0.9	22	2.3	20
26%	2008	10/29	30%	0.9	41	8.0	33	55%	0.9	27	3.1	24
29%	2008	10/24	20%	0.6	44	7.5	36	42%	0.7	33	4.5	29
30%	2007	11/26	9%	n.a.	50	11.2	39	23%	n.a.	44	8.1	36
30%	2007	11/26	27%	n.a.	42	7.7	35	50%	n.a.	29	4.9	24
32%	2007	11/23	36%	n.a.	32	6.9	25	57%	n.a.	29	3.5	25
32%	2008	11/23	55%	1.7	27	3.5	24	70%	0.6	23	3.0	21
33%	2007	12/24	17%	n.a.	45	8.5	36	58%	n.a.	27	3.3	24
38%	2008	12/20	0%	0.1	55	11.7	43	25%	0.1	43	6.7	36
41%	2007	15/22	47%	n.a.	32	6.4	26	41%	n.a.	35	4.7	31
41%	2007	16/23	13%	n.a.	47	9.5	37	57%	n.a.	25	4.6	21
41%	2007	16/23	13%	n.a.	44	9.7	35	39%	n.a.	32	4.4	27
44%	2008	16/20	31%	0.1	40	5.8	34	60%	0.5	28	3.0	25
46%	2008	16/19	44%	0.6	38	5.9	32	37%	0.9	33	3.6	30
49%	2008	17/18	35%	0.9	43	7.7	35	61%	0.4	28	3.4	25
51%	2007	19/18	26%	n.a.	40	7.6	33	44%	n.a.	29	3.5	25
Total		249/551	27%	0.8	41	7.8	33	49%	0.7	30	4.3	26

n.a. = not available.
Note: Each row represents one working group. The table shows the average results for female and for male students in each of the workgroups.

Table A.2
Dealing with clustering in limited numbers of groups.

	Dropout period 1	Dropout February	Dropout period 3	Dropout end of year	Absenteeism	Credits year total	Credits math courses	Credits other courses
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cluster and use $t-2$ distribution								
Share of females	−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 females	−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% (*) significance level, respectively. The bottom panel shows the same results, in which standard errors have been calculated using Bias Reduced Linearization (Bell & McCaffrey, 2002). Corresponding critical values for testing are 2.58, 1.96 and 1.65.

A.3. Questionnaire

- What grade do you give for the atmosphere in your workgroup on a scale of 1 (very bad) to 10 (excellent)?
- How frequently did it occur during the tutorial sessions that you attended, that fellow students distracted you? (Scale 1)
- How much attention did your fellow students on average pay during the tutorial sessions that you attended? (Scale 2)
- To what extent do you agree with the following statement? "In my tutorial group, the atmosphere is competitive." (3)
- How often do your fellow students help each other during the tutorial sessions (with exercises etc.)? (1)
- How much attention did you generally pay during the tutorial sessions that you attended? (2)
- How often do you think that you appear competitive to your fellow students during tutorials? (1)
- How often do you help your fellow students during the tutorial sessions (with exercises etc.)? (1)
- How often (next to attending classes) do you study together with fellow-economics students? (4)
- During tutorials, how often do you discuss topics with your fellow students that are not related to your studies? (1)
- How often does it occur that you do not understand something during a tutorial, yet do not ask a question because you do not want to appear stupid? (1)

Scales:

1. Never/almost never/now and then/often/very often
2. Very much/quite a lot/some/a little/none
3. Completely disagree/disagree/disagree nor agree/agree/completely agree
4. Often/regularly/sometimes/never

References

Bell, R., & 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., & Salvanes, K. G. (2013). Under pressure? The effect of peers on outcomes of young adults. *Journal of Labor Economics, 31*, 119–153.

Booij, A., Leuven, E., & Oosterbeek, H. (2013). Ability peer effects in university: Evidence from a randomized experiment, Mimeo.

Booth, A. L., Cardona-Sosa, L., & Nolen, P. (2013). *Do single-sex classes affect exam scores? An experiment in a coeducational university. Technical report* Centre for Economic Policy Research, Research School of Economics, Australian National University.

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

Carrell, S., Page, M., & West, J. (2010). Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Economics, 125*, 1101–1144.

Eisenkopf, G., Hessami, Z., Fischbacher, U., & Ursprung, H. (2011). *Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland. CESifo Working Paper No. 3592*.

Epple, D., & Romano, R. (2011). Peer effects in education: A survey of the theory and evidence. In J. Benhabib, A. Bisin, & M. Jackson (Eds.), *Handbook of social economics*. Amsterdam: Elsevier Science, North Holland.

Ficano, C. C. (2012). Peer effects in college academic outcomes – Gender matters. *Economics of Education Review, 31*, 1102–1115.

Graham, B., Imbens, G., & Ridder, G. (2009). *Measuring the effects of segregation in the presence of social spillovers: A nonparametric approach. NBER Working Paper No. 14860*.

Halpern, D., Eliot, L., Bigler, R., Fabes, R., Hanish, L., Hyde, J., et al. (2011). The pseudoscience of single-sex schooling. *Science, 333*, 1706–1707.

Han, L., & Li, T. (2009). The gender difference of peer influence in higher education? *Economics of Education Review, 28*(1), 129–134.

Hansen, Z., Owan, H., & Pan, J. (2006). *The impact of group diversity on performance and knowledge spillovers – An experiment in a college classroom. NBER Working Paper No. 12251*.

Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No. 7867*.

Hoxby, C., & Weingarth, G. (2005). *Taking race out of the equation: School reassignment and the structure of peer effects*. (unpublished).

Lavy, V., & 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., & 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.

Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences? *Economics of Education Review, 29*(6), 923–934.

Sacerdote, B. (2010). *Peer effects in education: How might they work, how big are they and how much do we know thus far? Handbook of economics of education* (Vol. 3). Amsterdam: Elsevier Science, North Holland.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.

Van Ewijk, R., & Sleegers, 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.