University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 285

Exposure to More Female Peers Widens the Gender Gap in STEM Participation

Anne Ardila Brenøe and Ulf Zölitz

April 2018

Exposure to More Female Peers Widens the Gender Gap in STEM Participation*

Anne Ardila Brenøe
(University of Copenhagen and IZA)

Ulf Zölitz
(University of Zurich and IZA)

This paper investigates how high school gender composition affects students' participation in STEM college studies. Using Danish administrative data, we exploit idiosyncratic within-school variation in gender composition. We find that having a larger proportion of female peers reduces women's probability of enrolling in and graduating from STEM programs. Men's STEM participation increases with more female peers present. In the long run, women exposed to more female peers earn less because they (1) are less likely to work in STEM occupations, and (2) have more children. Our findings show that the school peer environment has lasting effects on occupational sorting and the gender wage gap.

(JEL I21, J16, J31)
Keywords: gender, peer effects, STEM studies

* Anne Ardila Brenøe, University of Copenhagen and IZA. Øster Farimagsgade 5, 1353 Copenhagen. aab@econ.ku.dk. Ulf Zölitz: University of Zurich (Department of Economics and Jacobs Center for Youth Development), IZA, and Maastricht University. Schönberggasse 1, 8001 Zürich. ulf.zoelitz@econ.uzh.ch.
We are thankful for helpful comments and suggestions from Massimo Anelli, Mette Ejrnæs, Jan Feld, Mette Gørtz, Kjell Salvanes, and Denni Tommasi, as well as seminar participants at the Melbourne Institute of Applied Economic and Social Research, AEFP 2017, DGPE 2016, the University of Copenhagen, and the Copenhagen Education Network.

I. Introduction

Although women today attain more education than men do, large gender differences in the choice of study field persist in most OECD countries. Only 28 percent of students in Science, Technology, Engineering, and Mathematics (STEM) studies are female (OECD, 2016). As gender differences in ability do not seem to explain these differences (Kahn and Ginther 2017), we currently know little about why women remain underrepresented in STEM fields. It is important to gain an improved understanding of the origins of gender differences in study choices due to the potential consequences for both the individual and society. First, women with high math and science ability who do not participate in STEM forfeit higher lifetime earnings. Second, society as a whole may be less innovative and thereby have worse long-run economic growth when fewer women are part of the STEM workforce.

In this paper, we investigate how the gender composition in high school affects men's and women's decisions to choose STEM fields in higher education. High school peers represent a central aspect of teenagers' social environment as they interact on a daily basis for several years. For students facing one of the most crucial life choices, peers may therefore represent an important social force shaping specialization decisions. To investigate whether gender composition in high school affects the gender gap in STEM participation, we use Danish register data on all students entering the math track in high school between 1980 and 1994. The key advantage of this data set, in addition to its rich information on individuals' education and labor market outcomes, is that we can follow the entire student population over a period of 20 years after they entered high school. This allows us to identify the direct, delayed, and long-run consequences of high school peers.

Our strategy to identify the causal impact of gender composition on STEM choice builds on two empirical approaches that differ in their key identifying assumptions and interpretation of results. The first empirical approach exploits idiosyncratic cohort variation in the proportion of female students *within schools* across cohorts after taking out school fixed effects, cohort fixed effects, and school-specific time trends.[1] The second empirical approach extends this model by including cohort-by-school fixed effects, thereby examining whether the proportion of female peers *differentially* affects men and women *within the same school-cohort*. This identification strategy thus answers whether the gender gap in STEM choice changes as the gender composition changes.

[1] This identification strategy is similar to Anelli and Peri (forthcoming), Hill (2017), Hoxby (2000), and Lavy and Schlosser (2011), who exploit idiosyncratic variation in the proportion of female students within schools.

The key identifying assumption for our first strategy is that year-to-year variations in the proportion of female students are exogenous to factors affecting STEM choice, conditional on school fixed effects, cohort fixed effects, and school-specific time trends. To assess the credibility of this identifying assumption, we conduct an extensive set of balancing checks, testing whether changes in gender composition are associated with student characteristics. Using a large set of student background characteristics from the register data, we show that gender composition does not systematically relate to the characteristics of students selecting into the specific school cohort, conditional on school and cohort fixed effects.[2] While this balancing test provides strong support for our key identifying assumption, it remains theoretically possible that students sort into schools based on factors correlated with STEM choice that are both time variant and unobservable in the register data. Although it is difficult to think of mechanisms that would create these unobservable time-variant school selection patterns, our second empirical approach addresses this concern. The inclusion of fixed effects for each cohort-by-school cell alleviates potentially remaining concerns, as we can control for the exact level at which selection based on *time variant* and *unobservable* characteristics would take place.[3] Both of our empirical approaches yield qualitatively similar results.

Our results show that women exposed to a higher proportion of female peers become less likely to enroll in STEM fields and more likely to enter health-related studies in college. Men also behave more gender-stereotypically when more female peers are present: they become more likely to enroll in STEM studies and less likely to enter health-related studies. These peer effects in field of study choice are statistically and economically significant. A 10-percentage point increase in the proportion of female high school peers lowers women's probability of enrolling in STEM studies by 1.4 percentage points—which is equivalent to a 7 percent decrease from the baseline. For men, a similar change in the gender composition raises STEM enrollment by 0.9 percentage points (2.3 percent). These peer effects exacerbate gender differences not only in STEM enrollment but also translate into an increased gender gap in STEM degree completion. In our most conservative

[2] We also test and reject the possibility that the proportion of female students enrolling in a given school is autocorrelated over time. Put differently, we find no evidence that the proportion of female students in year *t-1* predicts the proportion female peers in year *t*.

[3] Importantly, results from regressions including cohort-by-school fixed effects no longer identify whether an increase in the proportion of female peers affects the levels of STEM enrollment for men and women, but instead identify gender differences in the response to changes of the peer environment. To our knowledge, this is the first paper to apply this strategy to identify a gap in group-specific responses.

model, which includes cohort-by-schools fixed effects, we find that 10 percentage points more high school female peers increase the gender gap in STEM degree completion by 2 percentage points, corresponding to a 17 percent increase.

We shed some light on possible mechanisms behind this finding by studying how peer gender affects student performance, measured as high school grade point average (GPA), and how the effects differ across subgroups with different levels of parental education. Since high-achieving students are generally more likely to enter STEM fields and it has been shown that peers affect performance in high school (e.g. Hoxby, 2000; Lavy and Schlosser 2011), one possible mechanism is that the gender composition affects men's and women's preparedness for STEM studies. We find evidence in support of this mechanism: having more female peers alters the gender gap in high school GPA in favor of men, which may give women—who consider their comparative advantage—reason to believe that they are less prepared for STEM college studies. When considering heterogeneous effects by parental background, we provide two pieces of evidence that suggest that information about college, in general, and STEM studies, in particular, can counter peer influences to some degree. First, students with college-educated parents are less affected by peers. Second, and more strikingly, the gender peer composition does not influence women with STEM-educated mothers, i.e. women who have a salient female role model at home.

Our long-run results on labor market trajectories show that the peer effects in study choice lead women and men to systematically different career paths. Not only are women exposed to more female high school peers less likely to choose STEM studies, they are also less likely to work in STEM occupations and they have lower earnings at age 36. A 10-percentage point increase in the proportion of female high school peers lowers women's probability of working in STEM occupations by 6 percent and increases the gender wage gap by 5 percent. These results imply that high school peers and their influence on college major choice have lasting and economically significant consequences for occupational segregation and earnings.

To provide a better understanding of *why* we observe this sizable impact of peers on earnings, we examine fertility as an underlying mechanism. The relationship between high school peers and fertility is important for the interpretation of our results for at least two reasons. First, entering a STEM career may reduce fertility if jobs in these fields are less family-friendly. It might be harder to combine having children with work obligations in these environments. In this case, changes in fertility should be interpreted as an unintended consequence of the high school composition that 'pushed' some individuals in or out of STEM careers. Second, if peers directly affect preferences for children, then the documented effects for the gender wage gap may simply be because having children reduces earnings for women (Kleven et al. 2018; Lundborg et al. 2017). In this case, lower female earnings should not be attributed to staying out of STEM fields per se, but are instead a consequence of the "career cost of children" (Adda et. al., 2017).[4]

Notably, we find that women exposed to more female peers have more children by age 36. We provide suggestive evidence that about half of the effect of female peers on earnings is due to increased fertility and the other half is due to STEM participation. For men, we also find that peers affect fertility, but we find no impact on earnings. This is consistent with Kleven et al. (2018), which documents a large reduction in women's earnings but no change in men's earnings after the arrival of children.

In the existing literature, only a handful of related studies investigate how the gender of peers in high school influences educational choices.[5] Related to our work, Lavy and Schlosser (2011) find that both male and female students take more science courses in high school when exposed to a high school cohort with more female peers. While Lavy and Schlosser (2011) provide intriguing evidence on the underlying mechanism, it is not possible to study long-run effects on study choice, occupational sorting, fertility, or earnings in their setting. In contrast to Lavy and Schlosser (2011), Anelli and Peri (forthcoming) show that the gender composition in high school has an effect on men's, but not on women's, study choice in college. The authors find that men attending high school classes with over 80 percent male peers are more likely to enroll in predominantly male college majors. Contrary to our findings, these effects do not persist into actual degree completion or labor market outcomes. At the college level, Hill (2017) presents suggestive evidence that women exposed to a university cohort with more female peers have a lower probability of majoring in STEM fields. Similarly, Zölitz and Feld (2017) show that women become less likely to major in male-dominated subfields when they are randomly assigned to university sections containing more female peers. On the contrary, Schneeweis and Zweimüller (2012) show that a larger share of female peers in lower secondary vocational school increases girls' propensity to choose male-dominated school types. Based on the existing literature, which presents mixed evidence from a variety of different settings, it is not clear how gender composition

[4] Related to this point, if peers shape individuals' preferences for children—even before those children are born—men and women may choose more family-friendly careers outside STEM that can facilitate their family planning.
[5] Starting from Hoxby (2000) a different related strand of literature investigates how gender composition affects student performance. For important studies on the impact of peer gender on performance, see Whitmore (2005), Lavy and Schlosser (2011), Giorgi, Pellizzari and Woolston (2012), Oosterbeek and Ewijk (2014) as well as Hill (2015).

affects specialization decisions.[6]

This paper contributes to the literature in three important ways. First, we are the first to document that gender composition in high school affects STEM participation in college. Second, we provide previously undocumented comprehensive evidence on the long-run occupational consequences of high school peers. Our ability to follow students in the Danish administrative data over the course of 20 years after high school entry distinguishes this study from existing work that mostly studies the short- or medium-run impact of peers. Third, and more broadly, this paper contributes to a better understanding of the origins of gender differences in educational choices and labor market outcomes. This paper shows that the gender composition of high school peers represents an important aspect of the social environment that shapes individuals' preferences for field of study, occupation, and fertility and thereby influences earnings.

II. Institutional Background and Data

In this study, we use Danish administrative data covering the entire population of first-year high school students enrolled in the math track from 1980 through 1994. The key advantage of our dataset is that it contains rich background information and allows us to follow individuals over the course of 20 years after high school entry. We link the administrative data on high school students to annual data on educational enrollment and degree completion, which also contains detailed information on the type, level, and field of education, as well as labor market outcomes up to 20 years after entering high school.

Throughout this paper, we focus on students within the high school math track—students for whom entering STEM fields in college represents a relevant career option. As admission to most STEM college programs requires specific high school STEM course prerequisites, students from other high school tracks rarely choose STEM studies.[7] Among students within the math track, 30 percent later decide to enroll in a STEM program. These students thus represent the most relevant margin for increasing women's STEM participation. In the rest of this section, we introduce the

[6] The related literature on the impact of single sex education also provides mixed results. While Jackson (2012) finds that single-sex secondary schools cause girls to take fewer math and science courses, other studies find no impact (Sohn, 2016) or a positive effect (Lee et al. (2014) and studies cited therein).
[7] For comparison, only 4.6 percent of STEM college graduates attended the high school language track.

institutional setting, describe the estimation sample, and present summary statistics on the key variables of the analysis.

A. *Institutional Background*

Children in Denmark enter primary school the year they turn 7 years and are required to attend school through grade 9.[8] It is optional to attend grade 10, which is a formal continuation of primary school. In their final year of primary school, students apply for secondary school. When applying for secondary education, students can choose between three-year academic high schools and vocational programs, which typically take four years.[9] The general academic high school, which represents the most popular type of academic high school, has two tracks: math and language.[10] In the high school application process, students specify their first, second, and third choice, each representing a combination of a specific high school and track. Students are qualified for high school admission if they have completed at least nine years of education with satisfactory results and if teachers state that they are qualified.[11] All applicants qualified according to these criteria are guaranteed admission in a high school in their county of residence. If there is insufficient capacity at all three preferred schools, the allocation committee in their home county admits them to another school after considering commuting time. [12] Schools experiencing capacity problems are concentrated in metropolitan areas. After high school completion, many students take one or two gap years before entering college.[13]

In the college application process, students apply for a specific field of study and a specific institution and can indicate up to eight institution-specific study programs. A diploma from an academic high school is required for admission. Admission depends on high school GPA; however, most STEM programs have no or very low GPA cutoffs and almost all eligible students

[8] For the cohorts we study, it was not mandatory to attend a kindergarten class (grade 0), but most children did so.
[9] Academic high schools fall broadly into three branches: general, commercial (HHX), and technical (HTX).
[10] During the period we study, about 18 percent of each birth cohort enroll in the math track and about 45 percent of students within the math track are female. Both the share of math-track students and the share of women within the math track were relatively constant over the period our estimation sample covers.
[11] If these conditions do not hold, students can still qualify for school admission if they pass an entry exam.
[12] According to conversations with school principals active during our observation period, admission committees did not consider the gender of applicants during the admission process.
[13] We define college as professional and academic tertiary education.

who apply are admitted. While GPA does not restrict students' STEM study choice, certain high school courses, such as advanced Mathematics and intermediate Physics and Chemistry, are prerequisites for STEM college majors.

B. *Estimation Sample*

We exclude students with missing values for gender and age (0.8 percent of students). We exclude students who were not between 14 and 19 years old when entering the general high school (less than 0.01 percent). We further restrict the estimation sample to schools in which at least 95 percent of students in a given cohort are 14–19 years old and exclude schools with very small cohort sizes of less than 10 students in a given year (6.1 percent of students). We apply these restrictions to exclude schools that mainly offer evening education or single courses, which target older part-time students who are, in many cases, working at the same time. Finally, we restrict the sample to schools that exist and admit students for at least four consecutive years (excluding 0.6 percent of students). None of these data restrictions qualitatively changes the results.[14]

C. *Summary Statistics*

Table 1 provides an overview of the summary statistics. Our estimation sample consists of 182,211 students attending 127 different schools over a period of 15 years, resulting in a total of 1,877 school-cohort observations. Forty-five percent of the students are female and the average cohort size is 108 students.

[Table 1 here]

Panel A in Table 1 shows the student outcomes we consider in this paper. The primary outcomes of interest are indicators for whether the student enrolls in a STEM study field and whether their college degree is within STEM fields at the college level or higher.[15] To classify

[14] Results are available upon request.
[15] Throughout the paper, we use the field of the highest obtained degree to construct measures of STEM completion.

STEM study programs, we follow the International Standard Classification of Education (ISCED) classification system. STEM degrees are thus studies within the following ISCED fields: Natural Sciences, Mathematics, and Statistics (ISCED-05), Information and Communication Technologies (ISCED-06), and Engineering, Manufacturing, and Construction (ISCED-07). To examine which fields within STEM drive the effects we also split STEM into four subfields: 1) Biology, 2) Math and Physics, 3) ICT and Engineering, and 4) Manufacturing and Construction. Additionally, we consider the probability of completing the highest degree within Health Sciences (ISCED-091), Education (ISCED-01), Arts/Humanities (ISCED-02), Social Sciences (ISCED-031), and Business/Law (ISCED-04).

From the total sample, 79 percent of all students enroll in college after high school. Table 1 shows that only 21 percent of female and 38 percent of male high school students subsequently enroll in STEM studies. This gender gap persists in STEM completion rates: while only 14 percent of women graduate with a STEM degree, 25 percent of men do so. Labor market outcomes show that 20 years after high school entry, 11 percent of women and 26 percent of men work in a STEM occupation.[16]

Panel B in Table 1 provides an overview of the student demographic and parental background characteristics we use in the regression analysis as controls, and Panel C shows school-level variables. The key peer variable of interest is the proportion of female peers at the time of high school entry, which we construct at the cohort-school-track level excluding the individual himself or herself. As less than one percent of students change to another high school or track, this group of peers represents the social group in which students interact over a three-year period. Students are, on average, exposed to 45 percent female peers. A one standard deviation change in the proportion of female peers is equivalent to 7.0 percentage points.

Figure 1 shows the raw correlations between the proportion of female peers and the probability of completing a STEM degree. For women, a higher proportion of female peers is correlated with a lower probability of obtaining a STEM degree. For men, on the contrary, we observe a positive correlation between the proportion of female peers and STEM degree

---

The enrollment variables are indicators for the student's ever having been enrolled in the respective study at the college level or higher. If we instead consider the field of first or last enrollment, we find very similar results throughout.

[16] We use the Danish version of the International Standard Classification of Occupations (DISCO) and construct an indicator for working in STEM if the individual works in a high-skilled occupation within STEM for at least half the years observed, 11–15 and 16—20 years after high school entry, respectively. All results remain qualitatively the same when using indicators for whether the mode occupation is within STEM for the considered periods.

completion. These raw associations suggest a fairly linear relationship. While these correlations are purely descriptive, they foreshadow the results of our regression analysis.

[Figure 1 here]

III. Empirical Strategy

The fundamental threat to identification of peer effects arises from student sorting at various institutional levels. Parents select into neighborhoods, students select into schools, and within schools, students may select into classrooms or be assigned to tracks. As students are typically not assigned to schools at random, the existing peer effects studies try to overcome this identification problem by exploiting natural variation in cohort composition within a given school across time (Bifulco, Fletcher, and Ross 2011; Carrell and Hoekstra, 2010; Carrell, Hoekstra, and Kuka 2016; Hanushek et al. 2003; Hoxby 2000; Hoxby and Weingarth 2005; Lefgren 2004; Vigdor and Nechyba 2006). While this identification strategy addresses the issue of endogenous, time-constant, student sorting into schools, it is vulnerable to school-specific (dynamic) time trends that may alter both the peer composition and the outcome of interest. More recent peer effects studies respond to this concern with the inclusion of school-specific time trends—linear, quadratic, and cubic (Hill, 2017; Lavy and Schlosser 2011; Lavy, Schlosser, and Paserman 2012; Schneeweis and Zweimueller 2012). For identification, these studies exploit the deviation in peer composition from its long-term time trend within a school. This approach has the advantage of controlling for unobserved factors correlated with time trends in school composition that may confound peer effects in schools.

Our first empirical approach is similar to the approach in the literature discussed in the previous paragraph, and shares the key identifying assumption that the variation in the peer composition is exogenous after taking out school fixed effects, cohort fixed effects, and school-specific time trends. In our second empirical approach, we estimate an even more restrictive model by including fixed effects for each cohort-by-school cell, thereby alleviating all potentially remaining concerns regarding selection. In the following, we describe our empirical model, our two identification strategies, and the underlying assumptions in more detail.

A. *Empirical model*

Our main empirical model is:

$$Y_{isc} = \beta_1\, Female_i \times PropFemalePeers_{isc} + \beta_2\, Male_i \times PropFemalePeers_{isc}$$
$$+\beta_3\, Female_i + C_{isc}\gamma' + e_{isc}\,, \tag{1}$$

where Y_{isc} is the outcome of student i attending school s in cohort c. The main outcomes we consider are STEM participation, the individuals' earnings percentile by age and birth cohort, and fertility. In our data, each individual represents one observation. The treatment variable of interest is $PropFemalePeers_{isc}$, which represents the proportion of female peers individual i is exposed to in their school s and cohort c. As the primary objective of this paper is to test whether peer composition affects women and men are differently, we interact the proportion of female peers with the indicator variables $Female_i$ and $Male_i$ that refer to the students' own gender.

β_1 thus captures to which degree women's study choice and labor market outcomes are affected by the peer gender composition in their high school and β_2 captures the equivalent impact for men. β_3 captures the gender gap in outcomes conditional on controls. C_{isc} represents a vector of school and cohort fixed effects as well as individual and peer characteristics which we gradually add when estimating Equation (1). The inclusion of high school fixed effects accounts for time-invariant endogenous sorting into schools and cohort fixed effects control for confounding factors at the national level, affecting all students in a given cohort. In order to account for unobserved time-variant school characteristics correlated both with changes in the proportion of female peers and educational choices for students within the same schools, we add school-specific linear, quadratic, and cubic time trends to the vector C_{isc}.

In our more conservative models, the vector C_{isc} includes the following additional student and peer average characteristics, which do not significantly alter our estimates: six indicator variables for mother's and father's highest educational degree and 18 indicators for their field of education;[17] indicators for first- and second-generation immigrant; a "traditional family" indicator that equals one if the student lives with both parents at age 10; dummies for student age at the time of high school start; mother's age at birth and its squared term; an indicator for having a young

[17] For each parent, we include 9 dummies indicating whether their highest education is within ISCED fields 1–9.

mother (< 22 years at birth); an indicator for whether the child is firstborn; family size and its squared term; and a dummy for whether the individual is adopted.[18] Additionally we control for up to third-degree polynomials of cohort size, as peer composition may potentially be correlated with cohort size (Epple and Romano, 2011).[19] Finally, the vector includes the proportion of female students in the language track in the same high school cohort and controls for the high school curriculum experiment that took place in Denmark in the 1980s (for more details regarding the experiment, see Joensen and Nielsen, 2016). The main purpose of including this large vector of control variables is to test how sensitive our results are to the inclusion of these variables. In the spirit of Altonji, Elder, and Taber (2005), changes in the coefficients of interest that result from including controls may inform us about the degree to which omitted unobservable factors may affect our results. To allow students' outcomes to correlate within their group of peers, we cluster the standard errors, e_{isc}, at the school-cohort level.[20]

The key identifying assumption for our first approach to yield causal estimates of β_1 and β_2 in Equation (1) is that no omitted variable exists that fulfills *all* of the following four requirements:

(1) time-variant and school specific,
(2) not captured by school-specific time trends,
(3) correlated with both the peer composition and the outcome of interest, and
(4) not included in the extensive set of individual- and peer-level control variables observed in the administrative data sets.

[18] As individual controls, we additionally always control for dummies indicating the number of years after high school entry the individual was last observed in the education registry for our education models and dummies for the number of years observed within the given period for labor market outcomes. The inclusion of these controls reduces measurement errors given that the annual data does not record individuals living abroad.

[19] Moreover, we include indicators for whether the school experiences a change of more than 30 students or more than 50 percent in the cohort size compared to the previous cohort and up to two period lags of these variables. Given that a high school degree takes three years, we include these two-year lags to account for the possibility that a student's outcome was affected despite the fact that the inflow did not happen in their cohort. We furthermore control for the number of students from the cohort that are not in the age range of 14–19 years and for the number of students in the cohort that have missing gender and age information. Reasons for missing information on gender or age include that the person does not live in Denmark but attends a Danish school (e.g. lives close to the Danish-German border) or that the person resides in the country on Diplomat visa. None of these included control variables qualitatively changes our results.

[20] Angrist (2014) shows that with chance variation in peer groups, measurement error can bias peer effects estimates. Feld and Zölitz (2017) study this issue in more depth and show that classical measurement error can lead to overestimation of peer effects. Because we observe the students' gender in administrative registries, gender is arguably measured without error and our estimates should thus be free from upward bias arising from measurement error.

While it is difficult to think of any plausible mechanism that would create a violation of this type, the existence of such factors remains possible. To assess the credibility that such factors do not exist, we conduct an extensive set of balancing checks in section IV in which we test whether the peer composition in a school-cohort is systematically related to a large vector of high-quality measures of student background characteristics observable in the register data. While these balancing tests strongly support our key identifying assumption, we also provide results from a second empirical approach. Our second approach addresses the possibility of identification problems arising from unobservable, time-variant, and school-specific omitted variables not captured by school-specific time trends that may be correlated with both the peer composition and the outcome of interest.

In our second empirical approach, we extend Equation (1) by including an additional set of fixed effects for each cohort-by-school cell. The inclusion of these cohort-by-school fixed effects alleviates potential remaining concerns, as we control for the exact level at which selection based on *time variant* and *unobservable* characteristics would take place. Importantly, estimates from this type of model no longer identify whether an increase in the proportion of female peers will affect the overall number of women and men who choose a STEM program, but instead identify a gender difference in the response to changes in the peer environment. These estimates thus answer whether the proportion of female high school peers affects gender gaps in STEM participation and earnings.

IV. Balancing tests

To assess the plausibility of our key identifying assumption that *time variant* and *unobservable* factors are not driving our results, we test whether we observe systematic selection based on a wide range of observable student characteristics. One violation of our key identifying assumption would be, for example, if women select into a specific school based on the expectation of a higher or lower proportion of female peers within that school-cohort. In our first balancing test, we test whether students' own gender is correlated with the proportion of female peers, conditional on cohort and school fixed effects. This test closely follows the randomization check proposed by Guryan, Kroft, and Notowidigdo (2009) and controls for the school-level leave-out mean of the proportion of female peers across cohorts within the school to account for the mechanical relationship between own gender and peer gender. Table 2 shows that the proportion of female high school peers is not systematically related to students' own gender. The point estimate is precisely estimated and not distinguishable from zero. The inclusion of individual and school level controls as well as up to cubic time trends in Columns (2)–(5) does not significantly alter the point estimate.

[Table 2 here]

While Table 2 rejects sorting based on gender, it may still be possible that students sort into schools with a high proportion of female peers based on characteristics other than gender. The availability of a large set of high-quality measures of student background characteristics in the Danish administrative registries—typically not observable in other studies—allows us to rigorously test for this possibility.

In our next balancing test, we determine in how many cases student characteristics are significantly correlated with the proportion of female peers. Table 3 summarizes the significance of the point estimates from a total of 190 separate bivariate regressions, which test whether the proportion of female peers is related to student characteristics conditional on cohort and school fixed effects. Each column presents the results for a different set of control variables and school-specific time trends. Appendix Table A1 shows the full balancing test with all 190 coefficients. As expected when running a large number of regressions testing multiple hypotheses, some coefficients are statistically significant. In the absence of systematic sorting, we would expect 10 percent of coefficients to be statistically significant at the 10 percent significance level, 5 percent at the 5 percent level, and 1 percent at the 1 percent level simply due to chance. The share of significant coefficients is below the respective expectation for all three significance levels. Table 3 shows that 1.1 percent of estimates are significant at the 1 percent level threshold, 2.1 percent are significant at the 5 percent level, and that 8.9 percent are statistically significant at the 10 percent level. These balancing tests suggest that the proportion of female peers is as good as random and provide strong support for our key identification assumption. Without systematic cohort and school-specific sorting on this large set of observables, it appears highly unlikely that *unobservable time variant* factors create unobserved sorting patterns.

[Table 3 here]

Next, we test whether the proportion of female peers enrolling at a given school is autocorrelated over time. We do this by running 127 separate regressions that separately test, for each individual school, whether the proportion of female students at time *t* is correlated with the proportion of female peers that enrolled in *t-1*. While such school-level autocorrelation would not impose a threat to our identification strategy as it would be captured by the included school-specific time trends, the existence of such school-specific time dynamics may point to the existence of other unobservable time-variant confounders. Table 4 provides a summary of this exercise and reports the proportion of schools for which we find significant autocorrelation in the proportion of female students. For all significance levels, the share of schools for which the lag of female peers significantly predicts the proportion of female peers is close to what we would expect in the absence of autocorrelation. Across all models, 0.98 percent of the school-level regressions are significant at the 1 percent level, 3.35 percent are significant at the 5 percent level, and 8.86 percent are significant at the 10 percent level. Thus, we find no evidence that the proportion of female students enrolling in a given school is autocorrelated over time.

[Table 4 here]

As a final randomization check, we inspect whether the variation in the proportion of female peers, which we empirically exploit in this paper, is consistent with variation that we would expect with natural random fluctuations. Figure 2 plots the proportion of female peers at the school level after residualizing on cohort and school fixed effects and school-specific linear time trends. Figure 2 shows that these deviations in the proportion of female peers closely follow the normal distribution, which we plot for comparison. The shape of the distribution further supports the idea that the proportion of female peers is as good as random, conditional on the included controls.

In sum, the extensive set of balancing checks in this section provides strong support for our key identifying assumption. The evidence shown in Table 2, Table 3, Table 4, and Figure 2 suggests that the proportion of female peers is as good as random, conditional on cohort and school fixed effects.

V. Results

A. *Participation in STEM College Education*

Table 5 shows estimates of how the peer composition affects STEM enrollment (Panel A) and STEM degree completion (Panel B).[21] Column (1) shows the most basic model, which includes only the proportion of female peers, student gender, and the interaction between these variables. In Columns (2)–(6), we gradually include additional fixed effects and individual level controls. The specification in Column (6) includes school fixed effects, cohort fixed effects, a large set of peer- and student-level control variables, as well as linear, quadratic, and cubic school-specific time trends. Columns (2)–(6) show that the magnitude of the estimates is not particularly sensitive to the exact set of included fixed effects, controls, or time trends. Column (7) shows estimates from our most restrictive specification, which includes cohort-by-school fixed effects and shows the impact of peers on the gender gap in STEM participation.

[Table 5 here]

Our results show that women exposed to a higher proportion of female peers become less likely to enroll in and graduate from STEM college programs. Men's choices also become more gender-stereotypical: they are more likely to enroll in and complete STEM studies when they have a larger share of female high school peers. In our preferred specification in Column (4), a 10-percentage point increase in the proportion of female high school peers lowers women's probability of enrolling in STEM by 1.4 percentage points, corresponding to a decrease of 6.7 percent relative to the baseline. For men, we find that a similar change in the gender composition raises STEM enrollment by 0.9 percentage points—a 2.4 percent change from the baseline.[22]

Column (7) shows our most restrictive model, which includes cohort-by-schools fixed effects. We find that 10 percentage points more female peers in a high school cohort increase the gender gap in STEM enrollment by 2.3 percentage points—which is equivalent to a 14 percent

[21] We tested whether the gender composition of peers affects the probability of dropping out of high school and the probability of enrolling in or completing college. Table A2 in the Appendix shows no effect.

[22] In addition to the linear-in-shares models shown in Table 5, we have also estimated non-linear peer effects using six bins for the proportion of female peers. In this analysis, we find relatively linear effects over the range of support that we have in the data (Figure 3 and Figure 4).

increase of the gender gap. Because we include cohort-by-school fixed effects in the model, the coefficient in Column (7) identifies a change in the gender gap in STEM completion and not an absolute effect. Importantly, the effect size we identify is close to the difference between the coefficients of male and female students in the less restrictive models in Columns (1)–(6), which increases our confidence in the estimates obtained from the models without cohort-by-school fixed effects.[23] Consequently, our estimate identified from within cohort variation implies that exposure to more female peers within a given school cohort substantially increases gender differences in STEM choice and leads to more gender-stereotypical enrollment choices. These results are consistent with Zölitz and Feld (2017), which find that exposure to more female peers in university teaching sections decreases women's likelihood of choosing male-dominated majors and lowers their labor earnings during the first years after college graduation.

Do high school peers only affect enrollment decisions or do they have lasting effects on study completion as well? The distinction between study enrollment and completion is potentially important as Anelli and Peri (forthcoming) find that gender peer effects in high school affect men's initial study enrollment, but have no impact on study completion or labor market outcomes. We therefore next shed light on the persistence of effects by considering the impact on the probability of STEM graduation.

Panel B in Table 5 shows that the peer effects in the field of enrollment persist into actual degree completion rates. In our setting, peer effects in study enrollment are not offset by changes of college major or college dropout. Women exposed to more female peers in high school are significantly less likely to graduate with a college STEM major. A 10-percentage point increase in the proportion of female peers lowers women's STEM graduation probability by 1.0 percentage points —a 7 percent decrease from the baseline (Column 4). The same change raises men's probability of graduating in a STEM field by 0.9 percentage points, which is equivalent to a 3.6 percent increase from the baseline. Again, the point estimates of interest in Columns (2)–(6) are very similar across models and are insensitive to the exact set of included fixed effects and time trends. Column (7) in Panel B confirms that these results hold when including cohort-by-school fixed effects. Gender differences also remain present in graduation rates when we exploit whether the gender composition among peers differentially affects men and women within the same school-cohort. Column (7) shows that 10 percentage points more female peers increase the gender gap in

[23] To see this, compare the effect size of -0.265 in Column (7) with the estimate of -0.226 (-0.135 -0.091) in Column (4).

STEM completion by 2 percentage points, corresponding to 17 percent.[24]

Given the low baseline rates for women's STEM enrolment and graduation, the size of the peer effects we document in Table 5 are economically significant. Taken together, we find that a higher share of female peers makes both men's and women's initial choice of study field and field of graduation more gender-stereotypical.

To understand which STEM subfields women are less attracted to when more female peers are present, we next estimate separate models in which we split STEM into four subgroups, shown in Panel A of Table 6: (1) Biology, (2) Math and Physics, (3) ICT and Engineering and (4) Manufacturing and Construction. . A comparison of the point estimates in Columns (1)–(4) reveals that the coefficients for women are relatively similar across STEM subfields. The point estimate for ICT & Engineering is marginally smaller and less precisely estimated, but not significantly different, from the other subfields. For men, the effect of having more female peers is stronger for entering Math and Physics and ICT and Engineering, which are the most male-dominated, gender-stereotypical STEM subfields.

We next ask which other study fields become more attractive for women and less attractive for men when having experienced a larger share of female high school peers. In Panel B of Table 6, we examine how the peer composition affects graduation from (1) Health Sciences, (2) Education Studies, (3) Arts and Humanities, (4), Social Sciences, and (5) Business and Law. Only Health Sciences—a field heavily dominated by women—becomes significantly more popular among women who had more female peers in high school (Column 3).[25] For men, we also find that more female peers make choices more gender-stereotypical: a larger share of female peers in high school decreases men's probability of graduation with a college major within Health and Education Sciences. Columns (2)–(5) show that the high school gender composition does not influence women's probability of completing a college degree in Education, Arts and Humanities, Social Sciences, or Business and Law. Consequently, our results suggest that women exposed to more

[24] As a robustness check, we also test whether results differ between students who attend a high school that is one of several in the municipality and those who attend the only high school in the municipality. If our estimates were driven by unobserved, time-variant selection into high schools, we would expect effects to differ substantially based on how much choice students have at the local level. Table A3 in the Appendix reports split sample regressions by the number of high schools in the municipality. For women, point estimates are very similar in regions that have only one high school in the municipality. For men, we find that the effect of peers seems to be somewhat larger in municipalities with only one high school. While it is possible that peer effects may differ by the number of high schools in the municipality, we think that these results provide additional support for the validity of our peer effects estimates as we find the same effects for regions where students had a de facto only very limited school choice.

[25] Of all women who enter health sciences 50 percent study nursing and midwifery, 20 percent study medicine, and 13 percent therapy and rehabilitation.

female peers in high school substitute STEM studies with careers in the health sector. In Subsection C, we investigate the potential labor market consequences of these education choices in more detail.

[Table 6 here]

B. *Underlying Mechanisms and Heterogeneity*

Why does the gender composition in high school affect students' decision to enter STEM fields? We shed some light on the underlying mechanism in two ways. First, we investigate whether the high school gender composition affects final high school GPA, which students use to apply for college. Second, we split the sample based on parental educational level and field of education to learn more about whether some groups of students are more sensitive than others to the gender composition at their school. Considering heterogeneity in a parent's field of education might help us understand whether STEM role models at home moderate or perpetuate the influence of peers.

In order to assess the first proposed mechanism, we investigate whether gender composition directly affects students' study ability or preparedness to enter STEM studies. This appears plausible as the peer effects literature has shown that gender composition can impact students' performance (Hill, 2017; Hoxby, 2000; Lavy and Schlosser, 2011). If the gender composition affects student performance differentially depending on the student's own gender, this effect may in part rationalize the effects on the choice of college major. Table 7 provides estimation results supporting such a mechanism.

Column (1) in Table 7 shows that the gender composition does not affect women's GPAs; the point estimate is tiny and not statistically significant. On the contrary, male students achieve a higher GPA when they are exposed to a high school cohort with more female peers. Ten percentage points more female peers raises the GPA of male students by 1.26 percent of a standard deviation.[26] Our finding is consistent with Lavy & Schlosser (2011) and Hill (2017), which show that men in high school and college achieve better grades when there are more female peers in their cohort. Importantly, the vast majority of STEM college programs in Denmark do not have a binding high school GPA threshold for admission. This rules out the possibility that gender composition

[26] All results in Table 7 are robust to the inclusion of cohort-by-school fixed effects.

mechanically affects STEM enrollment through the impact on male students' GPA.

The fact that male students achieve a higher GPA when a higher proportion of female peers are present may in part explain why fewer women and more men enter STEM studies in cohorts with more women. Given their higher GPAs, men might feel better prepared for STEM studies, which generally attract students with better high school grades. In contrast, women do not perform differently in high school, but might infer from the gender gap in high school GPA that they are less prepared or "suited" to enter STEM studies than their male peers. These results are consistent with Zölitz and Feld (2017), who also find evidence of gender-specific performance responses that can rationalize students' specialization choices.

Table 7 further investigates whether the influence of female peers is similar for students with less- versus highly educated parents (Columns 3 and 4) and parents with STEM educations (Columns 5 and 6). Our motivation for splitting the sample by parental education is that students from nonacademic families may have less information about college majors and associated occupations and therefore be more sensitive to their peers' choices. If peers can provide information about study fields that is not available to students who have non-college educated parents, we would expect students from families with less-educated parents to be more sensitive to the peer composition. Similarly, students with a STEM-educated father or mother might have better information about STEM studies and careers. If parents serve as strong role models that shape the choices of their children, we would expect that students with a (same sex) parent in STEM are less sensitive to peer influences.

Column (3) shows estimates for the subsample of students who have parents without a college education, while Column (4) shows the same model for students where one or both parents have a college degree.[27] The results show that the influence of peers is twice as strong for women with parents without college education relative to women with college-educated parents. However, effects for men do not vary by parental education. One interpretation of these results is that women are more likely to follow the choices of their peers when they lack parents who have a greater capacity to help them find information about higher education or share their own college experiences. It is also possible that the study choice of parents who attended college provides an additional reference point that moderates the impact of peer effects in high school.

We next investigate whether having a father or mother with a STEM education mediates

[27] To facilitate comparison, Column (2) in Table 7 reports point estimates for the full sample from our preferred specification in Column (5) of Table 5.

effects of the peer composition. Column (5) shows that the effect of peers on STEM completion is similar for women with a father in STEM and somewhat weaker and not statistically significant for men with fathers in STEM. These results suggest that men who have a STEM father are less susceptible to peer influences. Column (6) shows that these results are mirrored for women with STEM mothers. Strikingly, women with STEM mothers are not significantly affected by the peer gender composition. The point estimate is in fact positive, which suggests that STEM mothers counteract the effect of peers on their daughters' specialization choice. While the group of individuals who have a mother in a STEM field is small, these results show that women who have a STEM mother as a role model are unaffected by peers. While these results remain suggestive, they could imply that access to a non-stereotypical same sex role model in the family is more powerful in shaping women's STEM interest than the influence of high school peers.

[Table 7 here]

C. *Long-Run Effects on Labor Market Outcomes*

Given our finding that high school gender composition affects the probability of enrolling in a STEM college major and completing such education, we now want to know whether these gender peer effects also persist into occupational choice and whether they have consequences for labor earnings.

[Table 8 here]

Table 8 shows the impact of high school peers on labor market outcomes 15 and 20 years after high school entry (Columns 1 and 2). Fifteen years after high school entry, the median age of individuals is 29 years and most should have completed higher education and entered the labor market.[28] Table 8 shows that peers' influence on STEM field choice are persistent and closely mapped into women's occupational choices. A larger share of female high school peers reduces

[28] Brenøe and Lundberg (2017) show that after age 30, the share of a cohort that has completed a college or university degree is almost constant, indicating that by age 31 most individuals have completed higher education.

the probability that women work in a STEM-related occupation, while it has no impact on men's occupational choice. A 10-percentage-point increase in the proportion of female peers decreases the probability that women work in a STEM occupation 20 years after high school entry by 0.50 percentage points—a 4.6 percent decrease from the baseline (Column 2). In contrast, men's probability of working within STEM fields is not affected by the peer composition. This suggests that, although men's study choice is affected, those who enroll in and graduate from STEM fields due to peer effects are not more or less likely to end up working in STEM jobs.

Given these long-run effects on occupational choice, we next test whether individuals' labor earnings are affected. Does the proportion of female high school peers contribute to the gender wage gap? Columns (3) and (4) shed light on this question and show estimates for earnings 15 and 20 years after high school entry. We find that the high school gender composition has lasting effects on women's but not on men's earnings. A 10-percentage-point increase in the proportion of female peers decreases women's earnings 15 years after high school entry by 0.39 percentile points—nearly a one-percent decline from the baseline—and has no impact on men's earnings. Twenty years after high school entry, we still do not find any effect for men but an even larger effect of 0.67 for women. This increase in magnitude for women is in line with a similar increase in the effect on their STEM occupational participation. The estimate from our second strategy (Panel B), which includes cohort-by-school fixed effects, suggests that having a 10-percentage-point-larger share of female high school peers increases the gender earnings gap by 0.75 percentile points, corresponding to an increase of 5.4 percent (Column 4). Thus, our results show that high school gender composition has lasting impacts on the gender wage gap, likely mediated by the impact on STEM participation.

D. *Long-Run Effects on Fertility*

Given that we find a substantial effect of the high school gender composition on women's earnings, we next examine fertility as an underlying mechanism. While fertility is an interesting outcome in itself, we believe that there are a least two reasons the relationship between high school peers and fertility is important for the interpretation of our earnings results. First, women entering a STEM career because of peers may have fewer children once they realize that STEM jobs are less family-friendly. The more competitive environment, longer working hours, and lower job flexibility might make it harder to combine children with work obligations. In this case, changes in fertility should be interpreted as an unintended consequence of the high school composition that "pushed" some individuals in or out of STEM careers. Second, if the high school gender composition directly affects fertility preferences, then the documented effects for the gender wage gap may simply be due to the fact that having children reduces earnings for women (Kleven et al. 2018; Lundborg et al. 2017). In this case, lower earnings for women should not be attributed to their staying out of STEM fields per se, but may instead be a consequence of the "career cost of children" (Adda et. al., 2017).

[Table 9 here]

Table 9 documents the medium- and long-run consequence of high school gender composition on fertility. Columns (1)–(4) provide a detailed analysis on the timing of fertility effects on the extensive margin. While Column (1) shows little effect on fertility during the first 5 years after high school entry, we see a clear impact on the probability of having any children within 10 years after high school entry when individuals are around 26 years old and for the most part have entered the labor market (Column 2). We find that having a larger proportion of female high school peers increases women's and decreases men's probability of having any children by the age of 26. Column (4) shows that 20 years after high school entry, women are no longer less likely to have any children, but we still find a persistent effect for men.[29]

In Columns (5)–(8) of Table 9, we estimate effects at the intensive fertility margin by testing whether female peers affect the number of children individuals have. We find that women, who had more female peers and thus were less likely to enter STEM careers, have more children by the age of 36. The increased fertility effect becomes visible shortly after the time of college completion and doubles in size within the first couple of years in the labor market (Column 7). By the time individuals are 36 years old, women exposed to 10-percentage-points more female peers have on average 0.02 more children, an increase of about 1.0 percent from the baseline of 1.66 children (Column 8). For men, we find that those exposed to more female peers have significantly fewer children by age 36.

Our results show that the proportion of female peers has not only significant impacts on

[29] Restricting the sample to older cohorts reveals very similar effects on fertility through older ages. Therefore, the reported affects are close to effects on complete fertility, especially for women.

study choice, occupational choice, and earnings, but affects fertility as well. Strikingly, a larger share of female high school peers increases women's number of children. If high school peers already affect men's and women's fertility preferences before choosing their field of study, women's shift from STEM to health-related studies (and the reverse for men) might partly be explained by changes in the desired number of children.

These findings on fertility raise the question of whether the negative effect of having more female high school peers on women's earnings documented in Table 8 is driven by: (1) women's not participating in STEM, or by (2) their increased fertility. Although we cannot provide causal evidence that distinguishes between these mechanisms, we still attempt to provide some suggestive evidence. In Table 10, we test whether we can reduce or eliminate the estimated effect of the proportion of female peers on the earnings by including controls for fertility and STEM participation. While the inclusion of these endogenous variables, commonly referred to as "bad controls," complicates the interpretation, we still believe that this exercise hints at whether the earnings results can exclusively be attributed to gender peer effects on STEM participation or fertility. Column (1) shows the estimated main effects on earnings we observed in Column 4 of Table 8. In Column (2), we add controls for the individual's fertility by including gender-specific dummies for the number of children of the individual. In Column (3), we do not control for observed fertility but instead we control for STEM participation using gender-specific dummies for working in a STEM occupation and having a STEM college degree, and an interaction term between STEM occupation and education. Finally, Column (4) includes both the fertility and STEM controls.

[Table 10 here]

The results in Table 10 suggest that including either the fertility or STEM controls reduces the estimated peer effect on women's earnings and the gender gap in earnings by 40–50 percent. Once both sets of controls are included, the estimated gender peer effect is no longer significant and is only 21 percent of its original magnitude in Panel A. These findings indicate that the effects of the high school gender peer composition on fertility and STEM participation equally contribute to the long-run consequences for women's earnings. However, given the "bad control" problem we face here, we caution a causal interpretation of this result.

These results suggest that fertility cannot fully explain the lower earnings of women who were exposed to more female peers. These results are consistent with Kleven et al. (2018), who use Danish administrative data (similar to our data) and document that men and women experience very similar trends in labor earnings before the arrival of their first child. Yet, Kleven et al. (2018) show while women experience a large decline in earnings at the time of their first childbirth and still have earnings 20 percent below their initial level ten years later, men do not experience any change in their earnings after having children.

Taken together, our results suggest that there are two key factors to explain why women with more exposure to female peers have lower earnings. First, entering STEM jobs, which are arguably more competitive and have less flexible working hours, reduces women's fertility. Second, exposure to more female peers in high school increases women's fertility, which reduces their earnings independently of their STEM participation.

VI. Conclusion

This study demonstrates that the gender composition of high school peers affects students' decisions to undertake STEM studies in higher education. Our results show that a higher proportion of female high school peers makes study choices more gender-stereotypical. With more female peers present, women become less likely to enter STEM fields and more likely to enter Health Studies. Men also behave more gender-stereotypical and become more prone to enter STEM studies when exposed to more female peers.

For women, these gender peer effects in study choice have remarkably persistent long run effects on occupational choice, which remain visible 20 years after high school entry. Women who by chance were exposed to more female peers are less likely to work in STEM occupations and have lower earnings 20 years after high school entry. We further show that the high school gender composition affects individuals' fertility and provide suggestive evidence that gender peer effects on STEM participation and fertility equally contribute to the effect on women's earnings.

In conclusion, our results suggest that gender peer effects in high school shape preferences for study fields and thereby lead students to systematically different career trajectories. Our evidence on the underlying mechanisms remains suggestive, but indicates that a higher proportion of female peers affects the gender gap in high school GPA and may therefore foster the gender gap in STEM preparedness, which gives students reason to believe that they have a comparative advantage in a more gender-stereotypical college major. We also find that women with STEM-educated mothers are unaffected by the gender composition, which suggests that salient female role models may be able to counteract peer pressures in high school.

We believe this paper broadens our understanding of where gender differences in educational choice originate. Our findings emphasize that the social environment directly affects students' decisions to specialize within STEM fields and educational and occupational choices more generally. Moreover, it highlights the possibility that manipulating the gender composition in a given environment through affirmative action policies to achieve gender balance may have adverse and unintended consequences for fertility, gender segregation in college majors, and the labor market.

REFERENCES

Adda, J., Dustmann, C. and Stevens K. (2017). The Career Costs of Children. Journal of Political Economy 125 (2): 293–337.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic schools. Journal of Political Economy, 113(1), 151-184.

Anelli, M., and Peri, G. (forthcoming). The Effects of High School Peers' Gender on College Major, College Performance and Income. The Economic Journal.

Angrist, J. D. (2014). The Perils of Peer Effects. Labour Economics, 30, 98–108.

Bifulco, R., Fletcher, J. M., & Ross, S. L. (2011). The Effect of Classmate Characteristics on Post-secondary Outcomes: Evidence from the Add Health. American Economic Journal: Economic Policy, 3(1), 25-53.

Brenøe, A. A., & Lundberg, S. (2017). Gender Gaps in the Effects of Childhood Family Environment: Do They Persist into Adulthood? European Economic Review.

Carrell, S. E., Hoekstra, M. L. (2010). Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids. American Economic Journal: Applied Economics, 2(1), 211-228.

Carrell, S. E., Hoekstra, M., Kuka, E. (2016). The Long-run Effects of Disruptive Peers. NBER Working Paper No. 22042.

De Giorgi, G., Pellizzari, M. and Woolston, W.G., (2012). Class Size and Class Heterogeneity. Journal of the European Economic Association 10(4): 795-830.

Epple, D. and Romano, R. (2011). Peer Effects in Education: A Survey of the Theory and Evidence. Handbook of Social Economics, 1 (11), 1053–1163.

Feld, J., & Zölitz, U. (2017). Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. Journal of Labor Economics, 35(2), 387-428.

Guryan, J., Kroft, K., and Notowidigdo, M. J. (2009). Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. American Economic Journal: Applied Economics, 1 (4), 34–68.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does Peer Ability Affect Student Achievement? Journal of Applied Econometrics, 18(5), 527-544.

Hill, A. J. (2015). The Girl Next Door: The Effect of Opposite Gender Friends on High School Achievement. American Economic Journal: Applied Economics, 7(3), 147-177.

Hill, A. J. (2017). The Positive Influence of Female College Students on their Male Peers. Labour Economics, 44, 151-160.

Hoxby, C. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation. NBER Working Paper No. 7867.

Hoxby, C. M., Weingarth, G. (2005). Taking Race Out Of the Equation: School Reassignment and the Structure of Peer Effects. NBER Working Paper No. 7867.

Jackson, C. K. (2016). The Effect of Single-sex Education on Test Scores, School Completion, Arrests, and Teen Motherhood: Evidence from School Transitions. NBER Working Paper No.22222.

Jetter, M., and Walker, J. K. (2016). Gender in Jeopardy!: The Role of Opponent Gender in High-stakes Competition.

Joensen, J. S., and Nielsen, H. S. (2016). Mathematics and Gender: Heterogeneity in Causes and Consequences. The Economic Journal, 126 (593), 1129–1163.

Kahn, S. and D. Ginther (2017). Women and STEM. Technical report, National Bureau of Economic Research.

Kleven, H., Landais C., and Søgaard J.E. (2018). Children and Gender Inequality: Evidence from Denmark. NBER Working Paper No. 24219.

Lavy, V., Paserman, M. D., & Schlosser, A. (2012). Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. The Economic Journal, 122(559), 208-237.

Lavy, V., and Schlosser, A. (2011). Mechanisms and Impacts of Gender Peer Effects at School. American Economic Journal: Applied Economics, 3 (2), 1–33.

Lee, S., Turner, L.J., Woo, S. and Kim, K., (2014). All or Nothing? The Impact of School and Classroom Gender Composition on Effort and Academic Achievement. NBER Working Paper, No. 20722.

Lefgren, L. (2004). Educational Peer Effects and the Chicago Public Schools. Journal of Urban Economics, 56(2), 169-191.

Lundborg, P., Plug, E., & Rasmussen, A. (2017). Fertility Effects on Female Labor Supply: IV evidence from IVF treatments. American Economic Review. 107 (6). 1611-1637.

OECD (2016). Education at a Glance 2016: OECD Indicators.

Oosterbeek, H., and van Ewijk, R. (2014). Gender Peer Effects in University: Evidence from a Randomized Experiment. Economics of Education Review, 38, 51–63.

Schneeweis, N., and Zweimüller, M. (2012). Girls, Girls, Girls: Gender Composition and Female School Choice. Economics of Education Review, 31 (4), 482–500.

Sohn, H. (2016). Mean and Distributional Impact of Single-Sex High Schools on Students' Cognitive Achievement, Major Choice, and Test-taking Behavior: Evidence from a Random Assignment Policy in Seoul, Korea. Economics of Education Review, 52, 155-175.

UNESCO Institute for Statistics (UIS). (2012). International Standard Classification of Education: ISCED 2011. UIS, Montreal, Quebec.

Vigdor, J., & Nechyba, T. (2006). Peer Effects in North Carolina Public Schools. In In Schools and the Equal Opportunities Problem, edited by Ludger Woessmann and Paul, MIT Press.

Whitmore, D. (2005). Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment. American Economic Review, 95(2), 199-203.

Zölitz, U., Feld, J. (2017). The Effect of Peer Gender on Major Choice. University of Zurich Working Paper No. 270.

TABLES

Table 1: Descriptive Statistics

	Women			Men		
	N	Mean	SD	N	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Student outcome variables						
Any college enrollment	81820	0.794	0.404	100391	0.784	0.411
Any STEM enrollment	81820	0.210	0.407	100391	0.378	0.485
Any Science/Math enrollment	81820	0.118	0.323	100391	0.131	0.338
Any Technology/Engineering enrollment	81820	0.104	0.306	100391	0.279	0.448
Any Health enrollment	81820	0.268	0.443	100391	0.063	0.242
Any college completion	81820	0.710	0.454	100391	0.650	0.477
Highest completed degree within STEM	81820	0.135	0.342	100391	0.253	0.435
Highest completed degree within Science/Math	81820	0.062	0.241	100391	0.056	0.229
Highest completed degree within Technology/Engineering	81820	0.073	0.260	100391	0.197	0.398
Highest completed degree within Health	81820	0.217	0.412	100391	0.047	0.211
Highest completed degree within Health	81820	0.217	0.412	100391	0.047	0.211
Highest completed degree within Education	81820	0.075	0.264	100391	0.040	0.197
Highest completed degree within Arts & Humanities	81820	0.058	0.234	100391	0.045	0.208
Highest completed degree within Social Sciences	81820	0.049	0.215	100391	0.062	0.242
Highest completed degree within Business, Admin, Law	81820	0.093	0.290	100391	0.134	0.341
STEM occupation 15 years after high school entry	81270	0.103	0.304	99752	0.246	0.431
STEM occupation 20 years after high school entry	80114	0.109	0.312	98126	0.255	0.436
Annual labor earnings 15 years after high school entry	81265	241.944	133.320	99747	311.796	187.452
Annual labor earnings 20 years after high school entry	80108	327.915	167.152	98121	454.849	272.011
Earnings percentile by cohort 15 years after HS entry	81265	52.2	22.9	99747	62.6	26.3
Earnings percentile by cohort 20 years after HS entry	80108	56.6	23.7	98121	70.6	26
Any children 5 years after HS entry	81820	0.019	0.136	100391	0.007	0.081
Any children 10 years after HS entry	81820	0.208	0.406	100391	0.108	0.310
Any children 15 years after HS entry	81820	0.611	0.488	100391	0.444	0.497
Any children 20 years after HS entry	81820	0.796	0.403	100391	0.683	0.465
# of children 5 years after HS entry	81820	0.020	0.151	100391	0.007	0.088
# of children 10 years after HS entry	81820	0.259	0.549	100391	0.127	0.392
# of children 15 years after HS entry	81820	1.015	0.965	100391	0.670	0.857
# of children 20 years after HS entry	81820	1.656	1.066	100391	1.317	1.080

Table 1: Descriptive Statistics – continued

Panel B: Student level background variables						
Mother has tertiary education	81820	0.3557	0.4787	100391	0.3911	0.488
Mother has upper secondary education	81820	0.3696	0.4827	100391	0.3648	0.4814
Mother has less than upper secondary education	81820	0.2558	0.4363	100391	0.2208	0.4148
Father has tertiary education	81820	0.3759	0.4844	100391	0.4305	0.4951
Father has upper secondary education	81820	0.3716	0.4832	100391	0.3526	0.4778
Father has less than upper secondary education	81820	0.2053	0.404	100391	0.1602	0.3668
First generation immigrant	81820	0.008	0.091	100391	0.012	0.107
Second generation immigrant	81820	0.006	0.075	100391	0.008	0.088
Child is adopted	81820	0.009	0.095	100391	0.007	0.084
Mother's age at birth	81052	26.429	4.769	99156	26.563	4.752
Mother <22 years at birth	81052	0.141	0.348	99156	0.133	0.340
Firstborn	81820	0.505	0.500	100391	0.516	0.500
Number of siblings	81689	1.472	0.908	100077	1.436	0.901
Lives with both parents at age 10	81820	0.861	0.346	100391	0.841	0.366
Panel C: School level variables						
Proportion female peers	81820	0.454	0.066	100391	0.447	0.067
Number of students in cohort	81820	107.878	30.776	100391	108.470	31.053
Cohort	81820	1987	4.338	100391	1987	4.345
Number of feeding municipalities	81820	6.330	3.304	100391	6.305	3.215
2+ high schools in municipality	81820	0.446	0.497	100391	0.459	0.498

NOTE: Annual labor earnings are measured in thousand Danish krones (1 USD = 6.6 DKK). Annual labor earnings are adjusted for 2015prices.

Table 2: Balancing Test I

Does Student Gender Predict Proportion of Female Peers?

	Dependent variable: Proportion female peers in high school cohort							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.0009	0.0014	0.0009	0.0012	0.0008	0.0012	0.0007	0.0011
	(0.0013)	(0.0010)	(0.0009)	(0.0009)	(0.0013)	(0.0010)	(0.0009)	(0.0009)
Observations	182211	182211	182211	182211	182211	182211	182211	182211
p-value (Female)	0.491	0.186	0.344	0.174	0.521	0.237	0.443	0.230
High school and cohort fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
School level controls	-	-	-	-	✓	✓	✓	✓
School-specific time trends	-	linear	quadratic	cubed	-	linear	quadratic	cubed

NOTE: The dependent variable in all columns is the proportion of female peers in the high school cohort of an individual. All columns include cohort fixed effects and school fixed effects. Following the Guryan, Kroft, and Notowidigdo (2009) correction method, we control for the leave-out mean of the proportion of female peers across cohorts within the school in all columns. *School level controls* included in Columns (5)-(8) include an indicator if any student in the cohort is older than 20 years at high school entry, dummies for number of students without information on gender (ranging from 0 to 2), indicators for large changes in cohort size compared to previous years, the female share in the language track, an indicator if the high school has no language track, indicators for exposure to experiment on course curriculum, and cubed cohort size.

Table 3: Balancing Test II

Does Proportion of Female Peers Predict Student Background Characteristics?

	(1)	(2)	(3)	(4)	(5)	(6) Across all models
Number of performed tests	38	38	38	38	38	190
Number significant at 1 percent level	1	1	0	0	0	2
Number significant at 5 percent level	1	1	0	1	1	4
Number significant at 10 percent level	4	5	2	4	2	17
Share significant at 1 percent	0.026	0.026	0.000	0.000	0.000	0.011
Share significant at 5 percent	0.026	0.026	0.000	0.026	0.026	0.021
Share significant at 10 percent	0.105	0.132	0.053	0.105	0.053	0.089
School level controls	-	✓	✓	✓	✓	
School-specific time trends	-	-	linear	quadratic	cubed	

NOTE: This Table is based on 190 separate OLS regressions shown in Appendix Table A1. All regressions include cohort fixed effects and school fixed effects. *School level controls* included in Columns (2)-(5) include an indicator if any student in the cohort is older than 20 years at high school entry, dummies for number of students without information on gender (ranging from 0 to 2), indicators for large changes in cohort size compared to previous years, the female share in the language track, an indicator if the high school has no language track, indicators for exposure to experiment on course curriculum, and cubed cohort size. Standard errors are clustered at the school-cohort level.

Table 4: Balancing Test III
School Level Autocorrelation in the Propotion Female Students

For what proportion of schools does the proportion of female students in t-1 significantly predict the proportion female students in t?

Proportion of school coefficients	(1)	(2)	(3)	(4)	Across all models
Significant at 1 percent level	1.57%	0.79%	0.79%	0.79%	0.98%
Significant at 5 percent level	2.36%	2.36%	3.15%	5.51%	3.35%
Significant at 10 percent level	6.30%	5.51%	8.66%	14.96%	8.86%
School-specific time trends	-	linear	quadratic	cubed	

NOTE: This table provides summary statistics of significance for 127 separate bivariate school-level regressions that only include the respective school-specific trend variable(s).

Table 5: The Impact of Peer Gender on STEM Enrollment and STEM Degree Completion

Panel A: STEM Enrollment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female * Proportion female peers	-0.211***	-0.155***	-0.139***	-0.135***	-0.138***	-0.148***	-0.233***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.027)	(0.036)
Male * Proportion female peers	0.056**	0.104***	0.086***	0.091***	0.092***	0.081***	
	(0.028)	(0.028)	(0.027)	(0.027)	(0.027)	(0.028)	
Female	-0.047***	-0.049***	-0.064***	-0.063***	-0.062***	-0.062***	-0.061***
	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Observations	182211	182211	182211	182211	182211	182211	182211
Mean dependent variable women	0.210	0.210	0.210	0.210	0.210	0.210	0.210
Mean dependent variable men	0.378	0.378	0.378	0.378	0.378	0.378	0.378
p-values of test for gender equality of "Proportion female peers"	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	-
School fixed effects	-	✓	✓	✓	✓	✓	
Cohort fixed effects	-	✓	✓	✓	✓	✓	
Cohort controls	-	✓	✓	✓	✓	✓	
Individual & peer level controls	-	-	✓	✓	✓	✓	✓
School-specific time trends	-	-	-	linear	quadratic	cubed	-
Cohort-by-School fixed effects	-	-	-	-	-	-	✓

Panel B: STEM Completion

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female * Proportion female peers	-0.104***	-0.112***	-0.097***	-0.097***	-0.099***	-0.110***	-0.197***
	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.030)
Male * Proportion female peers	0.122***	0.106***	0.093***	0.092***	0.093***	0.082***	
	(0.025)	(0.024)	(0.023)	(0.023)	(0.023)	(0.024)	
Female	-0.016	-0.019	-0.032**	-0.033**	-0.031**	-0.031**	-0.030**
	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Observations	182211	182211	182211	182211	182211	182211	182211
Mean dependent variable women	0.135	0.135	0.135	0.135	0.135	0.135	0.135
Mean dependent variable men	0.253	0.253	0.253	0.253	0.253	0.253	0.253
p-values of test for gender equality of "Proportion female peers"	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	-
School fixed effects	-	✓	✓	✓	✓	✓	
Cohort fixed effects	-	✓	✓	✓	✓	✓	
Cohort controls	-	✓	✓	✓	✓	✓	
Individual & peer level controls	-	-	✓	✓	✓	✓	✓
School-specific time trends	-	-	-	linear	quadratic	cubed	-
Cohort-by-School fixed effects	-	-	-	-	-	-	✓

NOTE: The dependent variable in all columns of Panel A is an indicator for whether the student ever enrolled in a STEM program in college within 20 years after high school entry. The dependent variable in all columns of Panel B is an indicator for whether the student's highest completed education is at least at the college level and is within STEM. Colum (7) does not include peer-level variables because these are highly collinear with the cohort-by-high school fixed effects. Standard errors clustered at the school-cohort level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Impact of High School Gender Composition on Graduation in Various Fields

Panel A: STEM Subfields	Biology	Math / Physics	ICT / Engineering	Manufacturing / Construction
	(1)	(2)	(3)	(4)
Female * Proportion female peers	-0.030***	-0.026**	-0.019	-0.023**
	(0.010)	(0.010)	(0.015)	(0.010)
Male * Proportion female peers	0.011	0.030***	0.039**	0.012
	(0.007)	(0.011)	(0.019)	(0.011)
Female	0.039***	0.012**	-0.083***	-0.000
	(0.005)	(0.006)	(0.010)	(0.006)
Observations	182211	182211	182211	182211
Mean dependent variable women	0.035	0.027	0.044	0.029
Mean dependent variable men	0.015	0.040	0.153	0.045
p-values of test for gender equality of "Proportion female peers"	0.001	<.0001	0.008	0.013

Panel B: Other Fields of Study	Health Sciences	Education	Arts / Humanities	Social Sciences	Business / Law
	(1)	(2)	(3)	(4)	(5)
Female * Proportion female peers	0.110***	0.022	0.001	-0.019	0.017
	(0.024)	(0.015)	(0.014)	(0.013)	(0.019)
Male * Proportion female peers	-0.077***	-0.037***	0.014	0.004	0.004
	(0.015)	(0.012)	(0.011)	(0.013)	(0.019)
Female	0.085***	0.007	0.020***	-0.001	-0.045***
	(0.012)	(0.008)	(0.007)	(0.007)	(0.011)
Observations	182211	182211	182211	182211	182211
Mean dependent variable women	0.217	0.075	0.058	0.049	0.093
Mean dependent variable men	0.047	0.040	0.045	0.062	0.134
p-values of test for gender equality of "Proportion female peers"	<.0001	0.001	0.407	0.144	0.576

NOTE: All models control for linear school-specific time trends, cohort fixed effects, school fixed effects, cubed cohort size, indicators for large cohort size changes compared to previous years as well as a large set of individual and leave-out-mean peer controls shown in Panel B of Table 1. All results are robust to the inclusion of cohort-by-school fixed effects. Standard errors clustered at the school-cohort level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Mechanisms and Heterogeneity

Subgroup:	Full sample		No parent has college degree	1+ parent has college degree	Father has STEM education	Mother has STEM education
Dependent variable:	Std. High School GPA	STEM completion	STEM completion	STEM completion	STEM completion	STEM completion
	(1)	(2)	(3)	(4)	(5)	(6)
Female * Proportion female peers	-0.006	-0.097***	-0.128***	-0.067**	-0.112***	0.072
	(0.062)	(0.022)	(0.029)	(0.031)	(0.038)	(0.100)
Male * Proportion female peers	0.126**	0.092***	0.088***	0.094***	0.052	0.238**
	(0.058)	(0.023)	(0.032)	(0.030)	(0.040)	(0.101)
Female	0.061	-0.033**	-0.020	-0.047***	-0.064***	-0.031
	(0.038)	(0.013)	(0.019)	(0.018)	(0.023)	(0.057)
Observations	159603	182211	83783	98428	60544	9955
Mean dependent variable women	-0.011	0.135	0.165	0.103	0.158	0.207
Mean dependent variable men	0.007	0.253	0.280	0.217	0.298	0.307
p-values of test for gender equality of "Proportion female peers"	0.110	<.0001	0.008	0.001	0.001	0.192

NOTE: The dependent variable in column (1) is the grade point average (GPA) at the end of high school, standardized with a mean of zero and standard deviation of one, and is observed for those students who completed general academic high school. All models control for school-specific linear time trends, cohort fixed effects, school fixed effects, cubed cohort size, indicators for large cohort size changes compared to previous years, as well as a large set of individual and leave-out-mean peer controls shown in Panel B of Table 1. All results are robust to the inclusion of cohort-by-school fixed effects. Standard errors clustered at the school-cohort level are in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 8: Impact of High School Gender Composition on Labor Market Outcomes

Panel A: Cohort and School fixed effects	Working in STEM occupation		Earnings percentile	
	11–15 years after HS entry (1)	16–20 years after HS entry (2)	11–15 years after HS entry (3)	16–20 years after HS entry (4)
Female * Proportion female peers	-0.032*	-0.050**	-3.898***	-6.655***
	(0.019)	(0.020)	(1.362)	(1.471)
Male * Proportion female peers	0.020	0.030	-1.222	0.082
	(0.022)	(0.022)	(1.431)	(1.433)
Female	-0.124***	-0.117***	-9.725***	-11.538***
	(0.012)	(0.013)	(0.855)	(0.911)
Observations	181022	178240	181012	178229
Mean dependent variable women	0.103	0.109	52.178	56.633
Mean dependent variable men	0.246	0.255	62.567	70.615
p-values of test for gender equality of "Proportion female peers"	0.048	0.004	0.148	0.001

Panel B: Cohort-by-School fixed effects	Working in STEM occupation		Earnings percentile	
	11–15 years after HS entry (1)	16–20 years after HS entry (2)	11–15 years after HS entry (3)	16–20 years after HS entry (4)
Female * Proportion female peers	-0.059**	-0.088***	-3.239*	-7.508***
	(0.027)	(0.029)	(1.892)	(2.032)
Female	-0.122***	-0.113***	-9.462***	-11.196***
	(0.012)	(0.013)	(0.864)	(0.917)
Observations	181022	178240	181012	178229
Mean dependent variable women	0.103	0.109	52.178	56.633
Mean dependent variable men	0.246	0.255	62.567	70.615

NOTE: The dependent variable in columns (1) and (2) is an indicator for working in a STEM occupation; it takes the value one if the individual for at least half the period works in a STEM occupation within the Danish version of ISCO codes 21, 25, 31 or 35 (Science and Engineering Professionals, Information and Communications Technology Professionals, Science and Engineering Associate Professionals, or Information and Communications Technicians). The dependent variable in columns (3) and (4) is the average of the individual's labor earnings percentile during the five-year period, calculated by year of birth and age using the entire Danish population as a reference group. All models in Panel A control for school-specific linear time trends, cohort fixed effects, school fixed effects, cubed cohort size, indicators for large cohort size changes compared to previous years, as well as a large set of individual and leave-out-mean peer controls shown in Panel B of Table 1. All models in Panel B control for cohort-by-school fixed effects and individual controls. Standard errors clustered at the school-cohort level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Impact of High School Gender Composition on Fertility

Panel A: Cohort and School fixed effects	Dependent variable: Any children				Dependent variable: Number of children			
	5 years after HS entry	10 years after HS entry	15 years after HS entry	20 years after HS entry	5 years after HS entry	10 years after HS entry	15 years after HS entry	20 years after HS entry
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female * Proportion female peers	0.009	0.065***	0.023	0.013	0.011	0.103***	0.210***	0.168***
	-0.008	-0.022	-0.029	-0.025	-0.008	-0.03	-0.053	-0.062
Male * Proportion female peers	-0.009*	-0.065***	-0.085***	-0.051**	-0.007	-0.079***	-0.177***	-0.152**
	-0.005	-0.018	-0.027	-0.025	-0.005	-0.023	-0.048	-0.06
Female	0.004	0.040***	0.114***	0.081***	0.005	0.047***	0.161***	0.181***
	-0.004	-0.012	-0.017	-0.015	-0.004	-0.016	-0.03	-0.035
Observations	182211	182211	182211	182211	182211	182211	182211	182211
Mean dependent variable women	0.019	0.208	0.611	0.796	0.02	0.259	1.015	1.656
Mean dependent variable men	0.007	0.108	0.444	0.683	0.007	0.127	0.67	1.317
p-values of test for gender equality of "Proportion female peers"	0.027	<.0001	0.003	0.044	0.042	<.0001	<.0001	<.0001

Panel B: Cohort-by-School fixed effects	Dependent variable: Any children				Dependent variable: Number of children			
	5 years after HS entry	10 years after HS entry	15 years after HS entry	20 years after HS entry	5 years after HS entry	10 years after HS entry	15 years after HS entry	20 years after HS entry
Female * Proportion female peers	0.017***	0.126***	0.096***	0.056*	0.017*	0.181***	0.366***	0.294***
	-0.008	-0.027	-0.037	-0.032	-0.009	-0.035	-0.067	-0.079
Female	0.004	0.042***	0.120***	0.085***	0.005	0.048***	0.172***	0.194***
	-0.004	-0.012	-0.017	-0.015	-0.004	-0.016	-0.03	-0.035
Observations	182211	182211	182211	182211	182211	182211	182211	182211
Mean dependent variable women	0.019	0.208	0.611	0.796	0.02	0.259	1.015	1.656
Mean dependent variable men	0.007	0.108	0.444	0.683	0.007	0.127	0.67	1.317

NOTE: Five years after high school entry, individuals should be at the beginning of their college studies. Ten years after high school entry, individuals are around 26 years old and should have completed college education. Fifteen years after high school, individuals have been in the labor market for approximately 5 years if they attended college. Twenty years after high school, individuals are about 36 years old. All models in Panel A control for school-specific time trends, cohort fixed effects, school fixed effects, cubed cohort size, indicators for large cohort size changes compared to previous years as well as a large set of individual and leave-out-mean peer controls shown in Panel B of Table 1. All models in Panel B include cohort-by-school fixed effects and individual controls. Standard errors clustered at the school-cohort level are in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 10: Impact of High School Peers on Earnings—Including Fertility and STEM Controls

Panel A: Cohort and School fixed effects — Dependent Variable: Earnings Percentile

	Main Result	Fertility Controls	STEM Controls	Fertility and STEM Controls
	(1)	(2)	(3)	(4)
Female * Proportion female peers	-6.655***	-4.001***	-3.624***	-1.370
	(1.471)	(1.423)	(1.325)	(1.280)
Male * Proportion female peers	0.082	-1.022	-0.185	-1.206
	(1.433)	(1.371)	(1.299)	(1.256)
Female	-11.538***	-1.575*	-13.494***	-3.943***
	(0.911)	(0.900)	(0.817)	(0.806)
Observations	178229	178229	178229	178229
Mean dependent variable women	56.633	56.633	56.633	56.633
Mean dependent variable men	70.615	70.615	70.615	70.615
p-values of test for gender equality of "Proportion female peers"	0.001	0.119	0.054	0.924

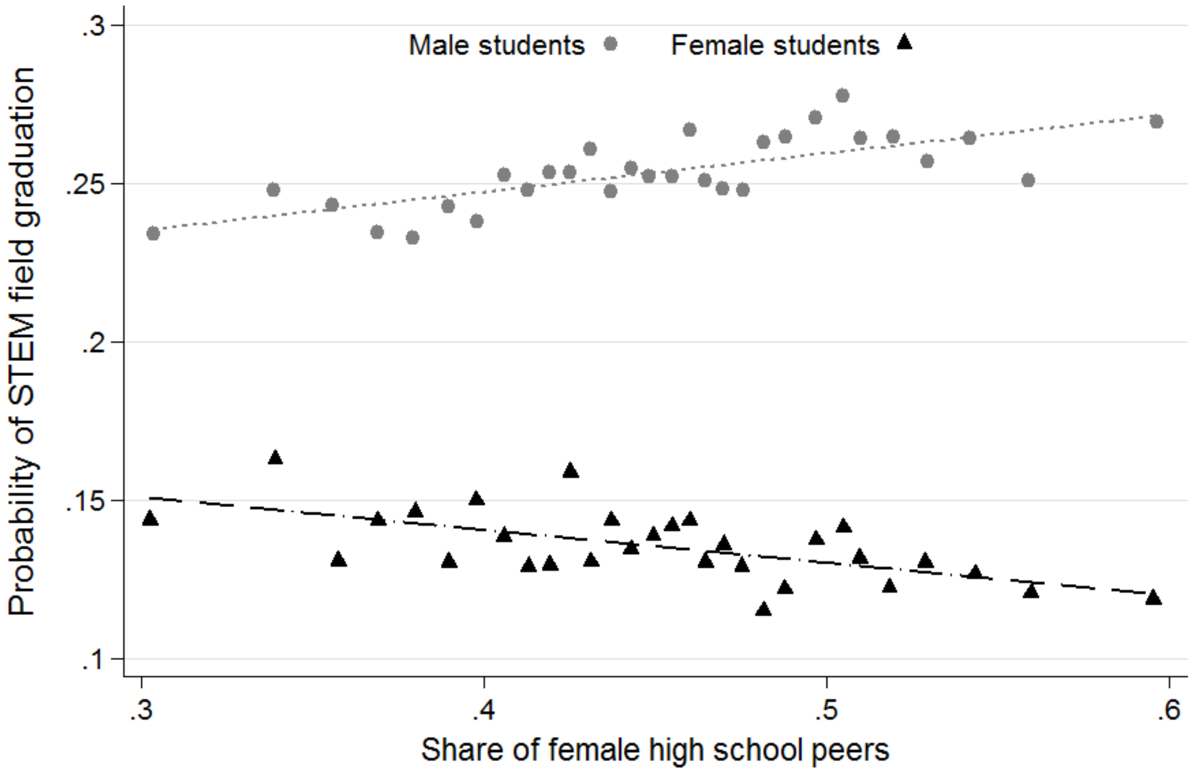
Panel B: Cohort-by-School fixed effects — Dependent Variable: Earnings Percentile

	Main Result	Fertility Controls	STEM Controls	Fertility and STEM Controls
	(1)	(2)	(3)	(4)
Female * Proportion female peers	-7.508***	-3.646*	-3.925**	-0.590
	(2.032)	(1.942)	(1.815)	(1.746)
Female	-11.196***	-1.306	-13.282***	-3.773***
	(0.917)	(0.904)	(0.821)	(0.808)
Observations	178229	178229	178229	178229
Mean dependent variable women	56.633	56.633	56.633	56.633
Mean dependent variable men	70.615	70.615	70.615	70.615

NOTE: The dependent variable in all columns is the average labor earnings percentile 16–20 years after high school entry, calculated by year of birth and age using the entire Danish population as a reference group. All models in Panel A control for school-specific linear time trends, cohort fixed effects, school fixed effects, cubed cohort size, indicators for large cohort size changes compared to previous years as well as a large set of individual and leave-out-mean peer controls shown in Panel B of Table 1. All models in Panel B control for cohort-by-school fixed effects and individual controls. *Fertility Controls* include gender-specific dummies for number of children 15 and 20 years after high school entry. *STEM Controls* include gender-specific dummies for working in STEM occupations and having completed a STEM college degree and their interactions. Standard errors clustered at the school-cohort level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

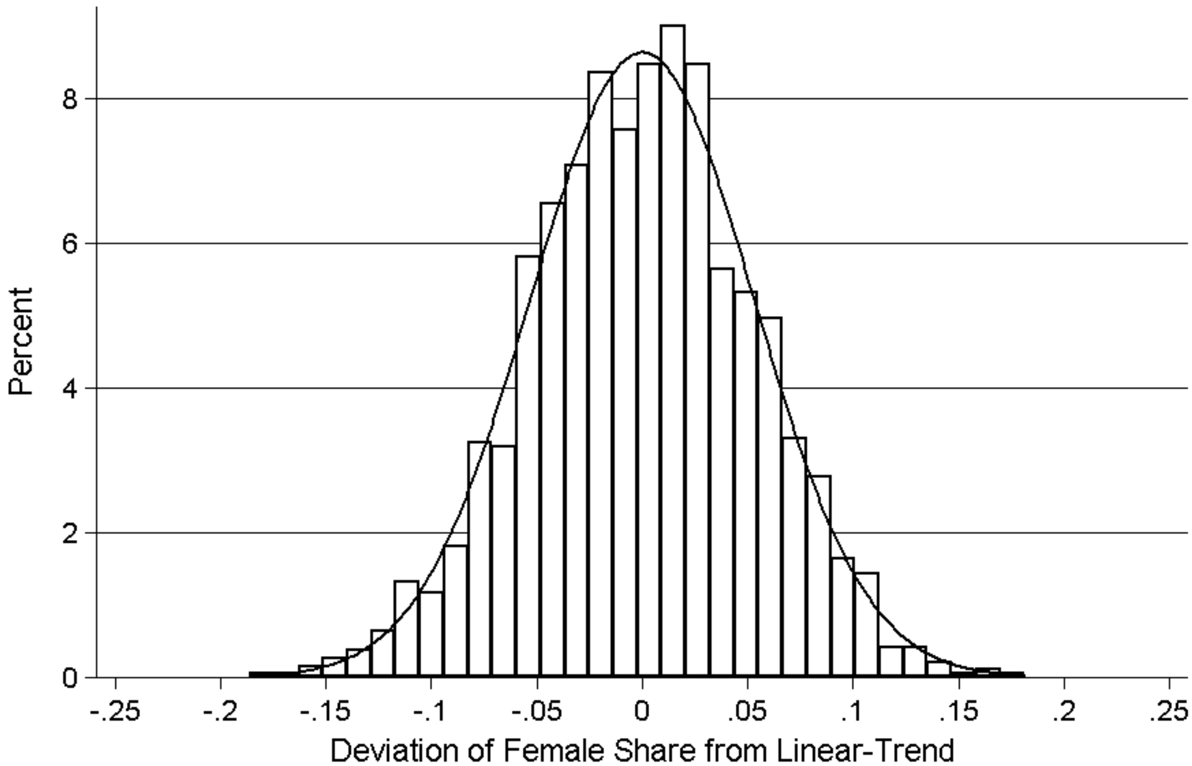
FIGURES

Figure 1: Correlation between Proportion of Female Peers and STEM Degree Completion



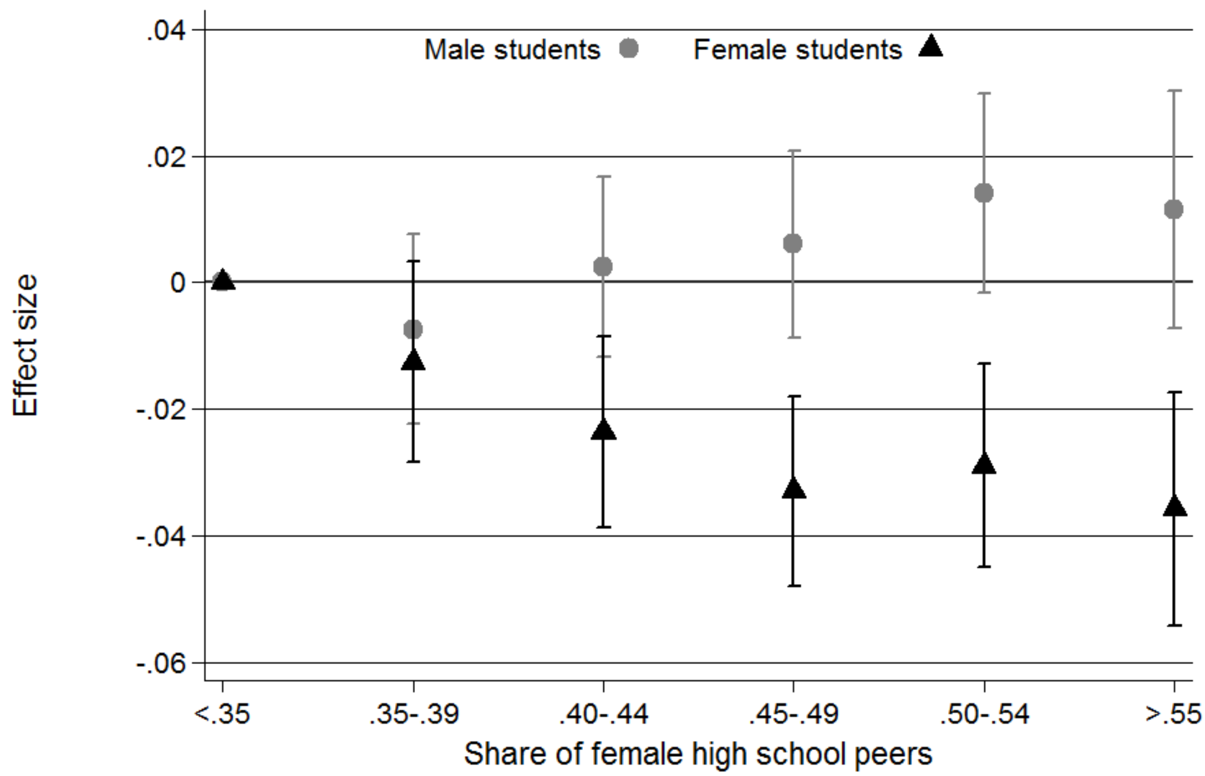
NOTE: The graph shows a bin scatter plot by gender using 30 bins. STEM Degree is measured as highest completed degree at the college level or higher 20 years after high school entry.

Figure 2: Year-to-Year Variation in the Proportion of Female High School Peers Within High Schools



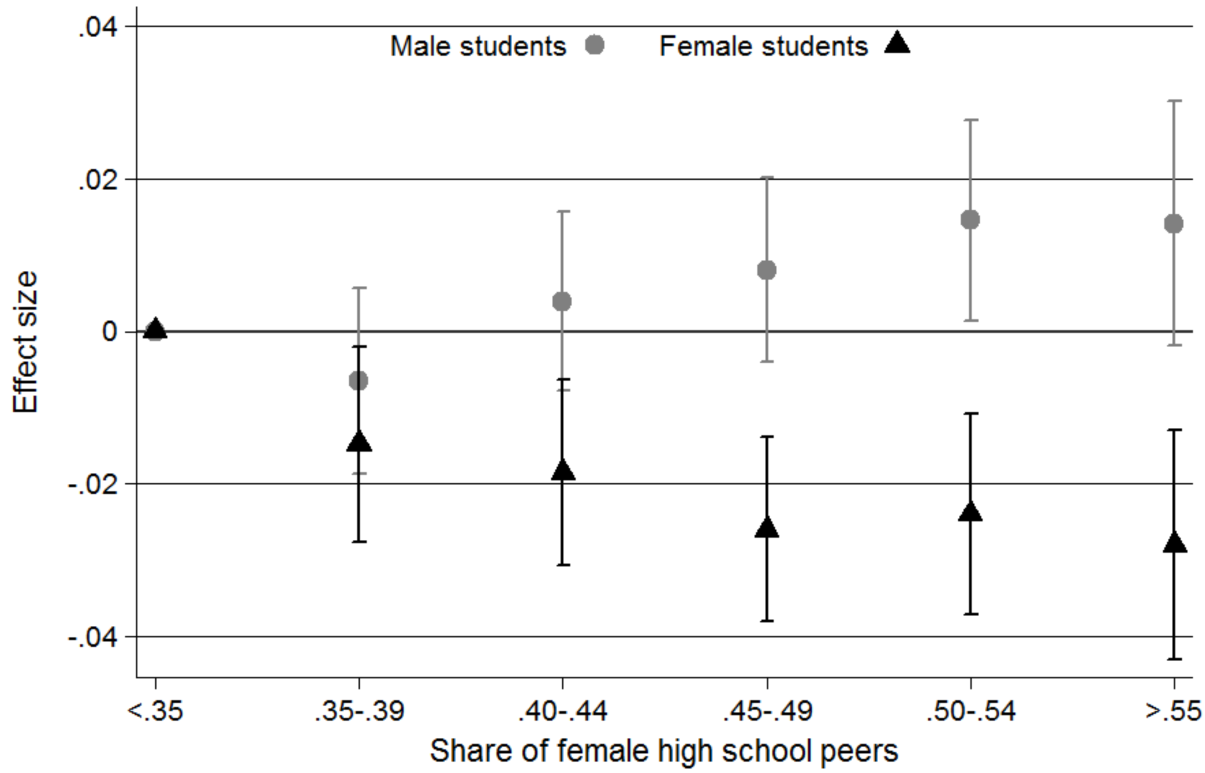
NOTE: This figure illustrates the year-to-year variable in the proportion of female high school peers within high schools, plotted together with the normal distribution. More precisely, it plots the predicted proportion of female peers at the school-cohort level from a regression regressing the proportion of female peers on cohort and school fixed effects and school-specific linear time trends. Each high school-cohort represents one observation.

Figure 3: High School Gender Composition and STEM Enrollment – Non-linear Effects



NOTE: This figure shows point estimates obtained from OLS regression. Instead of including the continuous measure of the proportion female high school peers interacted with the gender dummy, this regression includes five dummies for the high school gender composition interacted with gender. The dependent variable is STEM study enrollment. Vertical lines refer to the 95 percent confidence intervals. The model controls for school-specific linear time trends, cohort fixed effects, school fixed effects, cubed cohort size, indicators for large cohort size changes compared to previous years, as well as a large set of individual and leave-out-mean peer controls shown in Panel B of Table 1.

Figure 4: High School Gender Composition and STEM Graduation – Non-linear Effects



NOTE: This figure shows point estimates obtained from OLS regression. Instead of including the continuous measure of the proportion female high school peers interacted with the gender dummy, this regression includes five dummies for the high school gender composition interacted with gender. The dependent variable is STEM graduation. Vertical lines refer to the 95 percent confidence intervals. The model controls for school-specific linear time trends, cohort fixed effects, school fixed effects, cubed cohort size, indicators for large cohort size changes compared to previous years, as well as a large set of individual and leave-out-mean peer controls shown in Panel B of Table 1.

APPENDIX

Table A1: Complete Balancing Test Including All Individual Level Variables

Based on 190 separate regressions	(1)	(2)	(3)	(4)	(5)
Age at high school entry	-0.001	-0.002	0.002	0.008	0.018
	(0.032)	(0.032)	(0.029)	(0.029)	(0.029)
Mother has less than upper-secondary education	0.024	0.024	0.004	0.004	-0.011
	(0.020)	(0.020)	(0.018)	(0.019)	(0.019)
Mother has upper-secondary education	-0.005	-0.005	-0.006	0.001	0.003
	(0.023)	(0.023)	(0.021)	(0.022)	(0.022)
Mother has tertiary education	-0.020	-0.019	0.005	-0.002	0.013
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Mother education unknown	0.001	0.000	-0.003	-0.003	-0.006
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Father has less than upper-secondary education	0.019	0.017	-0.008	-0.013	-0.015
	(0.017)	(0.017)	(0.016)	(0.016)	(0.017)
Father has upper-secondary education	0.014	0.020	0.016	0.020	0.019
	(0.023)	(0.023)	(0.022)	(0.022)	(0.023)
Father has tertiary education	-0.018	-0.021	0.009	0.013	0.014
	(0.022)	(0.022)	(0.021)	(0.021)	(0.022)
Father education unknown	-0.015	-0.015	-0.017*	-0.019*	-0.018*
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Mother Education field	0.005	0.005	0.013	0.014	0.018
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)
Mother Humanities field	-0.015*	-0.015*	-0.011	-0.014*	-0.010
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Mother Social Sciences field	0.001	-0.000	0.004	0.005	0.006
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Mother Business, Admin, and Law field	-0.033*	-0.033*	-0.028	-0.026	-0.020
	(0.019)	(0.019)	(0.018)	(0.018)	(0.018)
Mother STEM field	-0.003	-0.002	0.002	0.003	0.002
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
Mother Life Sciences field	0.000	0.000	0.000	0.000	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mother Health and Welfare field	0.008	0.011	0.012	0.008	0.007
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Mother Service field	0.001	0.001	-0.000	0.001	0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Mother No field	0.034*	0.033	0.012	0.013	0.001
	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)
Individual level and high school level controls	-	✓	✓	✓	✓
School-specific time trends	-	-	linear	quadratic	cubed

Table A1— continued

	(1)	(2)	(3)	(4)	(5)
Father Education field	0.001	0.001	0.007	0.009	0.007
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Father Humanities field	-0.019***	-0.020***	-0.010	-0.009	-0.008
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Father Social Sciences field	-0.008	-0.009*	-0.008	-0.009*	-0.008
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Father Business, Admin, and Law field	0.010	0.012	0.022	0.026*	0.023
	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)
Father STEM field	0.016	0.017	0.011	0.005	0.008
	(0.021)	(0.021)	(0.020)	(0.020)	(0.021)
Father Life Sciences field	0.004	0.005	0.007	0.014**	0.013*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Father Health and Welfare field	-0.008	-0.008	-0.004	-0.004	-0.002
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Father Service field	0.005	0.005	0.004	0.001	-0.003
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Father No field	0.015	0.015	-0.014	-0.017	-0.017
	(0.017)	(0.018)	(0.016)	(0.017)	(0.017)
Child is adopted	0.004	0.004	0.003	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Lives with both parents at age 10	0.006	0.010	0.010	0.006	0.005
	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)
First-generation immigrant	0.001	0.000	-0.004	-0.001	-0.001
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
Second-generation immigrant	0.004	0.003	-0.004	-0.003	-0.004
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Firstborn	-0.039*	-0.041*	-0.038*	-0.024	-0.009
	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)
Number of siblings	0.007	-0.004	-0.028	-0.038	-0.059
	(0.039)	(0.039)	(0.038)	(0.039)	(0.041)
Number of siblings squared	-0.029	-0.100	-0.345	-0.402	-0.569**
	(0.267)	(0.266)	(0.257)	(0.265)	(0.282)
Mother's age at birth	0.267	0.308	0.274	0.063	-0.035
	(0.265)	(0.265)	(0.250)	(0.248)	(0.253)
Mother's age at birth squared	15.921	17.982	14.983	7.406	1.763
	(13.486)	(13.493)	(12.545)	(12.380)	(12.625)
Mother <22 years at birth	-0.013	-0.015	-0.011	-0.001	0.006
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Mother's age unknown	0.003	0.003	0.001	0.005	0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
School level controls	-	✓	✓	✓	✓
School-specific time trends	-	-	linear	quadratic	cubed

NOTE: Each cell in this table is estimated with a separate regression including including school and cohort fixed effects. The dependent variable in each cell is the proportion of female high school peers. *School level controls* included in Columns (2)-(5) are an indicator if any student in the cohort is older than 20 years at high school entry, dummies for number of students without information on gender (ranging from 0 to 2), indicators for large changes in cohort size compared to previous years, the female share in the language track, an indicator if the high school has no language track, indicators for exposure to experiment on course curriculum, and cubed cohort size. Standard errors in parentheses are clustered at the school-cohort level. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A2: The Impact of Peer Gender on High School Graduation, College Enrollment, and Higher Education Degree Completion

	(1)	(2)	(3)
	Completed academic high school	Ever enrolled in higher education	Completed higher education degree
Female * Proportion female peers	0.017	-0.010	0.000
	(0.017)	(0.024)	(0.027)
Male * Proportion female peers	0.017	-0.030	-0.035
	(0.017)	(0.022)	(0.026)
Female	0.024**	0.005	0.045***
	(0.010)	(0.014)	(0.016)
N	182211	182211	182211
Mean	0.909	0.789	0.677
p-values of test for gender equality of "Proportion female peers"	0.992	0.523	0.315

NOTE: The dependent variable in Column (1) is equal to one if the student completed academic high school within 5 years after high school entry. The dependent variable in Column (2) is equal to one if the student ever enrolled in college studies and the dependent variable in Column (3) is equal to one if the student ever completed any college education. All models control for school-specific time trends, cohort fixed effects, school fixed effects, cubed cohort size, indicators for large cohort size changes compared to previous years as well as a large set of individual and leave-out-mean peer controls shown in Panel B of Table 1. Standard errors clustered at the school-cohort level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Robustness Check—Main Results by Number of Schools in the Municipality

Subgroup	Full sample	Municipality with only 1 high schools	Municipality with 2+ high schools
Dependent variable:	STEM completion	STEM completion	STEM completion
	(1)	(2)	(3)
Female * Proportion female peers	-0.097***	-0.088***	-0.097***
	(0.022)	(0.030)	(0.032)
Male * Proportion female peers	0.092***	0.107***	0.064**
	(0.023)	(0.032)	(0.033)
Female	-0.033**	-0.036*	-0.037**
	(0.013)	(0.019)	(0.019)
Observations	182211	99599	82612
Mean	0.200	0.200	0.199
p-values of test for gender equality of "Proportion female peers"	<.0001	0.001	0.046

NOTE: All models control for school-specific time trends, cohort fixed effects, school fixed effects, cubed cohort size, indicators for large cohort size changes compared to previous years as well as a large set of individual and leave-out-mean peer controls shown in Panel B of Table 1. Standard errors clustered at the school-cohort level are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.