The long-run effects of peer gender on occupational sorting and the wage gap

Demid Getik[‡] Armando N. Meier[⋆]

August, 2024

Abstract

We study the impact of the early gender environment on inequality in the labor market. To this end, we link primary school data to occupations and earnings. We find that women exposed to more girls at critical ages earn more later on: A 10% increase in the share of girls leads to a reduction in the gender wage gap of 2.7%. We explore mechanisms and find a strong selection of women into less gender-stereotypical educational tracks and occupations, leading to higher earnings. The gender environment at an early age, therefore, leads to persistent changes in career trajectories and earnings.

JEL Classifications: I24, I26, J16, J24, J31, J70

Keywords: peers, school environment, gender, occupational sorting, wage gap

[‡] Demid Getik, Durham University and Center for Economic Demography at Lund University (CED), email: demid.getik@durham.ac.uk
[⋆] Armando N. Meier, University of Basel, email: armando.meier@unibas.ch

We thank three anonymous referees and the editors for detailed guidance. We also thank numerous scholars and seminar participants at the University College Dublin, Lund University, University of Basel, University of Hamburg, University of Bergen, Johannes Kepler University of Linz, Copenhagen Business School, as well as conference participants at the Copenhagen Education Workshop, the IZA Economics of Education Workshop, and the Workshop for Societal and Policy Problems for their comments.

1 Introduction

Women earn less than men, with more than 70% of the gender-wage gap not accounted for by traditional explanations such as educational attainment (Card, Cardoso and Kline, 2016; Goldin et al., 2017; Blau and Kahn, 2017). The persistent gap has spurred research on other potential explanations for inequalities in labor market outcomes. Recent literature proposes early-life gender socialization and related identity formation as one important explanation for the remaining inequality in occupational selection and wages (Olivetti and Petrongolo, 2016; Sloane, Hurst and Black, 2021). While a growing literature indicates that gender environments can shape identities and educational sorting (e.g., Schneeweis and Zweimüller, 2012; Black, Devereux and Salvanes, 2013; Brenøe and Zölitz, 2020), little is known about the long-run effects on labor market outcomes. One reason is a lack of data linking key features of early-life environments, such as the gender environment in primary school, to long-run labor market outcomes.

In this paper, we bridge this gap by studying whether and how a female-dominated environment at critical ages 6 to 16 shapes career trajectories in the long-run. We overcome the challenge of linking the early gender environment to labor market outcomes by using comprehensive Swedish register data ($N = 757,560$). We link the share of female peers in a student's primary school-cohort to grades, high-school tracks, college subjects, occupations, and earnings.

We find that more girls in primary school lead to higher earnings for women, lowering the gender wage gap. Changing from a 45% female to a 55% female cohort, corresponding to two more girls per classroom, leads to a $354 (SEK 1,717) increase in annual earnings of women at the age of 30.[1] This is equivalent to a 2.7% reduction in the gender wage gap. In the longer run, predicted lifetimes earnings are roughly $12,390 (SEK 60,236) higher for girls exposed to 10% more girls.[2] The effect on women's earnings is meaningful when compared to teacher value-added estimates.

Chetty et al. (2011) estimate that kindergarten students exposed to a teacher for 4 years who had more than 10 years of experience earn $524 more per year at ages 25 to 27. In com-

[1] We apply the 2009 exchange rate (1:7) and use inflation adjustments to compute 2024 dollar values throughout the paper. The year 2009 lies in the middle of the period when we measure income and corresponds to the dollar values used in (Chetty et al., 2011) to which we compare our estimates to.

[2] The data used in this paper is based on annual labor income which we refer to as earnings, income or wages.

parison, we find that a 10 percentage point shift in the gender composition of the cohort over around 7 years changes annual earnings by \$354 at age 30. The effects can also be compared to the association between parental education and children's earnings: A 10 percentage point shift in the gender composition amounts to almost 10% of this relationship.

We also examine the impact of a more female-dominated environment on occupational earnings potential only relying on variation across occupations. We find that women exposed to more girls in primary school enter occupations with higher earnings. A 55% female-dominated class, when compared to a 45% female class, leads to \$7,949 higher predicted occupational lifetime earnings and a reduction of the occupational gender earnings gap by 3.2%. A key reason for these findings is that women sort into less gender-stereotypical occupations after being exposed to more girls in primary school. One reason for the change in occupational selection is that girls select into less gender-stereotypical high-school specializations and college majors. Girls with more girls in the cohort initially choose more boy-dominated high-school tracks and then choose occupations dominated by men, leading to higher earnings.

To identify the causal impact of gender environment in primary school, we use idiosyncratic variation in the gender composition across cohorts within schools (following, e.g., Hoxby, 2000; Lavy and Schlosser, 2011; Carrell, Hoekstra and Kuka, 2018). The extensive administrative data allows us to check in detail whether the share of girls in a cohort is arbitrary. First, we observe that cohorts with more boys or girls are comparable based on an extensive range of parental characteristics, including parental age, educational attainment, and earnings. Second, we document that whether a girl or a boy ends up in a specific cohort is arbitrary. For instance, student gender does not predict the leave-one-out share of female peers (following Guryan, Kroft and Notowidigdo, 2009). Last, we also find that the differential effect on boys and girls are identical when including school–by–cohort fixed effects, holding constant any potential dynamic selection into specific school-cohorts (Brenøe and Zölitz, 2020). These and additional checks indicate that the gender composition in school-cohorts is largely arbitrary.

The paper offers a comprehensive evaluation of the impact of early gender environment on labor market outcomes, contributing to two broad strands of literature. First, we offer evidence for how gender socialization affects the gender wage gap. We therefore add to a large literature on occupational sorting and the gender pay gap (see, e.g., Altonji and Blank, 1999; Mulligan and Rubinstein, 2008; Olivetti and Petrongolo, 2016; Lundborg, Plug and Rasmussen, 2017; Kleven, Landais and Søgaard, 2019; Card, Colella and Lalive, 2021).

Previous literature has examined how educational attainment, fertility, occupational sorting, and discrimination affect gender gaps (Lundborg, Plug and Rasmussen, 2017; Kleven, Landais and Søgaard, 2019; Card, Cardoso and Kline, 2016), but has given less consideration to the role of early social environments and socialization:[3] "... much more validating empirical work will be needed in the near future for gender identity insights to have a long-lasting impact on how labor economists approach gender issues (Bertrand, 2011; p. 1545)."

The gender environment at school can shape gender identities and behavior by affecting cooperation and disruption (Lavy and Schlosser, 2011; Jackson, 2021), changing preferences (Schneeweis and Zweimüller, 2012), or motivation (Briole, 2021). Findings from psychological surveys and evidence on educational sorting indicate that girls in female-dominated environments are more self-confident, for instance, regarding their math skills, and less gender-conforming (Kessels and Hannover, 2008; Sullivan, 2009; Bertrand, 2011; Schneeweis and Zweimüller, 2012; Eisenkopf et al., 2015). This is consistent with our results, which show that women choose *less* gender-stereotypical jobs after exposure to more girls in primary school, suggesting that the early gender environment shapes gender roles in the long-run.

While women attain a higher level of education in high-income countries, they keep sorting into different college majors (Sloane, Hurst and Black, 2021) and the earnings gap persists (Olivetti and Petrongolo, 2016; Blau and Kahn, 2017). Our results suggest that, even beyond grades, socialization and resulting educational and occupational sorting are a key driver of the gender wage gap. Therefore persisting gender roles could play a critical part in explaining why the rise in educational attainment has not led to a commensurate reduction in gender inequality.

Second, we complement the literature on how school inputs and childhood environments affect later labor market outcomes (see, e.g., Chetty et al., 2011; Kirabo Jackson, Johnson and Persico, 2016; Carrell, Hoekstra and Kuka, 2018; Balestra, Eugster and Liebert, 2022; Bietenbeck, 2020; Fischer et al., 2020; Zölitz and Feld, 2021; Golsteyn, Non and Zölitz, 2021; Elsner, Isphording and Zölitz, 2021). Among others, previous research has examined the role of neighborhoods or the role of parents and siblings in shaping labor market outcomes and

[3] One exception is Slotwinski and Stutzer (2023) who show that voting rights increased labor force participation among women.

gender-conformity (see, e.g., Chetty et al., 2016; Almås et al., 2016; Brenøe and Lundberg, 2018; Brenøe, 2022).

Research examining the consequences of the gender environment in school has focused on educational sorting and outcomes (see, e.g., Lavy and Schlosser, 2011; Park, Behrman and Choi, 2012; Schneeweis and Zweimüller, 2012; Eisenkopf et al., 2015; Lu and Anderson, 2015; Anelli and Peri, 2019; Giardili, 2020; Brenøe and Zölitz, 2020; Borbely, Norris and Romiti, 2023). We discuss this literature in detail in Section 2. The literature generally shows that girls exposed to more girls do better at school, and choose less gender-stereotypical fields, such as science tracks, more often. For boys, the results are mixed, with the modal paper showing no statistically significant effect for boys. Our results on educational outcomes for both genders are consistent with these results.

While there is a growing literature on the effects of the gender environment on educational attainment and sorting, it is unclear whether and how effects on educational outcomes translate to long-run labor market outcomes. The three existing papers considering labor market outcomes offer a limited assessment of earnings and occupational sorting, and present conflicting results: Black, Devereux and Salvanes (2013) find positive effects on earnings for girls, Anelli and Peri (2019) do not find statistically significant effects on earnings, and Brenøe and Zölitz (2020) find negative effects on earnings percentiles for girls. We discuss the corresponding results and potential reasons for the discrepancies in Section 2.

We make three main contributions: First, we comprehensively assess effects on earnings and occupational sorting including on the gender wage gap, gender-stereotypical sorting, and long-run occupational earnings. Second, we can follow students from primary school through all educational stages and onto the labor market, allowing us to directly examine underlying mechanisms linking early educational attainment and sorting to labor market outcomes. To probe the importance of early formation of gender roles for labor market inequalities, we can study whether educational sorting into more or less gender stereotypical educational tracks translates to occupational sorting and earnings. Third, we examine students at earlier ages than previous research, starting from age 6 in compulsory school, rather than from age 14 in lower secondary school or after selection into competitive tracks occurred. This means we study the effects when children are at their most malleable (Bertrand, 2011) and at a point in time where no important sorting decisions have yet been made (Brown and Corcoran, 1997), reducing the risk of bias and increasing external validity.

In sum, we build on and contribute to previous literature by providing a comprehensive and unified evaluation of the impact of the gender environment at an early, malleable age on long-run labor market outcomes including occupational sorting. Our results demonstrate that even small differences in the early gender environment have long-run consequences for career and earnings trajectories.

2 The Literature on Peer Gender at School

Samples and Identification Strategies. Table 1 provides an overview of economics papers focusing on the effects of the share of girls at school on educational and labor market outcomes.[4] Papers cover the age ranges from 5 to 19, with most papers studying ages 12 to 18, and 11 of 17 papers exclusively using data from compulsory school.[5] Focusing on high-income countries, the papers use data from the US (4), Austria, France, Italy, Sweden, Denmark, Norway (2), Switzerland, Israel, China (3), and Ethiopia. The corresponding sample sizes range from 532 to the 757'760 of our study, with 4 papers analyzing more than 100'000 observations. The identification strategies are very similar across studies with 10 papers using cohort variation as we do (all of which include school fixed effects). Six papers use class variation and one paper uses desk assignments within the classroom. Overall, the identification strategies are very similar, relying predominantly on natural variation in composition, but the papers differ in the stages of schooling as well as educational tracks, ages, and countries analyzed.

[4] We only consider economics papers on coeducational schooling, mentioning gender in some form in the abstract, and mentioning either educational sorting, attainment, or labor market outcomes in the abstract.

[5] Because of the large age difference and strong selection when students attend college, we do not think the peer effect estimates for college students are straightforward to compare to the estimates in the summarized literature here. Papers looking at peer gender during college include Oosterbeek and Van Ewijk (2014); Hill (2017); Booth, Cardona-Sosa and Nolen (2018); Feld and Zölitz (2018); Shan (2021); Calkins et al. (2023), which find, if anything, positive impacts of more women for both genders.

Table 1: Overview of the Literature on Effects of Peer Gender at School

Study	*Location & cohorts*		*Schools & age*			*Data*			*Identification*		*Effects on*		*Findings*		
authors (year)	**region**	**cohorts**	**school level**	**type**	**age**	***N***	**share girls**	**regressor**	**variation**	**level**	**girls**	**boys**	**educational attainment**	**educational sorting**	**earnings**
Hoxby (2020)	USA / Texas	1990-1998	elementary school (grades 3 to 6)	compulsory	8 to 10	66'297	49%	share girls	natural	cohort	positive	positive	improved reading and math scores for boys and girls		
Whitmore (2005)	USA / Tennessee	1985-1986	kindergarden	compulsory	5 to 6	11'600	49%	share girls	quasi-random	class	positive	negative	better test scores for girls, lower test scores for boys in grade 3		
Lavy and Schlosser (2011)	Israel	1993-2000	elementary to high-school	compulsory and not compulsory	6 to 16	461'217	50%	share girls	natural	cohort	positive	positive	improved exam scores and scores in math, science, hebrew, and english of boys and girls, lower disruption and violence		
Schneeweiss and Zweimüller (2012)	Austria / Linz	1988-2006	low track secondary school (non-academic track), 58% of students	compulsory	10 to 14	7'472	45%	share girls	natural	cohort	positive	not stat. sign.		women choose more male-dominated high-school tracks	
Black, Devereux, and Salvanes (2013)	Norway / all	1973-1987	lower secondary schools	compulsory	14 to 16	417'780	49%	share girls	natural	cohort	positive	negative	negative impact on boys (lower education), no impact on women	lower academic track share for boys	higher labor force participation and log earnings for women
Hill (2015)	USA	1994-1995	high-school (grades 7-12)	compulsory and not compulsory	12 to 17	8'435	51%	share of opposite gender school-cohort friends	natural	cohort	positive	negative	GPA (particularly for girls) worse for kids with more opposite gender school cohort-friends and older than 16, no impact on college / high-school graduation		
Lu and Anderson (2015)	China / Jiangsu	2009	middle-school (grade 7)	compulsory	12 to 13	532	43%	indicator for whether girl is desk mate	random	seat	positive	not stat. sign.	increase in composite test scores for girls, no large impact on boys		
Hu (2015)	China	2013-2014	middle school (7th and 9th grade)	compulsory	12 to 13 and 14 to 15	9'565	48%	share girls	quasi-random	class	not stat. sign.	positive	improves math, Chinese, and English scores for boys		
Eren (2017)	USA / disadvantaged areas	2010-2012	high school	compulsory and not compulsory	14 to 17	3170	53%	share girls	random	class	positive	not stat. sign.	increases math test scores for girls		
Anelli and Peri (2019)	Italy / Milan	1979-1999	public college-preparatory high schools with two tracks (natural sciences with 60% men), excludes more common lower tracks	not compulsory	14 to 19	29'370	52% (33% vs. 60%)	share girls / indicator less than 20% girls	quasi-random	class	not stat. sign	not. stat. sign / negative	linear: none / <20% girls: none	linear: no impacts / <20% girls: +6-15 pp increase in initial choice of engineering & econ / business for men	linear: none / <20% girls: none
Schøne (2019)	Norway / all	2000-2005	lower secondary schools (grades 8-10)	compulsory	14 to 16	317'511	49%	share girls	natural	cohort	positive	unclear	worse language grades for boys and for girls, better math grades for girls	STEM choices in upper secondary school increase for both, boys and girls, choosing vocational track increases for boys	
Brenøe and Zölitz (2020)	Denmark / all	1980-1994	math track of high-school	not compulsory	15 to 19	81'820	45%	share girls	natural	cohort	negative	positive	no impact on women, men do better	lower enrollment of women in STEM, opposite for men	lower earnings percentile for women
Mouganie and Wang (2020)	China / city in Southern China	2003-2006	high-school, competitive educational track (60% admitted)	not compulsory	15 to 18	133'845	52%	share girls / share of high-performing girls	natural	cohort	not stat. sign / positive	not stat. sign.	linear: none / girls exposed to higher performing girls more likely to attain better universities	girls exposed to higher performing girls more likely to choose science high-school and college track	
Borbely, Norris, and Romiti (2021)	Ethiopia	2012-2013	primary school (grades 4-5)	not compulsory	7 to >12	5'077	51%	share girls	quasi-random	class	positive	not stat. sign.	less school absence, better math grades, motivation, and participation for girls, but not for boys		
Balestra, Sallin, and Wolter (2021)	Switzerland / St. Gallen	2008-2017	high-school (grade 8), better accomplished track	compulsory	13 to 15	31'625	52%	share girls / share of gifted girls	natural	class	positive	positive	improved scores in math and languages for girls and boys		
Briole (2021)	France	2008-2011	public middle schools (9th grade)	compulsory	14 to 15	81'642	51%	share girls	natural	cohort	positive	negative	increases test scores, high-school graduation, reduces dropout for girls, but decreases high-school graduation for boys	increases science track in high-school for girls	
Getik and Meier (2024)	Sweden	1979-1992	primary school	compulsory	6 to 16	757'760	49%	share girls	natural	cohort	positive	not stat. sign.	girls have better grades and enroll more in high-school, boys have worse grades	women choose more male dominated high-school and college tracks	higher earnings for women, lower gender wage gap

Notes: Summary of economics papers looking at the impacts of peer gender at school in coeducational settings, focusing on educational and labor market outcomes. Included papers mention gender in the abstract and have the share of girls or closely-related function of it as a key independent variable / regressor and shown separately in a table. Coho cohort started the first grade the paper considers. Age refers to the age students have in the studied grades. *N* refers to the largest number of observations available for educational outcomes. Quasi-random variation refers to settings where the variation has a more clearly random component than the settings with natural variation, e.g., when classes w almost randomly. Impact on girls and impact on boys is indicated as positive when students do better in school, choose more technical educational tracks, or have higher earnings.

Effects on Girls. To simplify the overview, we categorize papers that find that girls have better grades, choose more technical classes, or have better labor market outcomes after exposure to female peers as positive for girls. Out of 17 papers, 13 find positive effects for girls, 3 papers show no statistically significant effects, and 1 shows negative effects for girls. Accordingly, there is a broad consensus that girls previously exposed to more girls have higher educational attainment and are more likely to choose technical tracks.

Why do 4 papers show inconclusive or negative effects (Hu, 2015; Anelli and Peri, 2019; Brenøe and Zölitz, 2020; Mouganie and Wang, 2020)? Out of the 4 papers, 3 stand-out with characteristics distinct from all other papers (Anelli and Peri, 2019; Brenøe and Zölitz, 2020; Mouganie and Wang, 2020): i) they focus on non-compulsory schooling, ii) use samples from competitive or otherwise selective high-schools (math track, college preparatory high-school, and educational track), and iii) use data from the oldest age groups across all papers (ages 14 to 19). Two of the papers, Anelli and Peri (2019) and Brenøe and Zölitz (2020), also have a lower share of girls than usual driving their results: Anelli and Peri (2019) have one part of their sample with 40% share of girls, whereas Brenøe and Zölitz (2020) have a relatively low share of girls overall with 45%.

Of all other 14 papers without these characteristics, Hu (2015) is the only paper that finds no positive effect of exposure to girls at younger ages and during compulsory school. Some of the other papers also have students who are not in compulsory school, but they either examine much younger students or examine a school system with no tracking (Lavy and Schlosser, 2011; Hill, 2015; Eren, 2017; Borbely, Norris and Romiti, 2023). The only other paper studying a somewhat selective track considers a sample in compulsory school with a younger age range of 13 to 15 (Balestra, Sallin and Wolter, 2021). In all of the cases, the papers document positive effects of female peers on girls.

The comparison suggests that the impact of female peers is much less clear–and may even be negative (Brenøe and Zölitz, 2020)–when students are older and attend selective, non-compulsory school tracks. Likely reasons for this pattern include i) that older children are simply less influenced by the gender composition of their peers, ii) that girls who already selected into science / competitive tracks may be relatively less influenced, e.g., in terms of grades, information, and preferences, by fellow girls as they already prefer the science tracks and have better grades, iii) that these tracks are more dominated by boys to begin with, potentially leading to fewer same-gender role models (Balestra, Sallin and Wolter, 2021), smaller networks for girls as our results and the results by Anelli and Peri (2019) suggest, and to a different shape of the peer distribution, iv) teachers reacting less favorably to more girls in male-dominated tracks (Jackson, 2021), and v) differences in student interactions as students age, e.g., leading to less disruption.

In summary, the effects of female peers at school are clearly positive for girls at younger ages during compulsory school, but are more ambiguous for girls in selective, male-dominated, tracks preparing for college. Yet, how exactly the effects on educational sorting and attainment at younger ages translate to occupational sorting and the gender wage gap remains unanswered.

Effects on Boys. The impact of the share of girls on boys is much less clear. Across 17 papers, 5 show positive effects, 7 have not statistically significant results, 4 show negative effects and 1 paper shows positive effects in some, but negative effects in other dimensions. We could not find any systematic reasons for indicating why some papers find positive and some find negative effects.

We can only speculate about potential reasons which could include non-linearities (Anelli and Peri, 2019), impacts on mental health differing across cultures and school-systems (Getik and Meier, 2022), teachers reacting differently to different class compositions across contexts etc. Our results suggest that boys do best with a low to intermediate share of girls and it could be that some heterogeneity in results comes from different variances in the distribution of peers, with a higher variance possibly leading to worse outcomes for boys. The results from the current data also suggest that boys suffer in terms of network size, which may affect mental health (Getik and Meier, 2022) as well as access to information and role models. Given that there are stark differences across papers even from the same context (e.g., for the US and China), more research is needed to identify mechanisms for boys across different ages.

In conclusion, the effects on girls are likely similar in similar contexts, positive for younger age groups in compulsory schooling, but less strong for older age groups. The mixed evidence for boys, however, suggests that the nature of peer effects may differ even within very similar settings. We provide a more detailed discussion of potential mechanisms at play in our and other papers in Section 7.[6]

Differences and Contributions. Previous papers examining peer effects in focus on educational attainment and sorting. We know less about how those effects translate to long-run labor market outcomes.

Of a total of 17 previous papers, only 3 examine labor market outcomes (Black, Devereux and Salvanes, 2013; Anelli and Peri, 2019; Brenøe and Zölitz, 2020), focusing on 1 or 2 specific outcomes in each paper such as sorting into a STEM occupation. Black, Devereux and Salvanes (2013) find that women exposed to more girls in compulsory school have a higher labor force participation and higher log-earnings. The impact on boys is not statistically significant. Anelli and Peri (2019) find no statistically significant labor market impacts for boys or girls of being exposed to more girls in selective high-schools in Milan. In contrast, Brenøe and Zölitz (2020) find that women exposed to more girls have a lower earnings percentile and are less likely to sort into STEM occupations.

Our paper differs in three unique ways from these papers: First, we assess the impacts on labor market outcomes, including the gender wage gap, and the educational trajectories there in in more detail than previous papers which studied one or two labor market outcomes. In addition to earnings and employment specified in different ways, we document effects on occupational earnings potential, which is key to better understand the impact on lifetime earnings.

Second, for the first time in this literature, we consider whether people choose gender stereotypical jobs directly, allowing us to probe the relevance of gender roles in explaining sorting into higher wage jobs. We observe that across all educational stages and onto the labor markets women sort into less gender stereotypical tracks. These results offer a distinct and novel explanation for gender inequalities beyond educational attainment. The results show that educational sorting influenced by peers early in life partly explains labor market inequalities.

[6]There is also a sizable literature on the educational impacts of single-sex schooling, see, e.g., Park, Behrman and Choi (2012); Jackson (2012); Lee et al. (2014); Eisenkopf et al. (2015); Park, Behrman and Choi (2018); Jackson (2021). This literature is not directly comparable, as none of the papers has gender makeups that come very close to single-sex schooling. While there is also some heterogeneity in results for the literature on the effects of single-sex schooling on educational outcomes, many papers show positive effects across boys and girls.

Third, we look at younger individuals with exposure starting at age 6, whereas the other papers look at ages 14 (Black, Devereux and Salvanes, 2013; Anelli and Peri, 2019) and 15 (Brenøe and Zölitz, 2020). On average, we observe students 5 years earlier than Black, Devereux and Salvanes (2013) and Anelli and Peri (2019), and 6 years earlier than Brenøe and Zölitz (2020). This is an important difference as we observe exposure to peers in the critical time of early puberty and when students first enter school.

Previous research leaves the question unanswered of the effects of peer gender in primary school on occupational sorting and the gender wage gap. It also offers an incomplete picture of how educational sorting and attainment translate to labor market outcomes, which we can show with our data.

Future Avenues for Research. An obvious drawback of the literature is that it focuses mostly on high-income countries, with only 2 papers outside of the US, Europe, or China. Another drawback is that, despite the many papers examining educational attainment and outcomes, there is little evidence on the impact of peer gender early in school on later life outcomes including labor market outcomes with our paper as one exception. Apart from extending the coverage regarding age and countries, a more in-depth study of the peer effects for boys and corresponding mechanisms, for instance via noneducational channels such as crime, teachers, and social networks (Lavy and Schlosser, 2011; Hill, 2015; Jackson, 2021), is a first order concern.

3 Institutional Background and Data

3.1 Swedish Educational System

All Swedish children have to complete nine years of compulsory primary schooling (grundskolan), typically from ages 6 to 16, with a standardized national curriculum.[7] Admissions

[7]Since the Swedish term "grundskolan" describes a continuous period of primary schooling which is also compulsory in the country, we use the terms "primary" and "compulsory" interchangeably in the remainder of the paper.

to school is based on residence, with municipal schools responsible for accommodating all students in a given municipality and restricted to admission based on residential proximity.[8]

After starting school in the year when they turn seven, they usually pass through three stages of primary school: grades 1-3 (low), grades 4-6 (middle), and grades 7-9 (high).[9] There is no institutional reshuffling of cohorts after grade 7 (around age 13).[10] The exact grade configuration of schools can vary: During our study period, 50%-60% of schools provided all 9 grades of compulsory education (Holmlund and Böhlmark, 2019). In 1995, close to the modal cohort we observe, around 55% of schools had all grades and a total of around 65% had more grade levels than grades 7 to 9 (Holmlund and Böhlmark, 2019).

The institutional context leads to a likely arbitrary share of female peers in one's school-cohort as the place of residence and the age determine the assignment to specific school-cohorts. Nearly all students in our sample comply with the requirement and enter school in the year they turn seven.[11] Previous research using Swedish data shows that for a different time window, classroom and cohort gender composition are indeed orthogonal to parental and student characteristics (Getik and Meier, 2022). A battery of balance and robustness checks detailed below corroborate those findings, suggesting that gender composition across school-cohorts is largely arbitrary.

In the last year of primary school, students can apply for high-school admission within their municipality. Just under 90% of students in our sample complete high-school, which takes three years. There are currently 18 high-school programs that students in Sweden can choose from: 6 academic, and 12 vocationally oriented.[12] Graduation from an academic

[8] Most municipalities set a maximal acceptable travel distance. Some municipalities apply a measure known as relative distance (relativ närhet). This metric involves comparing the relative distance between a school and the next best alternative across students. See, for example, the explanation by the schooling authority for Stockholm. In the 1990s, charter schools (friskolor), which could apply additional criteria for admission, were introduced into the Swedish system. However, we only use data from schools we consistently observe from 1989 to 2002. Accordingly, we do not have charter schools in our sample.

[9] Note that, at the beginning of each of these stages, students are assigned to classes in which they normally remain for the duration of the stage.

[10] Within stages, the cohort composition remains very stable. Reasons for changes would include grade repetition, which is very rare in the Swedish primary-school context. OECD estimates suggest that only around 3.2% of students in Sweden repeat a grade at some point.

[11] Approximately 96% of students in our sample complete primary school at the age of 16, strictly 9 years after they are supposed to start, complying with the existing rules.

[12] The number and the list of available programs has changed over time. The academic programs available to students during the period covered by our data are: Natural Sciences; Social Sciences; Humanities; Arts; Healthcare; The Industrial Program. The vocationally oriented programs are: The Food Program; The Handicraft Program; Natural Resources; The Construction Program; Children Recreation; Electrical Engineering;

program provides the necessary basic qualification to enter a university, with the Natural Sciences track allowing for admission into the widest range of university programs. In the vocational programs, there is a possibility to fulfil extra requirements to attend university if the student chooses a sufficient load of academic courses. Students can then apply for specific programs at a university of their choice. Admission is based on high school grades or national test scores.

3.2 Data and Descriptive Statistics

Sample. We use administrative data on primary school cohort composition for the years 1989 to 2002, linked to labor-market outcomes at the age of 30 for the years 2003 to 2016. Data restrictions that we face come from two main sources. First, the data on primary school cohort composition starts in 1989. Second, we want to observe people up to age 30, or 14 years after completing compulsory schooling. Income data is available until 2016, which means we can use schooling data up to 2002.

The sample includes all Swedish primary school students who completed that stage of education between the years of 1989 and 2002 and whose school we observe consistently for the entire duration.[13] From this sample, we remove students for whom information about both parents is missing. Finally, following Brenøe and Zölitz (2020), we exclude cohorts with fewer than 10 students (around 1% of the sample). Those are excluded primarily to avoid conflation with non-regular education, such as evening or other special classes. The final dataset consists of 752,560 primary school students from 539 schools across Sweden.

Calculating the Share of Female Peers. To calculate the share of female peers in a school–cohort, we use data from the Primary School Outcomes Register (Registret över ämnesprov i årskurs 9). The data for each student is recorded once upon completion of

Vehicle Engineering; The Energy Program; Business Administration, The Hotel, Restaurant and Catering Program; and Media.

[13]Using a sample including also schools which we do not observe for all years yields similar results with the coefficient for female x share females being SEK 11,862 (*se* =3,861) and the gap between the impact on males and females being SEK 17,600 (*se* =5,472).

Table 2: Summary Statistics

Variable	N	Mean	SD	N	Mean	SD
Background Variables		**Men**			**Women**	
Mother High School	385,574	0.69	0.46	366,986	0.69	0.46
Father High School	385,574	0.71	0.45	366,986	0.71	0.45
Mother Vocational Degree	385,574	0.14	0.35	366,986	0.14	0.35
Father Vocational Degree	385,574	0.11	0.32	366,986	0.11	0.32
Mother College Degree	385,574	0.19	0.39	366,986	0.19	0.39
Father College Degree	385,574	0.19	0.40	366,986	0.19	0.40
Mother STEM Degree	385,574	0.01	0.09	366,986	0.01	0.09
Father STEM Degree	385,574	0.04	0.20	366,986	0.04	0.20
Log Family Income	385,574	11.95	2.79	366,986	11.97	2.76
Wage Mother (1000 SEK)	381,217	136.99	95.50	363,193	137.52	96.02
Wage Father (1000 SEK)	368,791	210.92	185.20	350,543	211.55	186.41
Mother Unemployed	385,574	10.57	30.74	366,986	10.44	30.58
Father Unemployed	385,574	10.96	31.24	366,986	11.06	31.36
First-Born Child	385,574	0.66	0.47	366,986	0.66	0.47
Number Siblings	385,574	1.40	1.14	366,986	1.39	1.15
Immigrant	385,574	0.08	0.27	366,986	0.08	0.27
2nd Generation Immigrant	385,574	0.21	0.41	366,986	0.21	0.41
Adopted	385,574	0.01	0.11	366,986	0.02	0.13
Age Mother	384,736	43.61	5.09	366,442	43.62	5.11
Age Father	379,537	46.50	5.79	360,898	46.50	5.80
Father Unknown	385,574	0.02	0.12	366,986	0.02	0.13
Mother Unknown	385,574	0.00	0.05	366,986	0.00	0.04
Single Mother	385,574	0.07	0.25	366,986	0.07	0.26
Single Father	385,574	0.01	0.10	366,986	0.01	0.09
Schooling Variables						
Primary-School Grade	385,574	-0.17	0.96	366,986	0.18	0.95
High-School Grade	308,584	-0.09	0.89	304,446	0.19	0.86
High-School Attendance	385,574	0.84	0.36	366,986	0.86	0.35
High-School Science Track	385,574	0.12	0.33	366,986	0.10	0.30
Share Females HS Track	325,678	0.38	0.26	316,015	0.62	0.15
Cohort Size	385,574	122.94	41.54	366,986	123.30	41.51
Schools in Municipality	385,574	42.16	57.80	366,986	42.28	57.93
Share Females	385,574	0.49	0.05	366,986	0.49	0.05
College Enrolment	385,574	0.25	0.43	366,986	0.28	0.45
Share Females College	168,494	0.50	0.22	220,660	0.66	0.15
Labour-Market Variables						
Share Females Occupation	333,413	0.34	0.26	318,702	0.66	0.23
Mean Occupation Wage	333,413	266.72	101.06	318,702	234.25	995.13
Log Occupation Wage	333,413	12.43	0.37	318,702	12.28	0.39
Wage (1000 SEK)	385,574	256.58	169.39	366,986	178.30	138.51
Log Income	342,678	12.34	0.94	323,448	11.84	1.17
Unemployed	385,574	8.44	27.79	366,986	11.13	31.45

Note: This table presents summary statistics for the key variables in the paper. The grades variables in both high- and primary school are standardized average grades. We do not observe all variables for the same number of people. Data for parental occupational earnings potential is not available to us at the time of school attendance as it starts after our observation period.

primary school. The register contains indicators for the school and year in which primary education was completed.[14]

Since we observe the gender of each student, we can then calculate the share of female peers for each school–cohort during primary school by counting the number of female students and dividing it by the total number of students.[15] A majority of schools offers all 9 years of compulsory schooling, but not all do. Data on grades offered by schools in 1995 (Holmlund and Böhlmark, 2019) indicate that, on average, students in our sample were exposed to the same gender composition for around 7 out of 9 years during primary school, equivalent to exposure starting at age 9. Accordingly, the majority of students is exposed to the share of female peers we compute for all of primary school.

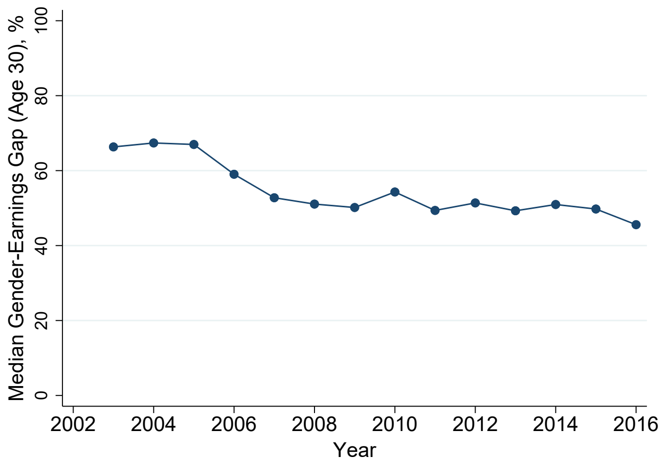
Educational Outcomes. To construct the educational variables, we use the data from the Primary, High School (Registret över slutbetyg från gymnasiet), and University (Universitets- och högskoleregistret) Registers. The first two registers record students' grades and outcomes at the respective stage, with the High School register also recording the study program. The University Registration register then provides information about university programs or courses in which they enrolled.

Labor Market Outcomes. We use the Income and Taxation register (Registret över inkomster- och taxeringar) to link a student's subsequent earnings to their former school–cohort composition. Using the Occupation Register (Yrkeregistret), we do the same for their subsequent chosen occupation. Note that occupation data is available for three years fewer than income data, which reduces the sample size when we consider occupation-related outcomes. Our two main dependent variables are the students' annual income and the median income in a former student's occupation, both measured at age 30. We choose age 30 for two main reasons. First, examining age 30 allows us to observe labor market outcomes for people who have likely completed their tertiary education as the average completion age for college

[14] Unfortunately, the data do not contain a classroom indicator. The classroom indicator variable is available only in the exams register, which starts in the year 2004, two years after the last cohort in our sample.

[15] A potential concern for our measure would be if students frequently changed schools within their municipalities. In a separate project at the Institute for Evaluation of labor Market and Education Policy (IFAU; working paper is forthcoming), we estimate the share of students who stay in the same school within a municipality if the school offers more than 3 consecutive years of education. This share is approximately 95%.

Figure 1: Raw Earnings Gap at Age 30



Note: The figure represents the gender gap in median annual earnings in relative terms at the age of 30 across the years for which we observe labor-market outcomes in our sample. The y-axis denotes the percentage difference between crude median male and female earnings in a given year. The underlying data is adjusted for inflation.

is 26. Second, we do not cede too many cohorts when we restrict the sample only to those we observe until age 30.

The annual individual earnings is labor income as recorded in the income register. This data directly comes from the Taxation Authority (Skatteverket). Occupational information comes from the Occupation Register and contains a three-digit code which identifies an individual's profession. Each digit represents a subsequent sub-division into categories of occupation, starting with eight main categories. We use the 186 unique occupations in our data to calculate a proxy for earnings potential by taking the median income for each occupation based on everyone aged 31 and older. We do the same to calculate the share of women in each occupation.

Descriptive Statistics. Figure 1 shows the development of the unadjusted relative gender gap in labor income over time. While the gap has also been decreasing in Sweden, the reduction of the gender gap has stagnated in recent years. Over the whole sample period, the average the raw difference between male and female earnings corresponds to roughly 45% at age 30.

Table 2 presents more comprehensive summary statistics. The first panel shows background variables related to family characteristics. Boys comprise approximately 51% of the sample, and there is little difference between the genders with respect to family characteristics. The second section shows school-level variables. An average school-cohort consists of around 123 students. Approximately 10% of the students complete the Natural Science track in high school, and around a quarter of students proceed to university. There is no meaningful difference with respect to cohort size and composition between the genders. However, girls in the sample obtain higher grades and are more likely to enroll in college. The last section describes the labor market outcomes we focus on at the age of 30. For both annual and occupational earnings, men out-earn women at that age. There also is gender homophily in occupational choice, with individuals working in occupations comprised of more than 60% of their own gender.

4 Empirical Strategy and Plausibility Checks

4.1 Specifications

We estimate the effects of cohort gender composition on labor market outcomes using the following main specification:

$$\begin{aligned} Y_{isc} = {} & \beta_1 \times Female_i \times ShareFemPeers_{isc} + \beta_2 \times Male_i \times ShareFemPeers_{isc} + \\ & \beta_3 \times Female_i + \alpha_{school} + \delta_{cohort} + X_i \gamma' + e_{isc} \end{aligned} \tag{1}$$

Y_{isc} is the outcome of interest for student i in school s and cohort c. The explanatory variable is $ShareFemPeers_{isc}$, which represents the proportion of female students in a given school-cohort. It is calculated as $\frac{ngirls_c}{size_c}$, where $ngirls_c$ is the number of girls in a given cohort and $size_c$ is the cohort size. The estimate for β_1 describes the effect of the share of female peers on women, β_2 describes the effect on men, and β_3 represents the gender gap in Y_{isc}. α_{school} denotes school-fixed effects, while δ_{cohort} denotes cohort fixed effects. X_i is a vector of school trends, individual- and school-level controls. The vector of controls includes parental education (measured by whether at least one of the parents went to college), log family income, and an indicator for immigrant background. It also includes cohort size (Epple and Romano, 2011) and the number of schools available in a given municipality. We cluster standard errors on the school level, thus allowing students' outcomes to correlate within schools.

In the tables we also show $\beta_1 - \beta_2$ which we call "Gap". It denotes the gender gap in the impact of female peers on each outcome — impact on women minus impact on men — and therefore shows by how much the gender gap changes with more girls in the cohort.

When analyzing peer effects at the school-cohort level, the primary threat to identification lies in potential sorting of students. Here, the institutional context ensures that there is little selection apart from geographic location and age into schools and cohorts. The inclusion of school fixed effects absorbs static heterogeneity in selection into schools, and cohort fixed effects control for national level changes which affect all students in a cohort. A large literature shows that peer composition, conditional on school and cohort fixed effects, is arbitrary in many contexts (see, e.g, Hoxby, 2000; Black, Devereux and Salvanes, 2013; Helene et al., 2015; Carrell, Hoekstra and Kuka, 2018).

While unlikely, dynamic selection across schools and cohorts may be a potential threat to identification. We do two things to address potential dynamic selection into school–cohorts across schools, over time. First, following the literature, we include school-specific time trends to analyze deviations from peer composition conditional on dynamic trends (see, e.g., Black, Devereux and Salvanes, 2013; Carrell, Hoekstra and Kuka, 2018; Brenøe and Zölitz, 2020). This allows to control parametrically for unobserved factors that may correlate with time trends in cohort composition.

We include time trends to account for unobserved factors at the school-level which may change over time and correlate with the evolution of the share of girls within schools as well as educational and earnings outcomes. One example for this could be changes in neighborhoods affecting school, such as an increase in crime. This may lead to a change in gender composition at school and also in educational attainment. Schneeweis and Zweimüller (2012) make another example: Imagine some schools advertise technical subjects, for instance, with a meet-and-greet with former students. This may then lead to more girls attending the school as well as increasing the chance that those girls choose technical high-school tracks later on. Including trends helps alleviate biases that could come from such behavior. Including trends parametrically is an intermediate step to fully account for such potential threats by nonparametrically including $school{\times}cohort$ fixed effects.

Second, we include fixed effects for each individual cohort in a given school (see, e.g., Brenøe and Zölitz, 2020). This more conservative specification including $school{\times}cohort$ fixed effects allows us to address remaining selection into different schools in different years. A drawback is that we are then only able to compute the impact on the gender gap in outcomes and not also on levels. This second specification then looks as follows:

$$Y_{isc} = \beta_1 \times Female_i \times ShareFemPeers_{isc} + \beta_2 \times Female_i + \alpha_{school} \times \delta_{cohort} + X_i\gamma' + e_{isc} \quad (2)$$

The underlying assumption for a causal interpretation in specification (1) including separate school and cohort fixed effects as well as school-specific trends, is that no omitted variable is simultaneously: (i) time-variant and cohort-specific; (ii) not captured by school fixed effects, cohort fixed effects, or linear time trends (iii) correlated with peer composition as well as labor-market outcomes, (iv) not included in our vector of controls. The existence of such a variable seems highly unlikely given the institutional setup. Still, to assess that empirically, we examine the relationship between high-quality and detailed observable characteristics from administrative registers and gender peer composition in the cohort. In addition, in specification (2), we include school-by-cohort fixed effects, which fully accounts for potential changes in sorting to schools over time. Taken together, the results of those checks do not suggest that the main identifying assumption is violated.

4.2 Plausibility and Balance Checks

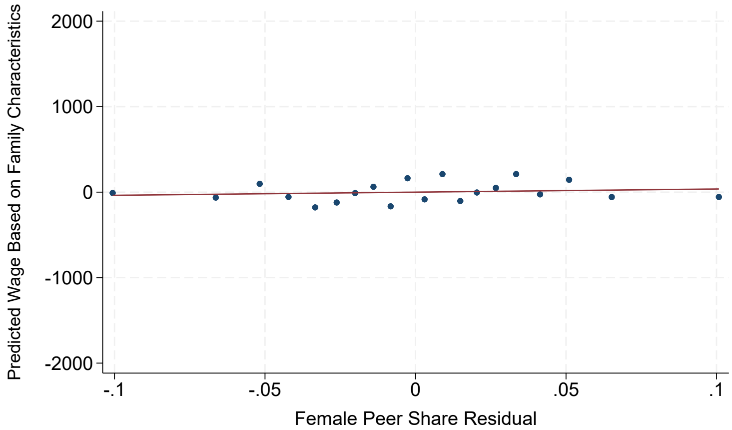
While specification (1) with separate school and cohort fixed effects addresses static selection into schools or cohorts, we now check whether remaining variation is likely arbitrary.

Balance of Family Characteristics Across the Peer Distribution. We provide a series of balance checks for high-quality background variables on parental and family characteristics from administrative data.[16] The 22 variables include, among others, detailed education and labor-market outcomes for each parent, as well as multiple family composition variables (all variables are shown in Table 3).

Figure 2 summarizes our findings by showing the association of our best prediction for the earnings of each individual based on family characteristics with the share of girls. If there was a consistent association between the predicted wage and the share of female peers, this would indicate that parents may be able to select into specific school–cohorts. However,

[16] We measure those just before children start the last phase of primary school.

Figure 2: Female Peer Share and Predicted Earnings Based on Family Characteristics



Note: The figure shows the relationship between one's predicted earnings at age 30, predicted solely based on family characteristics used in Table 3, and the residualized female peer share across school-cohorts, conditional on school and cohort fixed effects (that is, both variables are residuals from a regression on school and cohort fixed effects). The dots show the binned averages across 20 quantiles of the distribution. The figure indicates whether there is sorting based on family characteristics into different types of cohort along the whole peer-share distribution. The linear regression, line shown in blue, indicates no relationship: The coefficient is 369 with se=623 and p =0.55.

Figure 2 indicates that across the whole peer distribution, there is no systematic association between the background of individuals and the share of girls. Moreover, the coefficient and the quantiles are very close to 0 and a fraction of the main estimates as the comparison with Table 4 shows.

We do two more detailed checks. First, for each background characteristic, we examine whether there is a correlation with the share of female peers in the cohort across three specifications including: (1) separate school and cohort fixed effects, (2) fixed effects and school trends, and (3) fixed effects, trends, and controls. Across 66 bivariate regressions, we find only one coefficient that is statistically significant at the 10% level (Appendix, Table A.1). The results are in line with previous findings of Getik and Meier (2022) for Swedish primary schools, who find that the share of female peers across cohorts and across classrooms within cohorts is unrelated to parental characteristics including mental health as well as student mental health.

Another concern then is a potential nonlinear imbalance in family characteristics. To examine this, we run regressions with the preferred specification including quintiles of the peer share, see Table 3. Across 110 coefficients, there is 1 statistically significant coefficient at the 1% level, 4 at the 5% level, and 4 at the 10% level. Out of the 22 variables, the $F-$test for equality across the quintiles is marginally significant at the 5% level for only one of them.

Table 3: Balance Checks Across the Quintiles of the Share of Female Peers

Peer Quintile:	**1**	**2**	**3**	**4**	**5**	F p**-value**
Mother High School	0.002	-0.003**	-0.001	0.001	0.001	0.18
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Father High School	-0.001	-0.002	0.003*	-0.000	0.001	0.29
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Mother Vocational Degree	0.000	-0.001	0.001	-0.001	0.001	0.58
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Father Vocational Degree	-0.001	0.001	0.000	-0.000	0.000	0.9
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Mother College Degree	0.001	-0.001	-0.002	0.001	0.001	0.28
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Father College Degree	0.001	-0.004***	0.001	0.001	0.001	0.04
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Mother STEM Degree	-0.000	-0.000	-0.000	0.000	0.000	0.77
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Father STEM Degree	0.001	-0.001	-0.000	-0.001	0.001	0.24
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Log Family Income	-0.007	0.008	0.016	-0.007	-0.011	0.58
	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)	
Log Wage Mother	-0.000	0.001	-0.008	0.007	0.001	0.84
	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)	
Log Wage Father	-0.029**	0.020	0.029**	-0.015	-0.007	0.07
	(0.014)	(0.015)	(0.014)	(0.015)	(0.015)	
Mother Unemployed	-0.039	0.010	-0.017	0.035	0.012	0.99
	(0.108)	(0.107)	(0.114)	(0.107)	(0.118)	
Father Unemployed	-0.092	-0.075	-0.023	0.088	0.110	0.63
	(0.114)	(0.105)	(0.104)	(0.111)	(0.112)	
First-Born Child	0.001	0.001	0.001	-0.002	-0.001	0.72
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
Number Siblings	0.003	0.001	-0.006*	-0.003	0.006	0.22
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	
Immigrant	0.000	-0.000	-0.001	0.000	0.001	0.6
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
2nd Generation Immigrant	-0.000	0.001	-0.002*	0.000	0.001	0.45
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Adopted	-0.000	-0.000	0.001*	0.000	0.000	0.39
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Age Mother	0.003	-0.010	0.000	-0.006	0.013	0.92
	(0.016)	(0.015)	(0.016)	(0.016)	(0.017)	
Age Father	-0.007	0.011	-0.020	0.018	-0.002	0.7
	(0.019)	(0.018)	(0.019)	(0.018)	(0.019)	
Mother Unknown	-0.000	-0.000	-0.000	0.000	-0.000	0.71
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Father Unknown	-0.000	-0.000	-0.000	0.000	0.000	0.91
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Note: The table shows the estimated relationship between student family characteristics and different quintiles of the share of female peers in one's cohort (each quintile is regressed separately on each variable separately indicated in the first column). The specifications in the table incrementally include school and cohort fixed effects, school trends, and school-level controls. Those controls include cohort size and the number of schools in the municipality. The last column shows the F-statistic for the joint significance of the quintiles. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3: Residual Share of Female Peers Across School-Cohorts

Note: The figure above represents the distribution of the residualized female peer share across school-cohorts, conditional on separate fixed effects for schools and cohorts. The overlaid curve represents normal distribution.

Similarly to the linear specification above, these results suggest that there is no systematic correlation of family characteristics across the distribution of the share of female peers.

Gender and the Share of Female Peers in Cohorts. Similarly, a violation of the identifying assumption would occur if there was gender-based selection into cohorts. To test this, we do three related checks. First, we examine whether a student's own gender correlates with the leave-one-out share of girls in their cohort, following the methodology proposed by Guryan, Kroft and Notowidigdo (2009).[17] There is no statistically significant correlation between own gender and the share of female peers (Appendix, Table A.2).

Second, following Chetty et al. (2011), and Balestra, Eugster and Liebert (2022), we regress student gender on school-by-cohort fixed effects. The school-by-cohort fixed effects should be jointly insignificant if assignment to a school specific cohort is independent of student gender (Chetty et al., 2011). We find that this is indeed the case. Last, if the estimated impact on the gender gap remains similar when we include school-by-cohort fixed effects, it is unlikely that selection into specific school cohorts has a significant effect. The stable impact on gender gaps across specifications (see, e.g., Table 4), therefore, render it unlikely that selection into specific school-cohorts is a key driver of the main results.

[17] Across all specifications, we control for the school-level leave-one-out cohort mean of the share of girls to account for the mechanical relationship between peer and own gender. That is, we control for the average share of female peers in the other cohorts of a student's school.

Distribution of the Share of Girls. In another plausibility check of as-good-as-random assignment to school-cohorts, we examine the variation of the gender peer share that we eventually exploit. If gender peer share were as-good-as-randomly assigned at the school-cohort level, we would expect that the corresponding distribution of peer shares would look normally distributed, conditional on school and cohort fixed effects. Figure 3 suggests that the residual share of female peers is indeed well-behaved and follows a normal distribution. A simulation of us randomly assigning students to school cohort confirms that the distribution looks as-good-as random (Appendix, Figure A.1).

In sum, these tests indicate that the share of girls is likely arbitrary. For more details on plausibility and balance checks, see Appendix Section A.

5 Main Results: Effects on Earnings and Occupational Earnings Potential

5.1 Effects on Earnings and the Wage Gap

Main Effects. Figure 4 and Table 4 show that a higher share of female peers in primary school increases women's income at age 30, reducing the earnings gap. In Table 4, "Female × Share Females" shows the effects on the earnings level of females, "Male × Share Females" shows the effects on the earnings level of males. "Gap" shows the difference between the two effects and therefore indicates by how much an increase in the share of female peers reduces the gender gap. Finally, the female dummy shows the earnings gap, conditional on controls.

The effects on women earnings levels and the earnings gap are precisely estimated across specifications. A 10 percentage point increase in the share of girls (avg. = 49%, sd=0.05) results in an SEK 1,721 or approximately $354 increase in annual earnings for women at age of 30. This corresponds to a 2.7% reduction in the earnings gap. Neglecting wage trajectories across age, it leads to a $12,390 (35 years x $354) lifetime earnings difference up to the retirement age of 65.

In contrast, men do not to benefit from more girls in their cohort in terms of later earnings, with the estimate being less than half the magnitude and not statistically significant. Based on specification (3), a 10 percentage points higher share of female peers leads to a statistically insignificant reduction of SEK 671 or $138 annual wage among men.

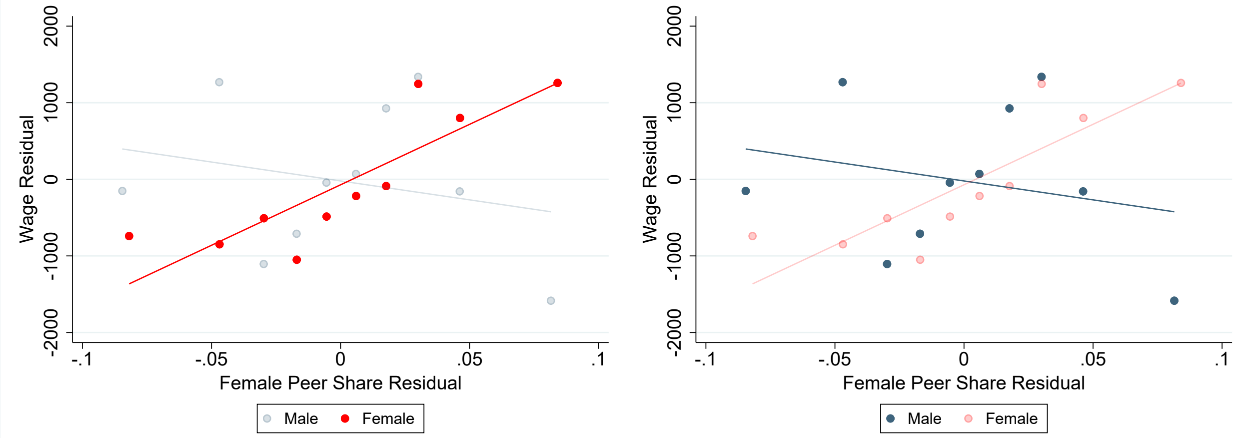
The estimates are robust to the inclusion of school and cohort-fixed effects in column (1) in Table 4, school-specific time trends in column (2), as well as to parental and school-level controls in column (3). The gender gap estimates remain stable even when including fixed effects for each school specific cohort in column (4) without controls and in column (5) with controls. Specifications (4) and (5) allow us to address potential remaining concerns about selection on the school-cohort level. While we cannot observe gender-specific effects in these specifications, we can still observe the difference between the effects on men and women. The robustness to the different specifications suggests that the estimates are not the results of static or dynamic sorting into schools, cohorts, or school specific cohorts (Oster, 2019).

Effect Size. The effects are modest, but meaningful when compared to previous literature and plausible benchmarks. Chetty, Friedman and Rockoff (2014) estimate that a one standard deviation improvement in teacher value-added over one year of exposure leads to a $494 wage increase at age 28. Chetty et al. (2011) find that kindergarten students exposed to teachers for 4 years who had more than 10 years of experience earn $364 more per year between the ages of 25 and 27. In comparison, we find that a 10 percentage point shift (2 standard deviations) in the gender composition over around 7 years, changes annual wages by $354. Further, we compare our estimates to the association of parental education and child wages. A 10 percentage point shift in the gender composition amounts to almost 10% of the relationship between parental college degree and child wage.

Examining the effects of the gender environment in secondary school, Black, Devereux and Salvanes (2013) find a 0.06 log point effect for a change from an all male to an all girls cohort, whereas we document a larger 0.12 log point effect (Appendix, Table C.4). These results align well, considering that in our case the duration of exposure to the peer environment is twice as long as the exposure in Black, Devereux and Salvanes (2013). Taken together, the coefficient estimates indicate a role of gender composition in primary school for female earnings levels and the gender wage gap.

Based on related evidence on the impact of gender environments on mental health (Getik and Meier, 2022), we expect that the classroom composition drives most of the effect of the gender environment captured by the cohort composition. The estimates here are, therefore,

Figure 4: Primary School Gender Composition and Earnings



Note: The figure shows the relationship between residualized earnings measured at age 30 and the residuals of the share of female peers in a given school-cohort (N = 752,560). The residuals stem from regressions of the respective variables on school and cohort fixed effects, school-specific trends, and controls (specification 3 in Table 4). The dots show the binned averages across the deciles of the distribution. The left section of the figure highlights the linear fit from OLS regressions for females ($\beta = 17,211$, $se = 5,172$), whereas the right section does so for males ($\beta = -6,714$, $se = 5,674$).

likely a lower bound of the true effect on girls given that we examine cohorts rather than classes. Unfortunately we do not have data on classrooms for the cohorts here, but we can use the evidence from Getik and Meier (2022) for an educated guess.[18] The previous findings show that the effects of classroom composition on mental health may be up to 50% larger than the estimates for the cohort: Meaning that a 10 percentage point increase in girls at the classroom level could lead up to a $531 increase in earnings at age 30 and a corresponding reduction in the earnings gap by 4%. However, given that we consider a different, long-run outcome here, we prefer the unadjusted estimates.

Effects Across the Earnings Distribution. We further consider how the female earnings shift across the distribution (Appendix, Table B.1). We find that women with more girls in the cohort are 4 percentage points less likely to be in the lowest quintile, but 4 percentage points more likely to be in the highest. The point estimates suggest a shift away from the left-hand side and towards the right-hand side of the earnings distribution for girls. In summary, the results indicate that women earn higher wages after socialization among girls.

[18]In Getik and Meier (2022) where we consider peer effects on mental health using data from Sweden from a later period, we do have data both on classroom (Table 1 of that paper) and cohort composition (Appendix, Table D.2 of that paper) which allows us to compare effects of the two. We find that effects of the gender environment on mental health are stronger when considering the classroom composition. The estimated coefficients are up to 50% larger for classroom composition than for cohort composition with smaller standard errors. The econometric issue that causes this discrepancy is likely a form of measurement error as the cohort composition is an imperfect proxy for the classroom composition.

Table 4: Primary School Gender Composition and Earnings

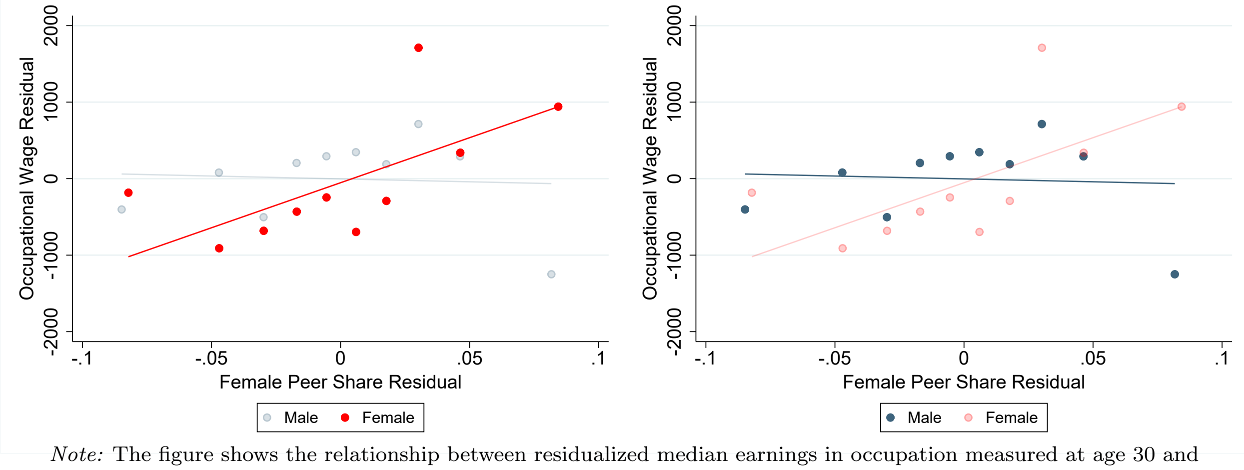
	Annual Earnings (Mean: 218,380; women: 178,229; men: 256,582)				
	(1)	(2)	(3)	(4)	(5)
Female × Share Females	16,822***	18,626***	17,211***		
	(5,062)	(5,199)	(5,172)		
Male × Share Females	-9,072	-7,704	-6,714		
	(5,912)	(5,776)	(5,674)		
Female	-90,886***	-91,077***	-90,002***	-91,072***	-89,681***
	(3,732)	(3,739)	(3,719)	(3,766)	(3,743)
Gap	26,098***	26,491***	24,197***	26,488***	23,552***
	(7,571)	(7,587)	(7,536)	(7,638)	(7,582)
School FE	X	X	X	-	-
Cohort FE	X	X	X	-	-
School Trends	-	X	X	-	-
Controls	-	-	X	-	X
School × Cohort FE	-	-	-	X	X
Observations	752,560	752,560	752,560	752,560	752,560
Schools	537	537	537	537	537
R-squared	0.09	0.09	0.10	0.10	0.11

Note: The table shows the estimated relationship between annual income at age 30 and the share of female peers in one's cohort. The income is recorded in Swedish crowns (SEK). The first row shows the coefficient estimates for women; the second row for men. The row "Female" shows the gross difference in annual earnings between the genders. The "Gap" row shows the difference in response to the share of female peers between the genders. The coefficients in the first three columns are based on the first specification that relies on school and cohort fixed effects. Columns (4) and (5) show estimates from our second specification, which include school-by-cohort fixed effects. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Effect Heterogeneity. We additionally examine heterogeneous effects by students' socioeconomic background, cohort size, or municipality size (parental income, employment and education; cohort size; the number of schools in the municipality; and the size of the municipality; Appendix, Table B.2).

The key motivation for this analysis is to see whether disadvantaged groups, such as children from parents without a college degree or with lower income, are more affected by their peer environment. For instance, girls with a disadvantaged background might benefit more from having more female peers who may be more likely to be role models, to have a similar background, or, coming from a more privileged background, to have information on non-stereotypical educational paths (see also Brenøe and Zölitz, 2020).

Figure 5: Primary School Gender Composition and Occupational Earnings Potential



Note: The figure shows the relationship between residualized median earnings in occupation measured at age 30 and the residuals of the share of female peers in a given school-cohort (N = 652,115). The residuals stem from regressions of the respective variables on school and cohort fixed effects, school-specific trends, and controls (specification 3 in Table 5). The dots show the binned averages across the deciles of the distribution. The left section of the figure highlights the linear fit from OLS regressions for females ($\beta = 11,040$, $se = 3,394$), whereas the right section does so for males ($\beta = -1,600$, $se = 3,084$).

Across six variables capturing parental education and employment, we do not see any clear patterns throughout. We find one statistically significant positive effect of a higher share of girls on girls from a higher income background as measured by log family income, but the other effects are not statistically significant and sometimes point in the opposite direction.

In addition, we examine heterogeneities across cohort sizes and differently sized municipalities. We did not expect differences here, as the Swedish schooling system is relatively homogeneous, but there may be a concern that the effects are driven by particularly small cohorts or municipalities as the variance in cohort shares will be bigger. However, across differently sized municipalities and cohorts, we do not have statistically significant heterogeneities. Jointly, the results suggest that the effects are likely not driven by a particular socio-economic group or area type.

5.2 Effects on Occupational Earnings Potential

Figure 5 and Table 5 indicate that a higher share of girls in primary school also diminishes the gap in occupational earnings potential. Here, the dependent variable is the median annual earnings in one's occupation after the age of 30. This is interesting as occupational earnings are a strong proxy for lifetime earnings.

The estimated impact on female income levels is stable and statistically significant across specifications. In the preferred specification in column (3), we find that a 10 percentage point increase in the share of female peers in primary school increases female occupational income potential by 1,104 SEK per year. This translates to an approximately 3.2% reduction in the gender gap in occupational earnings potential. Neglecting wage trajectories across age, this is equivalent to a SEK 38,640 (35 years x 1,104 SEK) or $7,949 lifetime earnings difference up to the retirement age of 65. The results indicate that gender peer composition in primary school does not only affect current wages, but also has effects on lifetime earnings potential via occupational sorting.

The size of the occupational earnings estimate is approximately 2/3 of the overall impact on earnings shown in Table 4. The smaller magnitude is mechanical as the estimates only capture the effects across, and not within, occupational groups. Several explanations for these differences are possible: Differential wage growth, different entry positions for women vs. men in certain occupational groups, different wage gaps across occupations, different compensations for cognitive vs. non-cognitive skills, structural changes in occupations over time with compensation changes for different age groups, and premiums for entering a non-gender stereotypical occupation. Given the large number of occupational groups and limited power to look at results within groups, we are not able to assess which differences are most important. However, we do examine potential mechanisms for the overall earnings differences in more detail in Section 7.

6 Robustness Checks

We do several checks to assess the robustness of our main effects. We summarize the results in Figure 6 (for details, see Appendix, Section C).

Estimates from Split Samples by Gender. In a first robustness check, we estimate the effect of the share of female peers on annual earnings and occupational earning potential separately for women and men. A disadvantage of that approach is that it makes it impossible to estimate the impact on the gender gap using school-by-cohort fixed effects. An advantage of splitting the sample is that potential interactions of gender with fixed effects and controls are absorbed. Figure 6 shows the estimated effects on realized and earnings potential separately

Table 5: Primary School Gender Composition and Occupational Earnings

	Median Earnings in Occupation Mean: 250,868				
	(1)	(2)	(3)	(4)	(5)
Female × Share Females (234,250)	11,906***	12,661***	11,040***		
	(3,432)	(3,467)	(3,394)		
Male × Share Females (266,724)	-2,476	-2,112	-1,600		
	(3,118)	(3,111)	(3,084)		
Female	-40,579***	-40,772***	-39,701***	-40,679***	-39,477***
	(2,296)	(2,299)	(2,278)	(2,309)	(2,295)
Gap	14,633***	15,025***	12,868***	14,845***	12,422***
	(4,693)	(4,698)	(4,647)	(4,711)	(4,676)
School FE	X	X	X	-	-
Cohort FE	X	X	X	-	-
School Trends	-	X	X	-	-
Controls	-	-	X	-	X
School × Cohort FE	-	-	-	X	X
Observations	652,115	652,115	652,115	652,115	652,115
School-Cohorts	537	537	537	537	537
R-squared	0.20	0.20	0.22	0.21	0.23

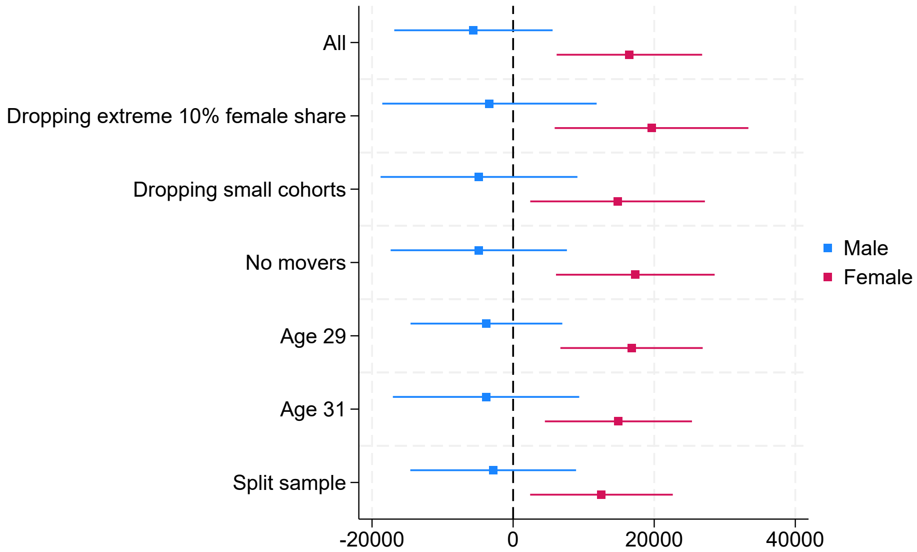
Note: The table shows the estimated relationship between median earnings in a given individual's occupation at age 30 and the share of female peers in one's cohort. These earnings are computed based on 186 unique occupations in our registers and are recorded in Swedish crowns (SEK). The first row shows the coefficient estimates for women; the second row for men. The row "Female" shows the gross difference in annual earnings between the genders. The "Gap" row shows the difference in response to the share of female peers between the genders. The coefficients in the first three columns are based on the first specification that relies on school and cohort fixed effects. Columns (4) and (5) show estimates from our second specification, which include school-by-cohort fixed effects. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

by gender. The estimates for the effects on women's earnings remain positive and statistically significant throughout. The coefficient sizes are not statistically distinguishable from the main estimates. Again, the point estimates for men are negative and statistically insignificant. Estimating the impact of the share of female peers with split samples therefore does not substantially affect the coefficient estimates.

Different Ages for Measuring Current Earnings and Using Logs. In a second robustness check, we examine what happens when we use ages 29 and 31 as the ages at which we measure actual and potential earnings.[19] Across all three ages 29, 30, and 31, we find that earnings of women are higher when they were exposed to more girls in their cohort, leading

[19]Note that we lose some observations at age 31, since we observe one fewer cohort up to this age.

Figure 6: Robustness of the Results to Different Specifications



Note: The figure shows the relationship between earnings measured at age 30 and the share of female peers in a given school-cohort for each of the sub-samples. The squares represent the size of the estimates, and the horizontal lines the 95% confidence interval. The estimates for each gender are produced using the pooled sample, with the exception of the "split sample" coefficients. The regressions correspond to those in Tables C.1, C.2, and C.3 in Appendix C.

to a reduction in the earnings gap. Thus, the results are not sensitive to using a specific cut-off age around age 30. We also examine the impact on log earnings. We see robust and statistically significant effects of the share of female peers (Appendix, Table C.4).

Movers. Individuals who stay in the same municipality do not have an institutional reason to change schools. However, could people who move from one municipality to another drive the results? To examine this concern, we restrict split the sample to non-movers. We define those as students who lived in the same municipality throughout the whole of primary school as well as the year before starting school. Just over 83% of students do not move across municipalities and therefore likely remain in the same school.[20] The coefficient estimates for non-movers does not differ substantially, suggesting that the results are not driven by students moving.

Outliers in Gender Peer Share and Small Cohorts. The results might be driven primarily by the tails of the distribution. To examine the possibility of outliers driving the

[20]We cannot directly observe moving within municipalities. However, in schools offering all three educational stages, 95% of students remain in the same school. Note also that there is no effect heterogeneity across municipalities with more and fewer schools. Regarding private schools, as those started to be implemented in Sweden in the 1990s, they are not in our sample.

results, we first estimate the main specification without the top and bottom 5% of cohorts with the highest share of female peers. In our data, these cut-offs roughly correspond to cohorts having 60% or more students of the same gender. We then also drop the 10% of cohorts with the lowest number of students. Dropping extreme observations results in qualitatively equivalent estimates.

Placebo Check: Previous and Past Cohorts. We do a placebo check examining whether the share of female peers in other cohorts affects outcomes. More specifically, we examine whether the share of female peers in the previous or the past cohort affects labor-market outcomes. We find that none of the corresponding regression coefficients are statistically significant at the 5% level (Appendix, Table C.5). Across 24 coefficient estimates three coefficients are statistically significant at the 10% level. The coefficient estimates are substantially smaller than the impact of the current cohort and have opposing signs. Taken together, the results indicate that our estimates capture idiosyncratic variation coming from the current cohort rather than from previous or past cohorts.

7 Mechanisms: Occupational and Educational Sorting

Why do we observe an increase in women's earnings after they were exposed to more girls in primary school? Previous research highlights the impact of the gender composition on educational attainment and choices (e.g., Schneeweis and Zweimüller, 2012; Giardili, 2020; Brenøe and Zölitz, 2020). However, less is known about how gender composition affects later occupational sorting. Table 6 shows the impact throughout a student's educational and labor market career, its effects on educational attainment as well as educational and labor market sorting. This table guides our exploration of what drives the impact of female peers on labor market outcomes.

7.1 Occupational and Educational Sorting

Educational Attainment, Networks, and Sorting. There are two main educational stages between primary school and labor market outcomes that we can observe: high school (gymnaiset) and tertiary education in university or vocational training. We can trace each

Table 6: Summary of Main Mechanisms

Dependent variable:	**Female**	**Male**	**Gap**
Primary-School Grades	0.082**	-0.141***	0.223***
	(0.038)	(0.038)	(0.053)
High-School Grades	0.011	-0.045	0.056
	(0.039)	(0.039)	(0.046)
Enter High School	0.037***	-0.002	0.040**
	(0.014)	(0.016)	(0.018)
Science Track High School	0.030**	0.004	0.026
	(0.013)	(0.014)	(0.018)
Cohort-Mates in High School	1.638**	-1.335**	3.000***
	(0.709)	(0.675)	(0.688)
College Enrolment	0.007	0.006	0.002
	(0.017)	(0.017)	(0.025)
STEM Enrolment	0.007	0.004	0.004
	(0.009)	(0.013)	(0.016)
Female Share High-School Track	-0.037*	-0.027	-0.010
	(0.022)	(0.022)	(0.041)
Female Share College	-0.021**	-0.001	-0.021
	(0.011)	(0.016)	(0.023)
Female Share Occupation	-0.042**	0.008	-0.050
	(0.017)	(0.017)	(0.031)
Married by 30	-0.020	0.006	-0.025
	(0.015)	(0.014)	(0.020)
Has Children by 30	-0.038**	0.013	-0.051**
	(0.017)	(0.015)	(0.023)
Unemployed	0.000	0.005	-0.005
	(0.011)	(0.009)	(0.015)
School FE	X	X	X
Cohort FE	X	X	X
School Trends	X	X	X
Controls	X	X	X

Note: The table estimates the effect of the share of female peers in the cohort on the observable intermediate outcomes that represent potential channels through which future income can be affected (population mean provided in parentheses next to the variable). The coefficient estimates below "Female" represent the interaction of the share of female peers with being a girl. The coefficient estimates below "Male" represents the impact of female peers on boys. "Gap" shows the difference in response to the share of female peers between the genders Both the primary- and high school grades variables are standardized. The share variables represent the proportion of females, entry, enrolment and completion variables are binary. The fertility variable takes on a value of 1 if the student had a child prior to the age of 30. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. We include all the observation for which we have data on. Standard errors (in parentheses) are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

student's school path: primary- and high school grades, entry into high school and the study track there, as well as tertiary education choices.

We first document a pronounced effect on primary school grades. An increase in the share of female peers improves grade point averages for girls and worsens boys' grades. The estimate suggests that a 10 percentage point increase in the share of female peers results in an approximately 6% widening of the grade gap in favor of girls. We also find that girls with more female peers are more likely to attend high school as well as select the natural sciences high school track. We do not observe direct effects of primary school composition on grades in high school or effects on college enrollment for either gender in line with Anelli and Peri (2019).

We further study what happens to peer networks. We find that girls with more girls in their cohort during primary school later on go to high-school with more of their previous same-gender peers. The opposite happens for boys. Accordingly, more same-gender peers in primary school lead to stronger same-gender networks in high-school, which might be responsible for a part of the positive effect on girls.

Last, we examine whether girls with more girls in the cohort chose less gender stereotypical subjects. Consistent with the findings of Schneeweis and Zweimüller (2012) and Giardili (2020), we observe that girls exposed to more female peers have a smaller share of girls in the track they choose in high school. There appears to be no corresponding selection effects on boys. We also observe a similar pattern with respect to university programs.

In sum, girls with more female peers therefore have a higher educational attainment in primary school and are less likely to choose gender stereotypical high school or college tracks. This finding is also interesting in light of the network formation of girls: The results suggest that girls form groups with other girls early on and then proceed with them to study less gender-stereotypical subjects.

Occupational Sorting. While we previously show that women exposed to more female peers select into better-paying occupations, we have not yet explored the type of jobs into which they sort. To study occupational selection we focus on four factors: 1) unemployment,[21] 2) the share of females in a given occupation (analogous to the educational study tracks), 3) sorting into specific occupational groups, and 4) gender gaps in the selected occupation.

[21] The unemployment indicator takes value 1 if a person received any unemployment benefits during the year he or she turned 30.

First, we observe no statistically discernible impact on unemployment or having a positive income for either men or women (see Appendix Tables D.4 and D.5). The results indicate that the effects we detect do not come from lower unemployment among women.

Second, women who had more female peers in primary school have a smaller share of women in the occupation that they pursue at the age of 30. Consistent with high-school track choices, where girls exposed to more girls select less-gender stereotypical tracks, women therefore sort into less gender-stereotypical jobs with higher earnings.

Third, for a better understanding of the effects on occupational sorting, we study the 8 most general categories of occupations available in the occupation register (Yrkeregistret, first digit of the occupational categorization, Appendix, Table D.1). A higher share of female peers results for women in a decreased likelihood of working in a service and care profession and an increased likelihood of having an occupation that requires academic education. Consistent with the results in Table 6, there are no statistically significant effects on men.

We also examine the 9 occupation categories at the next level of aggregation that attract more than 4% of our sample (second digit of the occupational categorization including 186 occupations, Appendix, Table D.2). A higher share of girls leads to a higher likelihood of working in a technology- or data-related occupation for women. The results on occupation types are consistent with the selection of girls into better-paying and less female-dominated occupations.

Fourth, we observe that women end up sorting into occupations with a slightly higher gender wage gap (see Appendix, Table D.3). Together with the finding that they earn more in those occupations, the results suggest that women exposed to more girls in their cohort may be trailblazers: They end up in more highly qualified, traditionally male-dominated jobs with higher gender gaps.

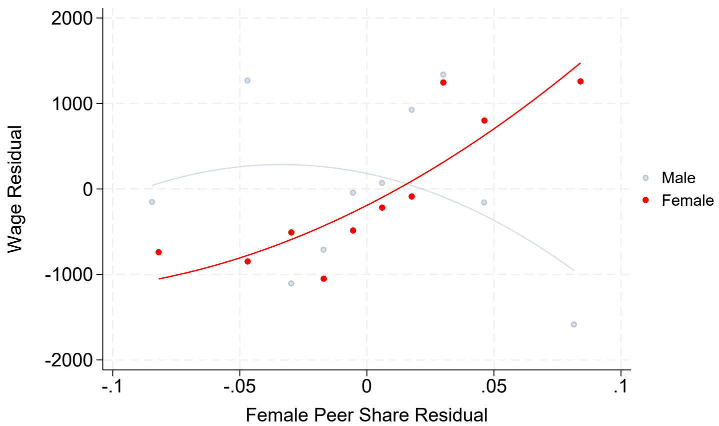
Fertility and Marriage. Additionally, register data allows us to observe fertility by the same age. Similarly to Black, Devereux and Salvanes (2013), we find that women experience an approximately 2.5% reduction in the relative likelihood of having a child with a 10 percentage point increase in the share of girls. Finally, there is no statistically significant impact on the likelihood of being married.

Exploring What Explains the Effects on the Gender Gap. We explore the extent to which we can attribute the higher earnings among women to the above mechanisms. We first naively controlling for each mechanism separately and then successively add the mechanisms jointly (Appendix, Table D.6 and Table D.7). This naive approach suggests that the biggest factor explaining the effects on the wage gap is not fertility or grades, but occupational selection, explaining up to 40% of the effect on girls. This makes sense given the findings that women select less gender stereotypical and higher paying occupations.

7.2 Non-linear Effects of Peer Gender

At which point does the effect of girls start for women and for men? We examine different ways of specifying non-linearities to probe what kind of cohort composition affects boys and girls most.

Figure 7: Primary-School Gender Composition and Earnings, Quadratic Specification



Note: The figure shows the relationship between residualized earnings measured at age 30 and the (quadratic) residuals of the share of female peers in a given school-cohort (N = 752,560). The residuals stem from regressions of the respective variables on school and cohort fixed effects, school-specific trends, and controls (specification 3 in Table 4). The dots show the binned averages across the deciles of the distribution. The linear coefficient for men is -6,276 (*se*=5,807), the quadratic coefficient for men is -93'749 (*se*=70,036), the linear coefficient for women is 15,102 (*se*=4,871), the quadratic coefficient for women is 56,139 (*se*=60,654).

Quadratic Specification. Figure 7 shows the quadratic relationship between earnings and the share of girls. Visually, there is no strong curvature for girls, but some for boys, with boys predicted to do worse with many girls in the cohort. However, the the coefficient estimates for the quadratic terms for both, men and women, are statistically not significant.

Quintile Specification. To allow for more flexibility, we examine the impact of different quintiles of the share of female peers on wages in Table 7. We estimate the following specification:

$$\begin{aligned} Y_{isc} = {} & Female_i \times QuintilesFemPeers_{isc}\beta_1' + Male_i \times QuintilesFemPeers_{isc}\beta_2' + \\ & \beta_3 \times Female_i + \alpha_{school} + \delta_{cohort} + X_i\gamma' + e_{isc} \end{aligned} \tag{3}$$

The only difference to the main specification is that instead of the share of girls, we now include a vector of indicators for the quintiles of the share of girls: $QuintilesFemPeers_{isc}$. It leaves out the third, middle quintile which captures balanced cohorts. The vector with estimates for β_1 describes the effect of the quintiles on women and β_2 describes the effect on men relative to the middle quintile.

Balanced and female-dominated cohorts increase women's wages more than male-dominated cohorts. The highest quintile of 53% to 75% share of female peers increases wages for women the most. However, in female-dominated cohorts, men also experience the highest reduction in earnings statistically significant at the 10% level.

We compare quintile 1 and 5 and also run joint F-tests of all coefficients for both, girls and boys. For girls, both tests are statistically significant (joint test of all coefficients with F=5.31, $p < 0.01$). Both tests for boys are not statistically significant at the 10% level (test for equality of all bins $p = 0.15$, test for equality of the highest bin compared to the lowest bin $p = 0.11$). So while there is an indication that men do worse in quintile 5 compared to the middle quintile ($p < 0.1$), the statistical support for more general non-linearities for men is not very strong. Additional checks show, that once we take out any linear effects, no strong non-linear effects of the quintiles remain for either boys or girls (Appendix, Table D.8). This suggest that most of the effects in Table 7 come from an underlying linear relationship.

Taken together, the results indicate that girls do better with more girls along the whole peer distribution and particularly when they dominate. For boys, however, we have weak evidence suggesting that when the share of girls is high they do worse. This may also be a reason for the mixed findings in the previous literature summarized in Section 2 as the moments of the peer distribution may affect the genders differently.

Table 7: Annual Earnings Effects by Gender Peer Share Quintile

Peer Share Quintile:	Male	Female
	(1)	(2)
Quintile 1 (.11-.45)	-99.12	-393.35
	(934.81)	(845.87)
Quintile 2 (.45-.48)	-681.97	-1850.53**
	(911.22)	(791.72)
Quintile 3 (.48-.5)	ref.	ref.
Quintile 4 (.5-.53)	631.91	150.02
	(863.64)	(803.88)
Quintile 5 (.53-.75)	-1654.09*	1347.34*
	(958.19)	(804.77)

Note: The table shows the estimated relationship between annual earnings at age 30 and the share of female peers in one's cohort expressed as 5 bins (quintiles) of that share, with quintile 3 (the middle one) as the reference category. The mean female share in a given bin is shown in parentheses. The outcome is recorded in Swedish crowns (SEK). Each row represent the corresponding quintile of the female share in cohort. All outcomes are estimated using the preferred specification from column (3) in the main tables. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7.3 Summary and Discussion of Mechanisms

Summary of Mechanisms. Our exploration of the impact of female peers across educational stages and early career reveals six main findings:

1. Girls exposed to more girls have higher primary-school grades and are more likely to attend high school. Conversely, boys' primary-school grades worsen. Accordingly, a higher share of female peers widens the grade gap.

2. Women exposed to more girls do not have a lower likelihood of unemployment.

3. Women exposed to more girls choose less gender-stereotypical educational paths and jobs.

4. Women exposed to more girls have lower fertility by age 30.

5. Occupational and educational sorting plays the biggest role, even conditional on educational attainment.

6. Women benefit from more girls along the whole peer distribution, whereas men may be affected negatively when women dominate.

The results provide a fairly consistent picture. Early exposure to more girls has lasting effects on educational and career trajectories. Women exposed to more girls consistently choose less gender-stereotypical educational paths and jobs. Hence, women socialized among girls seem to be less bound by gender norms and enter occupations with higher lifetime earnings.

Underlying Mechanisms of Gender Composition. Several reasons have been highlighted for why the gender composition affects students. There are two broad categories following Jackson (2021): direct effects coming from student interactions and indirect effects coming from reactions to the gender composition by teachers or schools.

Direct effects can come from interactions per se such as disruption (Lavy and Schlosser, 2011; Hill, 2015), violence (Lavy and Schlosser, 2011; Getik and Meier, 2022), romantic relationships (Hill, 2015), cooperation (Lu and Anderson, 2015), and friendship (Hill, 2015). They can also come from broad changes in preferences and beliefs related to self-confidence (Kessels and Hannover, 2008; Sullivan, 2009; Eisenkopf et al., 2015; Anelli and Peri, 2019), aspirations (Luo and Yang, 2023), competitiveness (Anelli and Peri, 2019; Schøne, von Simson and Strøm, 2020), norms (Eisenkopf et al., 2015; Schneeweis and Zweimüller, 2012), and rule-following (Briole, 2021). Last, differences in information from interactions can come via networks, (Anelli and Peri, 2019) and role-models (Mouganie and Wang, 2020; Balestra, Sallin and Wolter, 2021). All of those can all ultimately affect attendance (Eren, 2017; Borbely, Norris and Romiti, 2023) and well-being (Getik and Meier, 2022).

It is unclear how a higher share of girls plays out for many of the factors, although based on previous evidence more girls likely lead to less disruption. Most other effects are likely asymmetric, that is, more peers of the same gender are beneficial for that gender, but detrimental for the others. The evidence here and in a previous papers of ours in the same context suggests that students benefit from larger same-gender networks and better well-being when around same-gender peers (Getik and Meier, 2022). The results here also indicate that girls change to less gender-stereotypical paths, suggesting a role for changes in norms and information.

Unfortunately we do not have data on teachers or school resources to examine indirect effects. However, previous empirical evidence and theory (Jackson, 2021), suggests that impacts are likely asymmetric as teachers adapt their material better and have more time to focus on the dominant gender (Hill, 2015). Not-statistically significant estimates for boys suggests a combination of asymmetric and overall positive effects. To learn more about the relative weight of direct and indirect factors, detailed survey data combined with long-run administrative data would be invaluable.

8 Conclusion

We examine the impact of early gender socialization on long-run labor market outcomes. To this end, we use unique register data that links cohort composition during primary school to earnings and occupational data for up to 15 years after graduation. We exploit arbitrary variation in the share of girls in a student's cohort to estimate the long-run effects of gender environment at a critical age. We find that a higher share of female peers in primary school increases women's later earnings, thereby reducing the gender gap. A key mechanism behind the findings is that women exposed to more girls select into less gender-stereotypical jobs with higher lifetime earnings potential. This is preceded by them selecting into less gender stereotypical educational tracks in high school and college.

Because of data limitations and institutional reasons, existing evidence on how early gender socialization affects long-run labor market outcomes is still limited. Using detailed Swedish administrative data, we can comprehensively study the effect the primary school environment on later earnings and career trajectories. Our evidence suggests that earnings and occupational selection depend on the social environment in primary school: more girls in the cohort lead to consistently less gender-congruent behavior among girls. Gender roles likely play an important role for explaining why the gender wage gap has persisted in spite of advances in educational attainment. The evidence we present indicates that the early gender environment in school shapes gender roles with persistent impacts on labor market outcomes.

References

Almås, Ingvild, Alexander W. Cappelen, Kjell G. Salvanes, Erik Sørensen, and Bertil Tungodden. 2016. "What explains the gender gap in college track dropout? Experimental and administrative evidence." *American Economic Review*, 106(5): 296–302.

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Chapter 48: Race and gender in the labor market." In *Handbook of Labor Economics*. Vol. 3, 3143–3259. Elsevier.

Anelli, Massimo, and Giovanni Peri. 2019. "The effects of high school peers' gender on college major, college performance and income." *Economic Journal*, 129(618): 553–602.

Balestra, Simone, Aurélien Sallin, and Stefan C. Wolter. 2021. "High-ability influencers? The heterogeneous effects of gifted classmates." *Journal of Human Resources*, forthcoming.

Balestra, Simone, Beatrix Eugster, and Helge Liebert. 2022. "Peers with Special Needs: Effects and Policies." *Review of Economics and Statistics*, 104(3): 602–618.

Bertrand, Marianne. 2011. "New perspectives on gender." In *Handbook of Labor Economics*. Vol. 4, 1543–1590. Elsevier.

Bietenbeck, Jan. 2020. "The long-term impacts of low-achieving childhood peers: Evidence from project STAR." *Journal of the European Economic Association*, 18(1): 392–426.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2013. "Under pressure? The effect of peers on outcomes of young adults." *Journal of Labor Economics*, 31(1): 119–153.

Blau, Francine D., and Lawrence M. Kahn. 2017. "The gender wage gap: Extent, trends, & explanations." *Journal of Economic Literature*, 55(3): 789–865.

Booth, Alison L., Lina Cardona-Sosa, and Patrick Nolen. 2018. "Do single-sex classes affect academic achievement? An experiment in a coeducational university." *Journal of Public Economics*, 168: 109–126.

Borbely, Daniel, Jonathan Norris, and Agnese Romiti. 2023. "Peer Gender and Schooling: Evidence from Ethiopia." *Journal of Human Capital*, 17(2).

Brenøe, Anne. 2022. "Brothers Increase Women's Gender Conformity." *Journal of Population Economics*, 35: 1859–1896.

Brenøe, Anne A, and Ulf Zölitz. 2020. "Exposure to more female peers widens the gender gap in STEM participation." *Journal of Labor Economics*, 38(4): 1009–1054.

Brenøe, Anne Ardila, and Shelly Lundberg. 2018. "Gender gaps in the effects of childhood family environment: Do they persist into adulthood?" *European Economic Review*, 109: 42–62.

Briole, Simon. 2021. "Are girls always good for boys? Short and long term effects of school peers' gender." *Economics of Education Review*, 84: 102150.

Brown, Charles, and Mary Corcoran. 1997. "Sex-based differences in school content and the male-female wage gap." *Journal of Labor Economics*, 15(3): 431–465.

Calkins, Avery, Ariel J. Binder, Dana Shaat, and Brenden Timpe. 2023. "When Sarah Meets Lawrence: The Effects of Coeducation on Women's College Major Choices." *American Economic Journal: Applied Economics*, 15(3): 1–34.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women." *Quarterly Journal of Economics*, 131(2): 633–686.

Card, David, Fabrizio Colella, and Rafael Lalive. 2021. "Gender Preferences in Job Vacancies and Workplace Gender Diversity." NBER WP No. 29350.

Carrell, Scott E, Mark Hoekstra, and Elira Kuka. 2018. "The long-run effects of disruptive peers." *American Economic Review*, 108(11): 3377–3415.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American Economic Review*, 104(9): 2633–2679.

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How does your kindergarten classroom affect your earnings? Evidence from project star." *Quarterly Journal of Economics*, 126(4): 1593–1660.

Chetty, Raj, Nathaniel Hendren, Frina Lin, Jeremy Majerovitz, and Benjamin Scuderi. 2016. "Childhood environment and gender gaps in adulthood." *American Economic Review*, 106(5): 282–288.

Eisenkopf, Gerald, Zohal Hessami, Urs Fischbacher, and Heinrich W. Ursprung. 2015. "Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland." *Journal of Economic Behavior and Organization*, 115: 123–143.

Elsner, Benjamin, Ingo E. Isphording, and Ulf Zölitz. 2021. "Achievement Rank Affects Performance and Major Choices in College." *Economic Journal*, 131(640): 3182–3206.

Epple, Dennis, and Richard E. Romano. 2011. "Chapter 20 - Peer Effects in Education: A Survey of the Theory and Evidence." In *Handbook of Social Economics*. 1053–1163.

Eren, Ozkan. 2017. "Differential Peer Effects, Student Achievement, and Student Absenteeism: Evidence From a Large-Scale Randomized Experiment." *Demography*, 54(2): 745–773.

Feld, Jan, and Ulf Zölitz. 2018. "Peers from Venus and Mars-higher-achieving men foster gender gaps in major choice and labor market outcomes." *mimeo*, UZH.

Fischer, Martin, Martin Karlsson, Therese Nilsson, and Nina Schwarz. 2020. "The Long-Term Effects of Long Terms Compulsory Schooling Reforms in Sweden." *Journal of the European Economic Association*, 18(6): 2776–2823.

Getik, Demid, and Armando N. Meier. 2022. "Peer gender and mental health." *Journal of Economic Behavior & Organization*, 197: 643–659.

Giardili, Soledad. 2020. "Single-sex primary schools and student achievement: evidence from admission lotteries." Mimeo, University of Edinburgh.

Goldin, Claudia, Sari Pekkala Kerr, Claudia Olivetti, and Erling Barth. 2017. "The expanding gender earnings gap: Evidence from the LEHD-2000 census." *American Economic Review*, 107(5): 110–114.

Golsteyn, Bart H. H., Arjan Non, and Ulf Zölitz. 2021. "The impact of peer personality on academic achievement." *Journal of Political Economy*, 129(4): 1052–1097.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo. 2009. "Peer effects in the workplace: Evidence from random groupings in professional golf tournaments." *American Economic Journal: Applied Economics*, 1(4): 34–68.

Helene, Jannie, Grøne Kristoffersen, Morten Visby, Helena Skyt, and Marianne Simonsen. 2015. "Disruptive school peers and student outcomes." *Economics of Education Review*, 45: 1–13.

Hill, Andrew 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, Andrew J. 2017. "The positive influence of female college students on their male peers." *Labour Economics*, 44: 151–160.

Holmlund, Helena, and Anders Böhlmark. 2019. "Does grade configuration matter? Effects of school reorganisation on pupils' educational experience." *Journal of Urban Economics*, 109: 14–26.

Hoxby, Caroline M. 2000. "Peer effects in the classroom: Learning from gender and race Variation." NBER WP No. 7867.

Hu, Feng. 2015. "Do girl peers improve your academic performance?" *Economics Letters*, 137: 54–58.

Jackson, C. Kirabo. 2012. "Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in Trinidad and Tobago." *Journal of Public Economics*, 96(1-2): 173–187.

Jackson, C. Kirabo. 2021. "Can Introducing Single-Sex Education into Low-Performing Schools Improve Academics, Arrests, and Teen Motherhood?" *Journal of Human Resources*, 56(1): 1–39.

Kessels, Ursula, and Bettina Hannover. 2008. "When being a girl matters less: Accessibility of gender-related self-knowledge in single-sex and coeducational classes and its impact on students' physics-related self-concept of ability." *British Journal of Educational Psychology*, 78(2): 273–289.

Kirabo Jackson, C., Rucker C. Johnson, and Claudia Persico. 2016. "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms." *The Quarterly Journal of Economics*, 131(1): 157–218.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. "Children and gender inequality: Evidence from Denmark." *American Economic Journal: Applied Economics*, 11(4): 181–209.

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

Lee, Soohyung, Lesley J. Turner, Seokjin Woo, and Kyunghee Kim. 2014. "All or nothing? The impact of school and classroom gender composition on effort and academic achievement." NBER WP No. 20722.

Lu, Fangwen, and Michael L. Anderson. 2015. "Peer effects in microenvironments: The benefits of homogeneous classroom groups." *Journal of Labor Economics*, 33(1): 91–122.

Lundborg, Petter, Erik Plug, and Astrid Würtz Rasmussen. 2017. "Can women have children and a career? IV evidence from IVF treatments." *American Economic Review*, 107(6): 1611–1637.

Luo, Yiyang, and Songtao Yang. 2023. "Gender peer effects on students' educational and occupational expectations." *China Economic Review*, 77: 101898.

Mouganie, Pierre, and Yaojing Wang. 2020. "High-performing peers and female STEM choices in school." *Journal of Labor Economics*, 38(3): 805–841.

Mulligan, Casey B., and Yona Rubinstein. 2008. "Selection, Investment, and Women's Relative Wages Over Time." *Quarterly Journal of Economics*, 123(3): 1061–1110.

Olivetti, Claudia, and Barbara Petrongolo. 2016. "The Evolution of Gender Gaps in Industrialized Countries." *Annual Review of Economics*, 8(1): 405–434.

Oosterbeek, Hessel, and Reyn Van Ewijk. 2014. "Gender peer effects in university: Evidence from a randomized experiment." *Economics of Education Review*, 38: 51–63.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Validation." *Journal of Business Economics and Statistics*, 37(2): 187–204.

Park, Hyunjoon, Jere R. Behrman, and Jaesung Choi. 2012. "Single-sex education: Positive effects." *Science*, 335(6065): 165–166.

Park, Hyunjoon, Jere R. Behrman, and Jaesung Choi. 2018. "Do single-sex schools enhance students' STEM (science, technology, engineering, and mathematics) outcomes?" *Economics of Education Review*, 62: 35–47.

Schneeweis, Nicole, and Martina Zweimüller. 2012. "Girls, girls, girls: Gender composition and female school choice." *Economics of Education Review*, 31(4): 482–500.

Schøne, Pål, Kristine von Simson, and Marte Strøm. 2020. "Peer gender and educational choices." *Empirical Economics*, 59(4): 1763–1797.

Shan, Xiaoyue. 2021. "The Minority Trap: Minority Status Drives Women Out of Male-Dominated Fields." mimeo: Uni. Zurich.

Sloane, Carolyn M., Erik G. Hurst, and Dan A. Black. 2021. "College Majors, Occupations, and the Gender Wage Gap." *Journal of Economic Perspectives*, 35(4): 223–48.

Slotwinski, Michaela, and Alois Stutzer. 2023. "Women Leaving the Playpen: the Emancipating Role of Female Suffrage." *The Economic Journal*, 133(650): 812–844.

Sullivan, Alice. 2009. "Academic Self-Concept, Gender and Single-Sex Schooling." *Educational Research Journal*, 35(2): 259–288.

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

Zölitz, Ulf, and Jan Feld. 2021. "The Effect of Peer Gender on Major Choice in Business School." *Management Science*, 67(11): 6963–6979.

The long-run effects of peer gender
on occupational sorting and the wage gap

Demid Getik[‡] Armando N. Meier[⋆]

Supplemental Appendix

Table of Contents

[‡] Demid Getik, Durham University and Center for Economic Demography at Lund University (CED), email: demid.getik@durham.ac.uk
[⋆] Armando N. Meier, University of Basel, email: armando.meier@unibas.ch

A Balance

A.1 Linear Balance Checks

Table A.1: Linear, Bivariate Balance Checks

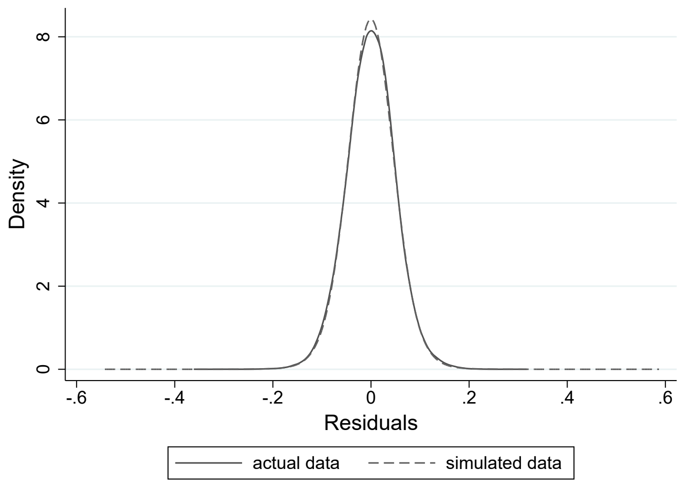
Dependent Variable:	(1)	(2)	(3)
Mother High School	0.015	0.009	0.009
	(0.013)	(0.011)	(0.011)
Father High School	0.007	0.010	0.010
	(0.012)	(0.011)	(0.011)
Mother Vocational Degree	0.016*	0.012	0.013
	(0.008)	(0.008)	(0.008)
Father Vocational Degree	0.006	0.005	0.005
	(0.007)	(0.007)	(0.007)
Mother College Degree	0.015	0.009	0.009
	(0.009)	(0.009)	(0.010)
Father College Degree	0.013	0.009	0.009
	(0.009)	(0.009)	(0.009)
Mother STEM Degree	0.002	0.003	0.003
	(0.002)	(0.002)	(0.002)
Father STEM Degree	-0.005	-0.004	-0.004
	(0.004)	(0.004)	(0.004)
Log Family Income	0.013	-0.076	-0.082
	(0.118)	(0.081)	(0.081)
Wage Mother (1000 SEK)	1.993	-0.800	-0.800
	(3.463)	(2.446)	(2.446)
Wage Father (1000 SEK)	1.875	1.275	1.275
	(6.445)	(4.769)	(4.769)
Mother Unemployed	-0.292	0.192	0.240
	(0.950)	(0.809)	(0.806)
Father Unemployed	0.561	1.049	1.072
	(0.870)	(0.815)	(0.817)
First-Born Child	-0.009	-0.011	-0.011
	(0.012)	(0.012)	(0.012)
Number Siblings	0.006	0.019	0.019
	(0.033)	(0.028)	(0.028)
Immigrant	-0.001	0.003	0.004
	(0.011)	(0.007)	(0.007)
2nd Generation Immigrant	0.007	0.006	0.007
	(0.012)	(0.010)	(0.010)
Adopted	-0.003	-0.001	-0.001
	(0.003)	(0.003)	(0.003)
Age Mother	0.083	0.088	0.088
	(0.143)	(0.129)	(0.129)
Age Father	0.119	0.125	0.128
	(0.151)	(0.141)	(0.141)
Mother Unknown	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)
Father Unknown	0.001	-0.000	-0.000
	(0.003)	(0.003)	(0.003)
School FE	X	X	X
Cohort FE	X	X	X
School Trends	-	X	X
Controls	-	-	X

Note: The table shows the estimated relationship between student family characteristics and the share of female peers in their cohort. The specifications in the table incrementally include school and cohort fixed effects, school trends, and school-level controls. Those controls include cohort size and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.2 Distribution of the Peer Share

We further test whether peer-gender variation within schools is consistent with random assignment by comparing the actual distribution to a simulated distribution of the female peer share. To this end, we do Monte Carlo simulations in which we assign students randomly to cohorts within their schools. We take the number and size of cohorts from the actual data. Similarly to Bietenbeck (2020), we then regress the share of female peers on school and cohort fixed effects in the simulated data and collect the residuals. We plot the simulated residuals from random assignment alongside the residuals from the actual data in Figure A.1. The distributions look very similar, a result consistent with as-good-as-random assignment of the share of female peers.

Figure A.1: Simulated and Actual Residual Share of Female Peers Across School-Cohorts



Note: The figure above represents the actual and simulated distribution of the residualized female peer share across school-cohorts, conditional on school and cohort fixed effects.

A.3 Student Gender and the Gender Peer Share

Table A.2: Effects of Own Gender on the Share of Female Peers

	Share of Female Peers		
	(1)	(2)	(3)
Female	-0.0006	-0.0007	-0.0006
	(0.0007)	(0.0006)	(0.0006)
School FE	X	X	X
Cohort FE	X	X	X
School Trends	-	X	X
Controls	-	-	X
Observations	752,560	752,560	752,560
Schools	537	537	537
R-squared	0.12	0.20	0.20

Note: The table shows the estimated relationship between the share of female peers in a cohort and a student's own gender. Following Guryan, Kroft and Notowidigdo (2009), we control for the school-level leave-one-out cohort share of females. That is, we control for the share of females in the rest of the school leaving out the cohort of the student under consideration. Controls include parental education, income, and mental health as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Gender and Cohorts. Following Chetty et al. (2011), and Balestra, Eugster and Liebert (2022) we do another check as follows: In the first step, we regress student gender on separate school and cohort fixed effects as well as on controls and we then retrieve the residuals from this regression. In the second step, we regress the residuals obtained in the prior regression on school-by-cohort fixed effects. We then do a joint F-test to determine whether the school-by-cohort fixed effects are jointly significant. Across three different specifications the F-statistics suggest no predictive power of school-by-cohort fixed effects for student gender: $F = 1.019$ without controls, $F = 1.020$ with school-level controls, and $F = 1.020$ with school-level and individual-level controls. The F-statistics are all not statistically significant at the 10% level.

B Effects on Earnings

B.1 Effects on Earnings Quintile

Table B.1: Gender Peer Share Effects by Earnings Quintile

Income Quintile:	**Male**	**Female**
	(1)	(2)
Quintile 1 (0 - 47778)	-0.000	-0.017
	(0.013)	(0.014)
Quintile 2 (47779 - 182656)	0.024*	-0.027
	(0.014)	(0.017)
Quintile 3 (182657 - 265422)	0.014	-0.016
	(0.014)	(0.015)
Quintile 4 (265423 - 342073)	0.005	0.020
	(0.019)	(0.016)
Quintile 5 (342074 - 9650019)	-0.042**	0.040***
	(0.016)	(0.013)

Note: The table shows the estimated relationship between annual earnings at age 30 expressed as 5 bins (quintiles) of earnings and the share of female peers in one's cohort. The boundaries of a given bin in terms of the respective annual earnings are recorded in parentheses. Each row illustrates the effect of the female peer share in one's cohort on the likelihood of being in a given income quintile. The outcomes are estimated separately for males and females. All outcomes are estimated using the preferred specification from column (3) in the main tables. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.2 Heterogeneity

Table B.2: Heterogeneity by SES Variables

	Female	**Male**
SES variable:	(1)	(2)
Female Share	18,626***	-7,704
	(5,199)	(5,776)
Parental Education	739	378
	(1,243)	(1,624)
Parent Went to College	-2,748	-7,529
	(10,652)	(12,932)
Parent Unemployed	-4,694	4,923
	(11,320)	(12,001)
Parental Unemployment Benefits	361	-572
	(1,337)	(1,421)
Log Family Income	3,647**	2,259
	(1,579)	(1,926)
Above-Median Family Income	5,250	-17,006
	(9,736)	(10,633)
Cohort Size	84	24
	(113)	(121)
Above-Median-Size Cohort	6,914	-450
	(10,731)	(12,251)
Schools in Municipality	17	-120
	(122)	(109)
Above-Median-Size Municipality	5,760	-4,897
	(10,006)	(12,031)
Observations	753,131	752,561
Schools	537	537

Note: The table presents heterogeneous effect of the share of female peers in the cohort on annual earnings at age 30. The first row replicates the main results of the paper presented in column (2) of Table 2. Column (1) presents the interaction effects for females, and column (2) for males. Parental Education variable is calculated as the total number of years outside of compulsory education; parental unemployment benefits is the amount of unemployment benefits received. Standard errors (in parentheses) are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C Robustness Checks

C.1 Sample Splits

Table C.1: Effects on Earnings when Splitting the Sample by Gender

	Annual Wage			Occupation Wage		
	(1)	(2)	(3)	(4)	(5)	(6)
Females	13,088***	14,918***	13,568***	10,941***	11,401***	9,782***
	(4,941)	(5,123)	(5,101)	(3,529)	(3,586)	(3,535)
Males	-6,854	-5,860	-4,428	-2,200	-1,517	-1,169
	(5,982)	(6,056)	(5,951)	(3,393)	(3,417)	(3,381)
School FE	X	X	X	X	X	X
Cohort FE	X	X	X	X	X	X
School Trends	-	X	X	-	X	X
Controls	-	-	X	-	-	X
Observations	752,560	752,560	752,560	652,115	652,115	652,115
Schools	537	537	537	537	537	537
R-squared	0.03	0.04	0.05	0.16	0.16	0.18

Note: The table shows the estimated relationship of the share of female peers with annual earnings at age 30 and occupational earning potential, with the sample split by gender. The outcomes are recorded in Swedish crowns (SEK). The first row shows the results for women; the second row for men. The first three columns show the relationship for the annual earnings; the last three columns for median occupation earnings. Occupational earnings are computed based on 186 unique occupations. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.2 Measurement Age

Table C.2: Effects on Labor-Market Outcomes at Different Age Cut-Offs

	Annual Wage			Occupation Wage		
Age:	29	30	31	29	30	31
Female × Share Females	17,781***	17,211***	15,396***	11,416***	11,866***	12,509***
	(5,057)	(5,172)	(5,264)	(3,422)	(3,608)	(3,821)
Male × Share Females	-5,247	-6,714	-4,991	-831	-2,198	-3,392
	(5,439)	(5,674)	(6,715)	(3,305)	(3,331)	(3,591)
Female	-80,827***	-90,002***	-95,994***	-36,871***	-39,198***	-41,382***
	(3,712)	(3,719)	(4,068)	(2,280)	(2,419)	(2,563)
Gap	23,138***	24,197***	20,556**	12,417***	14,305***	16,210***
	(7,433)	(7,536)	(8,179)	(4,635)	(4,937)	(5,228)
School FE	X	X	X	X	X	X
Cohort FE	X	X	X	X	X	X
School Trends	X	X	X	X	X	X
Controls	X	X	X	X	X	X
Observations	751,550	752,560	698,950	651,224	652,115	609,003
Schools	537	537	537	537	537	537
R-squared	0.09	0.10	0.10	0.18	0.21	0.20

Note: The table shows the estimated relationship between annual earnings at ages 29-31/occupational earning potential at ages 29-31 and the share of female peers in one's cohort. Occupational earnings are computed based on 186 unique occupations. The outcomes are recorded in Swedish crowns (SEK). The first row shows the results for women; the second row for men. The row "Female" shows the gross difference in annual earnings between the genders. The "Gap" row shows the difference in response to the share of female peers between the genders. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.3 Extreme Observations and Movers

Table C.3: Robustness to Excluding Extreme Observations or Movers

	Full Sample	**Extreme 10%**	**Small Cohorts**	**Movers**
	(1)	(2)	(3)	(4)
Females × Share Females	16,480***	19,615***	14,804**	17,320***
	(5,250)	(6,989)	(6,300)	(5,725)
Male × Share Females	-5,631	-3,358	-4,840	-4,871
	(5,714)	(7,735)	(7,100)	(6,359)
Female	-89,558***	-90,185***	-87,983***	-90,838***
	(3,746)	(5,114)	(4,516)	(4,180)
Gap	22,356***	23,281**	20,251**	22,192***
	(7,596)	(10,425)	(9,191)	(8,447)
School FE	X	X	X	X
Cohort FE	X	X	X	X
School Trends	X	X	X	X
Controls	X	X	X	X
Observations	742,833	681,621	592,172	622,924
School-Cohorts	537	537	478	537
R-squared	0.10	0.10	0.10	0.11

Note: The table shows the estimated relationship between annual earnings at age 30 and the share of female peers in one's cohort. The outcomes are recorded in Swedish crowns (SEK). The first row shows the results for women; the second row for men. The row "Female" shows the gross difference in annual earnings between the genders. The "Gap" row shows the difference in response to the share of female peers between the genders. The estimates in column (1) correspond to the estimates of our main results. In column (2), we exclude individuals who come from a cohort from an extreme in the distribution of the female peer share (top or bottom 5%). In column (4), we exclude individuals who come from a cohort from a cohort that lies in the bottom 10% of the cohort size distribution. In column (5), we only include non-movers. We define those to be individuals who resided in one municipality throughout the entirety of primary school (9 years). Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.4 Log Earnings

Table C.4: Effects of the Gender Peer Share on the Annual Log Earnings

	Log Annual Wage Mean: 12.1				
	(1)	(2)	(3)	(4)	(5)
Female × Share Females	0.11**	0.13***	0.12***		
	(0.04)	(0.04)	(0.04)		
Male × Share Females	-0.01	0.00	0.00		
	(0.04)	(0.04)	(0.04)		
Female	-0.56***	-0.56***	-0.56***	-0.56***	-0.56***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Gap	0.12**	0.13**	0.12**	0.13**	0.12*
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
School FE	X	X	X	-	-
Cohort FE	X	X	X	-	-
School Trends	-	X	X	-	-
Controls	-	-	X	-	X
School × Cohort FE	-	-	-	X	X
Observations	666,126	666,126	666,126	666,126	666,126
School-Cohorts	537	537	537	537	537
R-squared	0.07	0.07	0.08	0.08	0.08

Note: The table shows the estimated relationship between log annual earnings at age 30 and the share of female peers in one's cohort. The first row shows the results for women; the second row for men. The row "Female" shows the gross difference in annual earnings between the genders. The "Gap" row shows the difference in response to the share of female peers between the genders. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.5 Placebo Check

Table C.5: Effect of Gender Composition in Other Cohorts on Earnings

	Annual Wage			Occupation Wage		
	(1)	(2)	(3)	(4)	(5)	(6)
Males (t-1)	-9,757	-7,554	-7,120	2,630	4,075	3,375
	(5,961)	(6,070)	(5,971)	(3,459)	(3,562)	(3,450)
Females (t-1)	-8,903*	-8,808	-9,063*	3,743	2,001	1,562
	(4,871)	(5,154)	(5,134)	(3,495)	(3,630)	(3,552)
Males (t+1)	5,513	8,897	9,886*	-2,184	-1,120	-818
	(5,311)	(5,538)	(5,507)	(3,252)	(3,366)	(3,343)
Females (t+1)	-1,467	-490	790	3,240	2,545	3,274
	(4,569)	(4,831)	(4,814)	(3,543)	(3,603)	(3,605)
School FE	X	X	X	X	X	X
Cohort FE	X	X	X	X	X	X
School Trends	-	X	X	-	X	X
Controls	-	-	X	-	-	X

Note: The table shows the estimated relationship between annual earnings at 30 and the share of female peers in the previous (t-1) and the subsequent cohorts (t+1). Males refers to the sample only consisting of boys, females refers to the sample only consisting of girls. The first three columns present the relationship for annual earnings; the subsequent three columns for occupational earnings. Occupational earnings are computed based on 186 unique occupations. The specifications for each of the variables incrementally include school and cohort fixed effects, school trends, and a vector of controls. This vector includes parental education, income, and mental health as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D Mechanisms

D.1 Occupational Sorting

Table D.1: Effects of Female Peer Share on General Occupational Groups

Dependent variable	Female	Male	Gap
	(1)	(2)	(3)
Leading Role	0.006	0.004	0.003
	(0.006)	(0.006)	(0.008)
Requires Degree	0.013	0.017	-0.004
	(0.016)	(0.015)	(0.022)
Specialized Work	0.030*	0.005	0.025
	(0.016)	(0.015)	(0.021)
Office/Customer Work	0.001	-0.000	0.001
	(0.012)	(0.010)	(0.015)
Service/Care	-0.039*	0.014	-0.053
	(0.022)	(0.019)	(0.033)
Nature-related	-0.003	-0.001	-0.002
	(0.003)	(0.004)	(0.005)
Craft and Building	0.008	-0.016	0.024
	(0.012)	(0.017)	(0.026)
Unqualified Work	-0.008	0.007	-0.014
	(0.008)	(0.008)	(0.012)
School FE	X	X	X
Cohort FE	X	X	X
School Trends	X	X	X
Controls	X	X	X

Note: The table estimates the relationship between the share of cohort female peers and subsequent selection into different categories of occupation. This classification is based on the eight primary occupation categories in the Occupation register (Yrkeregistret). "Leading Role" corresponds to the "Ledningsarbete" category; "Requires Degree" refers to occupations requiring longer academic training (Arbete som kräver teoretisk specialkompetens); "Specialized Work" refers to occupations with shorter academic trainings (Arbete som kräver kortare högskoleutbildning eller motsvarande kunskaper); "Office/Customer Work" corresponds to "Kontors- och kundservicearbete". "Service/Care" to "Service-, Omsorg-, och Försäljningsarbete"; "Nature-related" to "Arbete inom jordbruk trädgård, skogsbruk och fiske". "Craft and Building" to "Hantverksarbete inom byggverksamhet och tillverkning". "Unqualified Work" refers to work that does not require special qualifications (Arbete utan krav på särskild yrkesutbildning). The dependent variable is a dummy for working in a given occupation. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are clustered at the school level.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.2: Effects of Female Peer Share on Occupational Choice, Most Common Occupations

Group:	**Female**	**Male**
	(1)	(2)
Technology/Data (.09)	0.019**	0.004
	(0.008)	(0.006)
Biology/Medicine (.04)	-0.007	-0.002
	(0.009)	(0.005)
Teaching (.08)	0.004	0.009
	(0.011)	(0.008)
Other Requiring Higher Education (.14)	0.001	-0.016
	(0.013)	(0.011)
Office Work (.05)	0.006	-0.002
	(0.008)	(0.007)
Service (.12)	-0.001	0.018
	(0.017)	(0.013)
Sales (.06)	-0.012	-0.008
	(0.009)	(0.008)
Building/Construction (.07)	0.004	-0.008
	(0.011)	(0.015)
Machinery/Transport (.05)	-0.001	-0.011
	(0.009)	(0.011)

Note: The table estimates the relationship between the share of cohort female peers and subsequent selection into different categories of occupation. This classification is based on the occupation categories in the Occupation register (Yrkeregistret) which comprise more than 4% of the sample. "Technology/Data" corresponds to the "Arbete som kräver teoretisk specialkompetens inom teknik och datavetenskap" category; "Biology/Medicine" to "[...] inom biologi, hälso- och sjukvård"; "Teaching" to "Lärararbete". 'Other Requiring Higher Education" to "Annat arbete som kräver teoretisk specialkompetens"; "Service" to "Kontors- och Kundservicearbete" "Sales" to "Service-, Omsorg-, och Försäljningsarbete"; "Building/Construction" to "Hantverksarbete inom byggverksamhet och tillverkning"; "Machinery/Transport" to "Process- och Maskinsoperatörsarbete, Transportarbete".The dependent variable is a dummy for working in a given occupation. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are clustered at the school level.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.3: Gender Gap in Median Earnings in Occupation at Age 30

	Earnings Gap in Chosen Occupation Mean: 48410				
	(1)	(2)	(3)	(4)	(5)
Female × Share Females	2,909**	2,968**	2,721*		
	(1,414)	(1,508)	(1,504)		
Male × Share Females	-1,279	-1,511	-1,415		
	(1,542)	(1,532)	(1,537)		
Female	-15,736***	-15,877***	-15,706***	-15,928***	-15,742***
	(1,109)	(1,138)	(1,141)	(1,146)	(1,148)
Gap	4,218*	4,502*	4,159*	4,590*	4,215*
	(2,268)	(2,329)	(2,334)	(2,342)	(2,346)
School FE	X	X	X	-	-
Cohort FE	X	X	X	-	-
School Trends	-	X	X	-	-
Controls	-	-	X	-	X
School × Cohort FE	-	-	-	X	X
Observations	652,115	652,115	652,115	652,115	652,115
School-Cohorts	537	537	537	537	537
R-squared	0.05	0.05	0.06	0.06	0.07

Note: The table shows the estimated relationship between the gender gap in median earnings in one's chosen occupation at the age of 30 and the share of female peers in one's cohort. The first row shows the results for women; the second row for men. The row "Female" shows the gross difference in annual earnings between the genders. The "Gap" row shows the difference in response to the share of female peers between the genders.These earnings are computed based on 186 unique occupations in our registers and are recorded in Swedish crowns (SEK). The coefficients in the first three columns are based on the first specification that relies on school and cohort fixed effects. Columns (4) and (5) record the estimates produced by our second specification, which include school-by-cohort fixed effects. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D.2 Unemployment and Indicator for Positive Income

Table D.4: Gender Peer Share and Unemployment

	Unemployed at 30 Mean: 9.75				
	(1)	(2)	(3)	(4)	(5)
Female × Share Females	0.02	-0.04	0.08		
	(1.13)	(1.11)	(1.11)		
Male × Share Females	0.71	0.71	0.70		
	(0.98)	(0.93)	(0.93)		
Female	3.00***	3.03***	2.97***	3.00***	2.91***
	(0.72)	(0.73)	(0.73)	(0.73)	(0.73)
Gap	-0.68	-0.74	-0.62	-0.68	-0.49
	(1.46)	(1.47)	(1.47)	(1.48)	(1.48)
School FE	X	X	X	-	-
Cohort FE	X	X	X	-	-
School Trends	-	X	X	-	-
Controls	-	-	X	-	X
School × Cohort FE	-	-	-	X	X
Observations	752,560	752,560	752,560	752,560	752,560
School-Cohorts	537	537	537	537	537
R-squared	0.03	0.03	0.04	0.04	0.04

Note: The table shows the estimated relationship between having been unemployed in the year when turning 30 and the share of female peers in one's cohort. The variable is 100 for unemployment and 0 otherwise. We classify someone as having been unemployed in that year if they received unemployment benefits at any point during the year. The first row shows the results for women; the second row for men. The row "Female" shows the gross difference in annual earnings between the genders. The "Gap" row shows the difference in response to the share of female peers between the genders. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.5: Gender Peer Share and Positive Income in Given Year

	Positive Income Mean: .89				
	(1)	(2)	(3)	(4)	(5)
Female × Share Females	-0.01	-0.01	-0.01		
	(0.01)	(0.01)	(0.01)		
Male × Share Females	-0.01	-0.01	-0.00		
	(0.01)	(0.01)	(0.01)		
Female	-0.01	-0.01	-0.00	-0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Gap	-0.00	-0.00	-0.01	-0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
School FE	X	X	X	-	-
Cohort FE	X	X	X	-	-
School Trends	-	X	X	-	-
Controls	-	-	X	-	X
School × Cohort FE	-	-	-	X	X
Observations	752,561	752,561	752,561	752,561	752,561
School-Cohorts	537	537	537	537	537
R-squared	0.01	0.01	0.02	0.02	0.03

Note: The table shows the estimated relationship between having been having a positive income in the year when they turn 30 and the share of female peers in one's cohort. The variable is 1 for positive income and 0 otherwise. The first row shows the results for women; the second row for men. The row "Female" shows the gross difference in annual earnings between the genders. The "Gap" row shows the difference in response to the share of female peers between the genders. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D.3 Attribution of Effects to Mechanisms

Note, our goal is to assess the extent to which these mechanisms can account for the effects of gender composition, and not to make adjustments to our estimates by controlling for endogenous variables.

Accounting for primary-school and high-school grades, dummies for high-school study tracks and college tracks, eight occupation dummies, and fertility, reduces the size of the coefficient estimate capturing the impact of the share of females peers on the gender gap by about 40% (Table D.7, column 5).[22] Comparing across the different mechanisms, including dummies for the occupations accounts for the largest reduction in coefficients (Table D.6, column 4). This specification also indicates that even within broad occupational categories, women earn more after being exposed to more girls. Importantly, fertility does not account for a sizable portion of the impact of female peers on the gender gap.[23]

[22]Controlling for educational attainment and choices does not affect the estimated impact of the gender environment on the selection into non-gender stereotypical occupations. Note also that controlling for study track dummies accounts for differential competitiveness across tracks.

[23]For all analysis, we reduce the sample of these analyses to observations for which we observe all variables that we control for.

Table D.6: Effects on Earnings, Including Intermediate Stage Controls

	Annual Wage Mean: 246,128				
	(1)	(2)	(3)	(4)	(5)
Female × Share Females	17,798***	17,569***	13,676**	12,159**	16,687***
	(5,737)	(5,680)	(5,425)	(4,918)	(5,535)
Male × Share Females	-6,560	-1,165	-6,460	-5,210	-6,328
	(6,006)	(5,873)	(5,793)	(5,428)	(6,007)
Gap	24,686***	19,056**	20,465***	17,634**	23,361***
	(7,948)	(7,915)	(7,440)	(6,986)	(7,809)
Grades	-	X	-	-	-
Study Tracks	-	-	X	-	-
Occupation	-	-	-	X	-
Fertility	-	-	-	-	X
School FE	X	X	X	X	X
Cohort FE	X	X	X	X	X
School Trends	X	X	X	X	X
Controls	X	X	X	X	X
Observations	538,099	538,099	538,099	538,099	538,099
Schools	537	537	537	537	537
R-squared	0.15	0.19	0.20	0.30	0.16

Note: The table shows the estimated relationship between annual earnings at age 30 and the share of female peers in one's cohort including intermediate controls discussed in the mechanisms section. The first row shows the results for women; the second row for men. The "Gap" row shows the difference in response to the share of female peers between the genders. The first column represents the main set of results. The second column includes primary- and high-school grades. The specification in column (3) includes dummies for high-school tracks and university programs. Column (4) includes dummies for the 8 occupational categories shown in Table D.1. Column (5) also includes a dummy for giving birth. The sample size in each specification is reduced to a sub-sample for which we can observe all the relevant variables. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.7: Effects on Earnings, Including All Intermediate Stage Controls

	Annual Wage Mean: 246,128				
	(1)	(2)	(3)	(4)	(5)
Female × Share Females	17,798***	18,528***	14,726***	11,540**	11,050**
	(5,737)	(5,641)	(5,362)	(4,829)	(4,746)
Male × Share Females	-6,560	-236	-3,049	-3,860	-3,891
	(6,006)	(5,962)	(5,783)	(5,413)	(5,442)
Gap	24,686***	19,071**	18,084**	15,655**	15,211**
	(7,948)	(7,914)	(7,362)	(6,819)	(6,812)
Grades	-	X	X	X	X
Study Tracks	-	-	X	X	X
Occupation	-	-	-	X	X
Fertility	-	-	-	-	X
School FE	X	X	X	X	X
Cohort FE	X	X	X	X	X
School Trends	X	X	X	X	X
Controls	X	X	X	X	X
Observations	538,099	538,099	538,099	538,099	538,099
Schools	537	537	537	537	537
R-squared	0.15	0.19	0.22	0.32	0.32

Note: The table shows the estimated relationship between annual earnings at age 30 and the share of female peers in one's cohort including the main intermediate controls discussed in the mechanisms section. The first row shows the results for women; the second row for men. The "Gap" row shows the difference in response to the share of female peers between the genders. The first column represents the main set of results. Column (2) additionally includes primary- and high-school grades. The specification in column (3) additionally includes dummies for high-school tracks and university programs. Column (4) also includes dummies for the 8 occupational categories shown in Table D.1. Column (5) also includes a dummy for giving birth. The sample size in each specification is reduced to a sub-sample for which we can observe all the relevant variables. Standard errors (in parentheses) are based on clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D.4 Non-linearities

Table D.8: Are There Important Non-linearities? Residual Annual Earnings Effects by Gender Peer Share Quintile

Peer Share Quintile:	Male	Female
	(1)	(2)
Quintile 1 (.11-.45)	-515.65	801.73
	(934.91)	(846.53)
Quintile 2 (.45-.48)	-831.75	-1399.98*
	(911.19)	(791.90)
Quintile 3 (.48-.5)	ref.	ref.
Quintile 4 (.5-.53)	778.88	-272.46
	(863.74)	(804.12)
Quintile 5 (.53-.75)	-1248.02	163.53
	(957.81)	(803.73)

Note: The table shows the estimated relationship between annual earnings residuals taking out the linear relationship between the annual earnings and gender peer share and the share of female peers in one's cohort expressed as 5 bins (quintiles) of that share, with quintile 3 as the reference category. The mean female share and the range for a given bin is shown in parentheses. The outcome is recorded in Swedish crowns (SEK). Each row represent the corresponding quintile of the female share in cohort. All outcomes are estimated using the preferred specification from column (3) in the main tables. Controls include parental education, income and family composition as well as class size, cohort size, and the number of schools in the municipality. Standard errors (in parentheses) are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$