# Gender Peer Effects in Post-Secondary Vocational Education [1]

**Fernanda Ramírez-Espinoza** [2]

November 7, 2021

[Click here for latest version]

**Abstract**

This paper presents evidence that women and men benefit from having a higher percentage of female peers in post-secondary vocational STEM programs. I use idiosyncratic variation in gender composition across cohorts within majors within branches (campuses) for identification. Having a higher percentage of female peers positively affects students in STEM majors, decreasing women's dropout rates and increasing GPA. The peer effect seems to be mediated by the gender of the instructors: as female students have fewer female instructors, the effect of having more female peers intensifies. For men, the effects are in the same direction but smaller and not statistically significant, suggesting that policies that increase the representation of women need not entail a trade-off for male STEM students.

Although educational attainment gaps have not only narrowed but have reversed in most high-income countries and Latin America (Goldin, 2002; Goldin, Katz, & Kuziemko, 2006; Duryea, Galiani, Nopo, & Piras, 2007), a high degree of occupational segregation remains: men and women are still concentrated in different occupations (Schneeweis &

[1] I thank DUOC UC for providing data, institutional knowledge, and important feedback to this work. I also want to thank Lawrence Katz, Ricardo Paredes, Kosuke Imai, Shom Mazumder, Eric Taylor, Felipe Barrera-Osorio, Virginia Lovison, Veronica Frisancho, and Mikko Silliman for their helpful comments.

[2] framireze@g.harvard.edu

Zweimüller, 2012). This is an essential point in that gender wage differences are partly attributable to the subjects that men and women choose to study. Consequently, studying the mechanisms through which women persist (or desist) in high-paying paths like STEM fields helps us understand how to close this persistent gender inequity.

In this paper, I hypothesize that female students' educational outcomes will be positively related to having more female peers in their cohorts in STEM majors. My analysis is motivated by previous literature that suggests that peer effects exist and are particularly salient in the STEM fields, where women have lower participation than in other fields (Hoxby, 2000; Bostwick & Weinberg, 2018; Sacerdote, 2011). The identification strategy to test this hypothesis uses observational data from 129,378 students of the largest post-secondary vocational institution in Chile. The outcome studied is first-year dropout and standardized Grade Point Average (GPA).

To identify the gender peer effect, I estimate a linear model of educational outcomes on gender composition and gender of the student, including an interaction term that captures how female students differentially respond to gender composition. To control for unobserved characteristics of the students that might be related to dropout rates and gender composition, I rely on deviation from long-term trends in gender peer composition within major-by-branch across the years. For example, to calculate the effect, I compare different cohorts of Automotive Mechanic in the Valparaiso campus, where students were exposed to a slightly different percentage of women year to year. To do so, I include a major-by-branch fixed effect and time trend. Additionally, I include control variables such as age, diagnostic math test scores, education of the mother, working status, financial aid status, and shift (day/night).

Many researchers and teachers have argued that peer composition is an important a determinant of student outcomes (Sacerdote, 2011). This issue seems to be particularly crucial for women: there is evidence that women respond more than men to peer influences, consistent with social psychology theories that peers affect female students more

(Han & Li, 2009).

There is evidence that the gender composition of peers affects outcomes and that these effects are different for boys and girls (Busso & Frisancho, 2021; Mouganie & Wang, 2020; Zölitz & Feld, 2020). Lavy and Schlosser (2011) find large positive effects from the percent of girls within a classroom, and they also interpret these effects as working through more than merely increasing peer average test scores. Along the same lines, (Hoxby, 2000) finds modestly large effects of peer background on own test scores, using idiosyncratic gender variation. Paredes (2018) finds that single-sex classrooms reduce the math gender gap by more than half in the Chilean context. She finds that this effect is driven by the gender composition of the classroom itself.

Although the gender peer effect literature in primary and secondary education is robust, post-secondary education remains a less-explored area. Some papers that study how culture may be connected to female underrepresentation have been published recently, like those authored by Lundberg (2017) and Wu (2017). Likewise, there are some studies on STEM graduate program admissions and persistence. For instance, Bostwick and Weinberg (2018) use a difference-in-difference approach and find that an increase in the percentage of female students differentially increases the probability of on-time graduation for women. This paper contributes to the emergent body of evidence of gender peer effects in post-secondary education.

Furthermore, studies on gender peer effects in vocational education are almost nonexistent, although it comprises a significant part of educational systems worldwide. In some developed countries, one-quarter of cohorts pursue professional programs. In the United States, certificate graduation rates are burgeoning — tripling in recent years (*Skills beyond school: synthesis report*, 2014). This study is situated in Latin America, where the post-secondary education sector is also growing fast. This trend is observed in countries with the highest secondary education completion rates, such as Colombia, Mexico, Brazil, Chile, and Peru. As other Latin American countries raise their secondary education completion rates, this paper sheds light on how gender peer effects play out in this specific context.

The results I present in this paper suggest that a 10% increase in the percentage of female peers within major-by-branch units, close to the mean idiosyncratic variation in the data, is associated with a reduction of 9.6% (1.9 percentage points) in female students' dropout rate and a 0.05 standard deviation increase in GPA. This result supports the hypothesis that female students' educational outcomes are positively related to having more female peers in their cohorts in STEM majors. For men, this relationship is of a smaller magnitude and not significant for dropout and GPA, suggesting that men in STEM programs are not harmed by having more female peers. These results are robust to the inclusion of controls and major-by-branch time trends.

Some may argue that the problem is not some particular gender dynamic in STEM majors but male-concentrated majors. If this were the case, I would observe a similar effect for male-concentrated STEM and Non-STEM majors. Nevertheless, when the effect is calculated for non-STEM majors that are male concentrated, I observe that the coefficients for percentage of female peers are statistically significantly different from STEM-major coefficients, and of opposite sign.

The causal identification strategy relies on the assumption that the variation in gender peer composition within major-by-branches is as good as random. To support this idea, I estimate an autoregressive model of gender peer composition with major-by-branch and year-fixed effects. In this context, peer composition in the previous year is not significantly correlated to gender peer composition in the current year, supporting the idea that the change of percentage of female peers is idiosyncratic within major-by-branch.

The paper is organized as follows. Section 1 summarizes the key features of the application process in the vocational education institution that will be studied, describes data sources, reports summary statistics, and tests for balance on the treatment variable and the plausibility of the identifying assumption. In Section 2, I describe the identification strategy used in this study. I present the main results in Section 3 and provide a preliminary mechanism exploration in Section 4. I conclude in Section 5 by interpreting my findings in the context of gender peer effects.

# 1 Context and Data

## 1.1 Institutional Context

Chile's education system comprises four levels of education: pre-school, primary education, secondary education, and tertiary education (also known as higher education). Three types of institutions can provide higher education: Universities, Professional Institutes (IPs), and Centers for Technical Training (CFTs). The most significant difference between universities and the vocational sector is the type and length of training they provide. Universities focus on formal academic training, while IPs and CFTs are vocational and focus on developing practical work skills. This is reflected in the length of courses — the average minimum length of university degrees for incoming students is about nine semesters (Arango, Evans, & Quadri, 2016). Students enrolling in universities or vocational institutions must choose a major when they are admitted, and most of their classes are with peers of the same or similar majors.

This paper uses information from DUOC UC [3], the largest vocational higher education institution in Chile. It tends to 19.3% of all vocational higher education students in the country, offering 75 majors in 9 areas of health, tourism, construction, and management. In 2018, they had 102,817 enrolled students in their 15 branches, present in three regions of the country. Twelve of these branches are located in the Greater Santiago Area (Figure 1), which covers an area similar to New York City and has 6,257,516 inhabitants. Given

[3]For more information, visit: http://www.duoc.cl/international-affairs

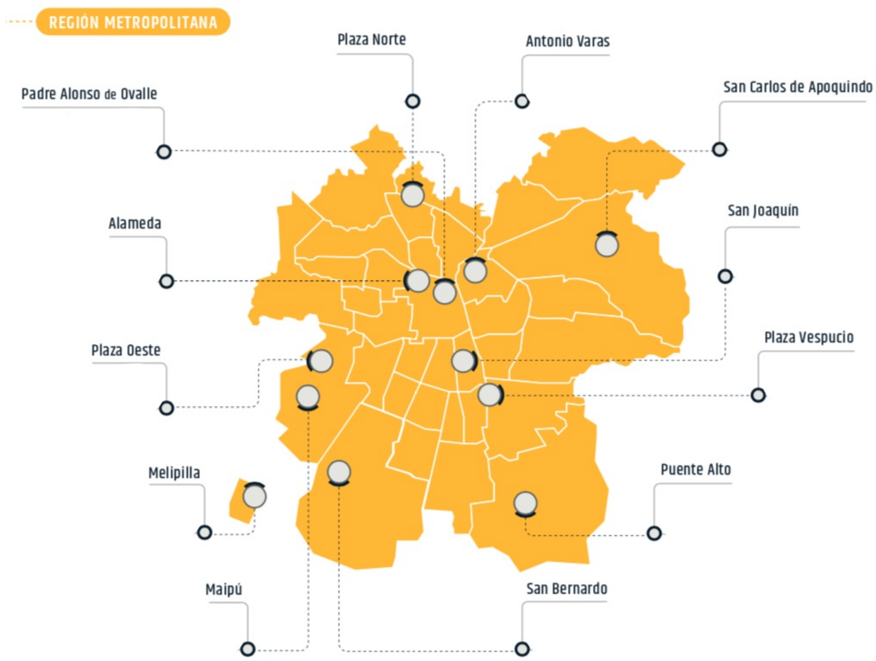


Figure 1: Map of DUOC branches in the Greater Santiago area

that some majors are taught in more than one branch, there are 323 Major-by-Branch combinations.

Although academic requirements to study in DUOC UC are not strict, the institution holds the highest level of accreditation in the system — a standard shared only with three of the best universities in the system. According to this metric, Duoc UC is the highest quality vocational higher education institution in the country.

## 1.2 The enrollment process

DUOC UC has a rolling admissions process that is "first-come, first-served." Besides some minimal academic requirements, there are no requisites for enrollment besides proof of payment (first come-first served). Therefore, neither the students nor the institution can accurately forecast the percentage of female peers they will have during their first semester. In theory, only the last student who enrolls could know how many female peers she/he would be exposed to, but the characteristics of who are already enrolled are not public, so the student does not have this information available when she/he makes the decision. Therefore, individuals students cannot know exactly how many female peers they will have.

Another essential institutional detail is how higher education studies are taught in the Chilean context. For the higher education system in general and DUOC UC in particular, students enroll directly into specific majors. This has two significant implications: first, if they wish to change majors, they must re-enroll to the new major and start this second major from scratch. The second, and most important in the context of the study, is that they always study with the same group of peers: those from their same year-of-entry, same major and same branch. Therefore, the concentration of female peers is the same for all students in that major in a particular branch, and the opportunity for inter-major gender peer influence is less likely.

As it was previously stated, DUOC UC has 15 branches in different parts of the country. Each of them offers a set of majors (e.g., Electrical Technician, Dental Technician, Gastronomy), and most of the majors are offered in more than one branch. For this study, I consider the relevant peer group as those studying the same major at the same branch in the same year of entry. Therefore, a first-year Electrical Technician student in branch A has a different peer group than a first-year Electrical Technician student in branch B.

## 1.3 Data

I use data from all enrolled students in Duoc UC from 2014 to 2018. The dataset includes all individuals that studied one of the 72 majors continuously offered between 2014 and 2018 in its 15 branches. This dataset contains information on 129,378 students, with characteristics like gender, age, and mother's education. It also includes information on their working status, the scores on a mathematics diagnostic test all students take before the beginning of their first academic year, if they attend school during the day or at night (night shift), gender of their instructors, first year dropout rates, and GPA. All this information allows testing for the effects and mechanisms described in Section 2.

The measurement of educational outcomes for this dataset is students' dropout during her first year of studies at Duoc UC and students' GPA. From the data, we get that 18.1% of students dropout in the first year. As shown in Table 1, STEM majors have a higher dropout rate than non-STEM majors (18.1% vs. 15.2%). STEM majors are markedly male concentrated: on average, they have 12.5% of women, whereas non-STEM majors have 60%. Compared to the mean, differences in the other covariates are small in magnitude, showing that the differences are not practically meaningful.

Table 1: Summary Statistics for STEM and Non-STEM students

| | STEM | Non-STEM | Difference |
|---|---|---|---|
| Dropout | 0.181 | 0.152 | 0.029*** |
| | | | (0.002) |
| GPA (in SD) | -0.11 | 0.148 | -0.258*** |
| | | | (0.006) |
| Percentage Female | 0.125 | 0.6 | -0.475*** |
| | | | (0.001) |
| Age | 21.243 | 21.632 | -0.389*** |
| | | | (0.113) |
| Diagnostic score | 48.819 | 45.936 | 2.883*** |
| | | | (0.121) |
| Mothers' education | 0.632 | 0.639 | -0.007** |
| | | | (0.003) |
| Works | 0.572 | 0.558 | 0.024*** |
| | | | (0.003) |
| Has financial aid | 0.661 | 0.666 | -0.005 |
| | | | (0.003) |
| Night Shift | 0.366 | 0.254 | 0.113*** |
| | | | (0.003) |
| N | 44,234 | 59,912 | 104,146 |

Notes: This table presents means and mean difference Standard errors are reported in parenthesis ($^{*}$p<0.1;$^{**}$p<0.05; $^{***}$p<0.01)

The "treatment" in this setting is each student's percentage of female peers in their first semester. This is calculated taking the number of female peers over the number of

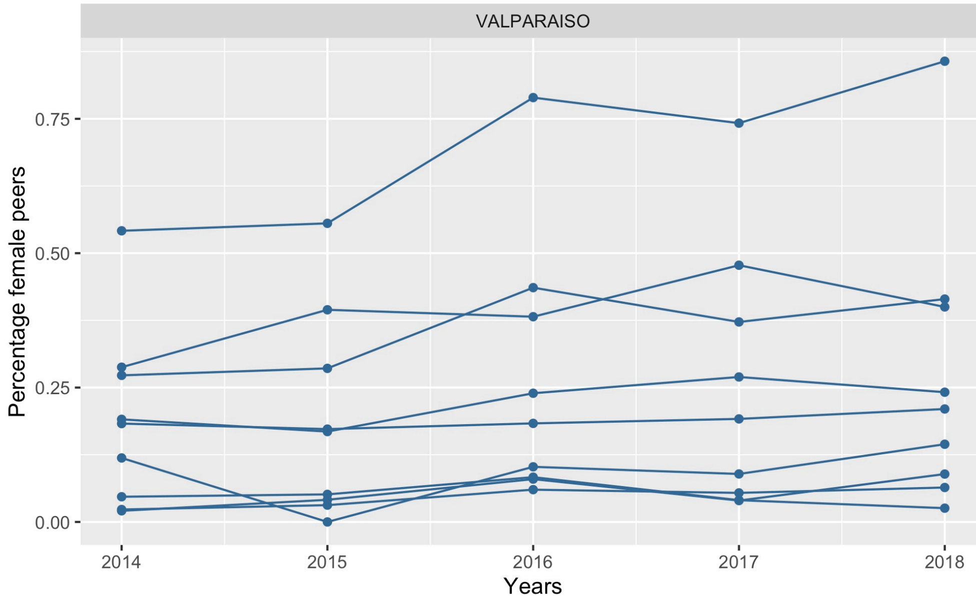


Figure 2: Within Major-by-Branch variation in percentage of female peers (Valparaiso branch)

total peers per major-by-branch on their first semester:

$$PercentageFemale_{itk} = \frac{\sum_{i' \neq i}^{n_{tk}} Female_{i'}}{n_{tk} - 1} \tag{1}$$

Where $n_{tk}$ is student $i$'s number of peers $n$ in the major-by-branch $k$ in cohort $t$.

The identifying variation used in this paper is the variation of the average percentage of female peers students have within the 323 major-by-branches. Figure 2 illustrates the variation of treatment within major for the one particular branch (located in Valparaiso). The year-to-year variation plotted in Figure 2 is the identifying variation in this context. The variation for all the institution branches can be seen in Figure A.

## 2 Causal Identification Strategy

In this context, the treatment is the percentage of female peers a student has in their program, in their branch, in their entrance cohort. Through a linear model that includes year-fixed effects, major-by-branch fixed effects, and time trends, I use year-to-year deviations from long-term trends in the percentage of female peers within Major-by-Branch to estimate the effect of interest.

As the treatment was not randomized, there is the potential for bias in estimating the treatment effect. If the change of percentage of women were somehow correlated with unobservables correlated with the outcome, the estimate of treatment effect would be biased. An example of an unobservable that could threaten the identification strategy is a program director concerned about gender imbalance making efforts to attract more women to STEM programs and implementing initiatives to reduce female dropout or improve their grades.

In this paper, the identifying assumption is that the treatment was assigned as good as randomly, conditional on the controls. If this is satisfied, the average treatment effect will be given by:

$$ATE = \mathbb{E}[\mu(Y_{it(\%fem+1\%)}/X_{itk}) - \mu(Y_{it(\%fem)}/X_{itk})] \tag{2}$$

Where $Y_{it(\%fem)}$ is the percentage of female peers student $i$ has in her cohort $t$ (treatment), and $X_{itk}$ are observable characteristics of student $i$ of cohort $t$ in major-by-branch $k$. Because the treatment $Y_{i(\%fem)}$ is continuous, instead of using the traditional binary treatment of $Y_{i(\%fem)}$ being equal to 0 or 1, I use the continuous version of finding the difference in outcome between $Y_{i(\%fem+1\%)}$ and $Y_{i(\%fem)}$ conditioned in all other observables.

To estimate the effect, I use a linear regression model. Because I am interested in knowing the effect of the percentage of female peers and how this effect is different for men and women, I will use an interaction term[4] between gender (male) and percentage of

[4] It is crucial to address the fact that when I include an interaction term with gender in the potential outcomes framework, there is an implication that both percentage of female peers and gender can be manipulated, analogous to treatment in a randomized experiment. If gender is not manipulable, then there is the danger of post-treatment bias stemming from the fact that almost all variables on which I will condition

female peers. The linear regression model I will use to estimate the effect of the percentage of female peers on women and men is:

$$Y_{itk} = \beta_0 + \beta_1 \% female_{tk} + \beta_2 Male_{itk} + \beta_3 \% female_{tk} * Male_{itk} + \beta_4 X_{itk} + \gamma_t + \delta_k + \psi_k year_{st} + \varepsilon_{itk} \tag{3}$$

$Y_{itk}$ is the outcome of interest (dropout or standardized GPA) for student $i$ in cohort $t$ in major-by-branch $k$, $\% female_{tk}$ is the percentage of female peers in major-by-branch $k$ in cohort $t$, $Male_{itk}$ is a dummy that takes value of 1 if the student $i$ in cohort $t$ in major-by-branch $k$ is male, $X_{itk}$ is a vector of student's covariates, $\gamma_t$ are years fixed effects, $\delta_k$ are major-by-branch fixed effects and $\psi_k$ is a set of major-by-branch-specific linear time trends. Errors are clustered to the major-by-branch-by-year level. The coefficient of interest is $\beta_1$, that indicates how percentage of female peers is related to dropout for women. If the assumptions hold, then $\beta_1$ would identify the causal effect of percentage of female peers on educational outcomes for women.

## 2.1 Evidence on the Feasibility of the Identifying Assumption Strategy

In the absence of a randomized treatment assignment, I need to assume that the percentage of women is not correlated with unobservables correlated with the outcome. As it involves unobservables, there is no definitive way to test this assumption. Nevertheless,

are determined after an individual's conception (Greiner & Rubin, 2011). A shift in focus from actual traits to perceptions can address this problem, as suggested by Greiner and Rubin (2011). Here, the treatment is not a gender switch from female to male, but a change in the perceptions about students' performance attached to their gender. As in a potential outcomes framework, we usually think of a state of the world we wish to achieve and an intervention that will get us closer to it. In this context, the goal would be that perceptions on performance are not related to a person's gender, and the mechanism proposed to achieve such a goal is changes in the percentage of female peers the student is exposed to.

I argue that the treatment is independent of potential outcomes and unobservables. I do this by analyzing how observables relate to treatment and extend this reasoning to the unobservables. Additionally, I present an autoregressive model to check for within major-by-branch gender peer composition trends.

I first check how the treatment is correlated to some of the observable covariates in Table 2 by regressing the covariate on the treatment using major-by-branch fixed effects. They are either not significant or practically zero, which is indicative that, at least in observables, the exogeneity assumption holds. Although certainly not conclusive, this is an indication that unobservables (conditional on major-by-branch fixed effects) might be perpendicular to the treatment as well.

As shown in Figure 2, the percentages of female peers are similar year to year. Nevertheless, the specification used is *conditional* on the major-by-branch, meaning that although the average level of percentage female peers might be predictable, the idiosyncratic change can be thought of as random. Therefore, as it is hard for students to manipulate the percentage of female peers they are exposed to by unilaterally switching to other majors (both because they do not know the exact number until they start classes when the enrollment process is finished and because most majors do not have a substitute that is accessible to the students), assuming that the small changes within major-by-branch year-to-year are random is plausible. If this were not the case, we would see that the percentage of female peers in one year is a good predictor of the percentage of female peers in the next. To check if the idiosyncratic variation in gender peer composition within programs has any trends, I estimate an autoregressive model of gender peer composition with major-by-branch fixed effects:

$$\%female_{tk} = \beta_0 + \beta_1 \%female_{(t-1)k} + \delta_k + \varepsilon_{tk} \tag{4}$$

Table 3 shows the results from estimating equation 4. In this context, peer composi-

Table 2: Balance tests

| | *Dependent variable:* | |
|---|---|---|
| | % of Female Peers (Treatment) | |
| | STEM | Non-STEM |
| Age ($\overline{X} = 21.4$) | -0.0001$^{**}$ | $-0.00003$ |
| | (0.00004) | (0.0001) |
| Diagnostic score ($\overline{X} = 46.5$) | 0.00002$^{**}$ | $-0.00000$ |
| | (0.00001) | (0.00002) |
| Mothers' education ($\overline{X} = 0.64$) | $-0.001^{***}$ | $-0.001$ |
| | (0.0004) | (0.001) |
| Works ($\overline{X} = 0.59$) | $-0.010^{***}$ | $-0.002$ |
| | (0.0004) | (0.001) |
| Has financial aid ($\overline{X} = 0.67$) | $-0.001^{*}$ | $-0.004^{***}$ |
| | (0.0004) | (0.001) |
| Night Shift ($\overline{X} = 0.3$) | $-0.002^{***}$ | $-0.002$ |
| | (0.0004) | (0.002) |
| Observations | 44,234 | 59,912 |
| Major-by-branch & Year Fixed Effects | Yes | Yes |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | |

Errors clustered to the Year and Major level

Table 3: Autocorrelation model for percentage of female peers

| | Percentage Female (t) |
|---|---|
| Percentage Female (t-1) | -0.082 |
| | 0.056 |
| Major by Branch and Year FE | Yes |
| Observations | 424 |

*Notes:* + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

tion in the previous year is not significantly correlated to gender peer composition in the current year, supporting the idea that within major-by-branch, the change of percentage of female peers is idiosyncratic.

## 2.2 Relevant Variation in the Treatment for Interpretation

To validly interpret the results of my fixed effects model, I need to find a plausible hypothetical change in the percentage of female peers supported in the data. To do so, I will use within-unit variation to motivate counterfactuals when discussing the substantive impact of the treatment (Mummolo & Peterson, 2018)

To identify a benchmark for an increase of female percentage that is feasible and has support in the data, I analyze the within-unit ranges of treatment for the major-by-branch units. Figure 3 shows the ranges of treatment: the mean difference between the maximum and minimum percentage of female peers within the 106 STEM Major-by-Branch units, shown in blue, is 11 percentage points (p.p.), with a standard deviation of 6 p.p. To interpret the results, I will use 10 p.p. as a benchmark for a feasible change in treatment.

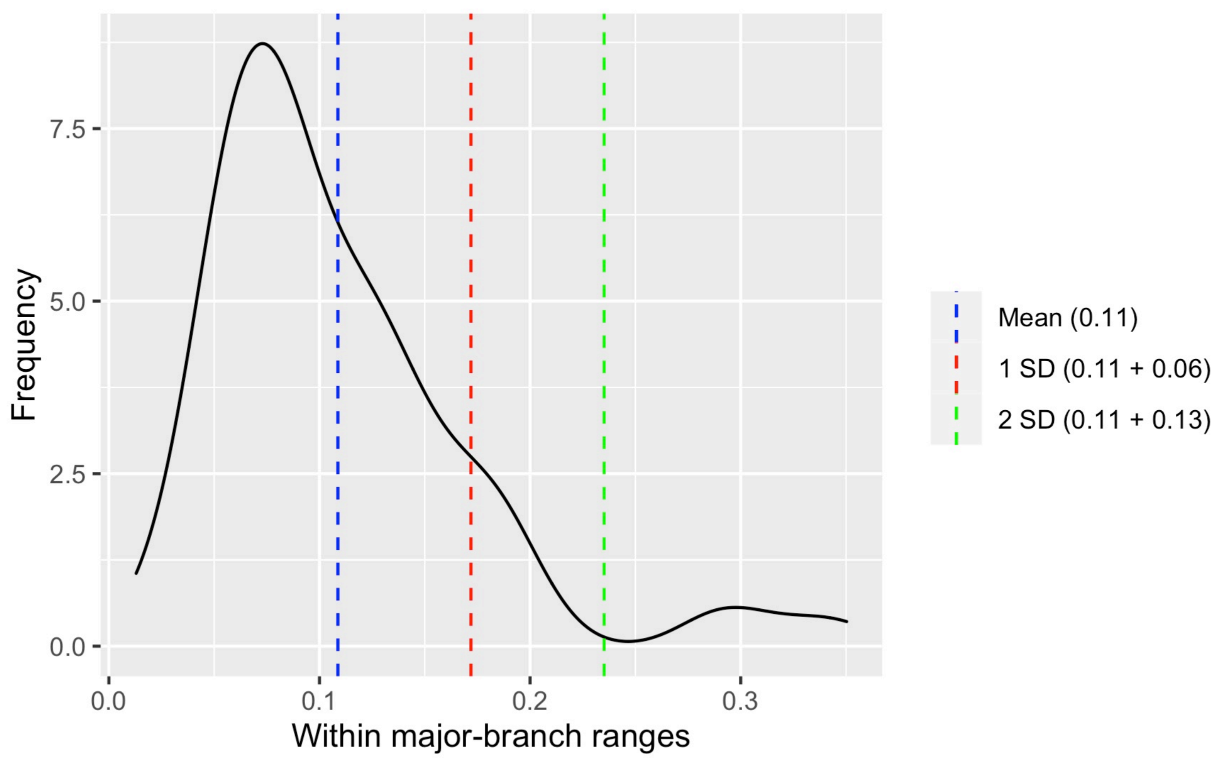


Figure 3: Within major-by-branch ranges of treatment

# 3 Results

Tables 4 and 5 report the estimation of the fixed effects model described in equation 3, for two outcomes: standardized GPA and first-year dropout. The sample includes 129,378 observations for dropout and 126,088 for standardized GPA, divided into STEM Major and Non-STEM Major students. Regressions shown in the columns (1) and (4) include Major-by-Branch and year fixed effects, but no controls or Major-by-Branch linear time trends. The rest of the estimations include student-level controls (age, diagnostic test score, mother's education, working status, funding status, and night shift) and a dummy variable for each control that takes a value of 1 if the control is missing for that student. Besides controls and year fixed effects and Major-by-Branch fixed effects, columns (3) and (6) include Major-by-Branch linear time trends. The errors are clustered at the Major-by-Branch and year level. The interaction term allows comparing the coefficient for percentage female for men and women in each sample.

Using dropout as the outcome, the coefficient of interest for women in STEM majors is negative and significant at the 5% level. As shown in Table 4, the coefficient for percentage of female peers for women in STEM remains negative and significant for the three specifications. Estimates for the model with controls and major by branch fixed effects and time trends (column (3)) indicate that an increase of 10 p.p. on the percentage of women within a STEM Major-by-Branch causes a decrease of 1.9 percentage points in female students' dropout rate. This reduction represents a 9.6% decrease in dropout rates for women in STEM majors. For men in STEM majors, the effect of the percentage of women on dropout is small and is not statistically significant. Similarly, for students in Non-STEM majors, the coefficient of interest on dropout is small and non-significant. This suggests that gender peer composition does not affect dropout or GPA for Non-STEM major students.

In the case of standardized GPA as the outcome, the estimations showed in Table 5 are consistent with what is observed for dropout. For STEM majors, the coefficients of all

Table 4: Estimates of the Effect of Percentage Female on Dropout

| | STEM Majors | | | Non-STEM Majors | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Percentage Female | -0.208** | -0.200** | -0.179* | -0.006 | 0.006 | 0.014 |
| | (0.068) | (0.070) | (0.069) | (0.028) | (0.029) | (0.028) |
| Male | -0.009 | 0.264* | 0.252* | 0.046** | 0.111 | 0.104 |
| | (0.011) | (0.122) | (0.124) | (0.011) | (0.100) | (0.100) |
| Percentage Female:Male | 0.138** | 0.145** | 0.155** | 0.001 | 0.019 | 0.026 |
| | (0.050) | (0.048) | (0.048) | (0.018) | (0.017) | (0.017) |
| Mean outcome for women | 0.186 | 0.186 | 0.186 | 0.146 | 0.146 | 0.146 |
| Major by Branch and Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | No | Yes | Yes | No | Yes | Yes |
| Major by Branch Time Trend | No | No | Yes | No | No | Yes |
| Observations | 52316 | 52316 | 52316 | 77062 | 77062 | 77062 |

*Notes:* + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Controls include age, diagnostic scores, mother's education work status, financial aid status, education shift, and missing dummies for all the variables. All controls are interacted with the gender dummy. Errors are clustered at the major by branch by year level.

models are positive and significant. For the model with all fixed effects and time trends (column (3)), the coefficient of interest is significant to the to the 5% level. Here, a 10 p.p. increase of the percentage of female peers within the Major-by-Branch is related to a 0.05 standard deviation increase in GPA.

Some could argue that the problem is not some particular gender dynamic in STEM majors but male-concentrated majors. If this was the case, we would observe a decrease on dropout and increase in standardized GPA of a higher female percentage for women on male-concentrated non-STEM majors. This can be explored with the data, comparing estimates of the main specification for male concentrated STEM majors versus male concentrated non-STEM majors [5]. The results of estimating the main specification (equation 3) for STEM male-concentrated majors and non-STEM male concentrated majors are

[5] Male concentrated majors are defined as majors that have less than 40% female students in their student body. Alternative definitions of male concentration are shown in Appendix B.

Table 5: Estimates of the Effect of Percentage Female on GPA (SD)

| | STEM Majors | | | Non-STEM Majors | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Percentage Female | 0.419+ | 0.452* | 0.555** | 0.088 | 0.087 | 0.047 |
| | (0.213) | (0.212) | (0.192) | (0.097) | (0.099) | (0.096) |
| Male | -0.066* | -0.225 | -0.215 | -0.294** | -0.667** | -0.606** |
| | (0.030) | (0.415) | (0.417) | (0.029) | (0.215) | (0.221) |
| Percentage Female:Male | -0.387** | -0.346* | -0.397** | 0.093+ | 0.062 | 0.041 |
| | (0.145) | (0.140) | (0.128) | (0.048) | (0.047) | (0.046) |
| Mean outcome for women | -0.047 | -0.047 | -0.047 | 0.225 | 0.225 | 0.225 |
| Major by Branch and Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | No | Yes | Yes | No | Yes | Yes |
| Major by Branch Time Trend | No | No | Yes | No | No | Yes |
| Observations | 50756 | 50756 | 50756 | 75332 | 75332 | 75332 |

*Notes:* + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. GPA is measured as a students' cohort percentile (0-100). Controls include age, diagnostic scores, mother's education work status, financial aid status, education shift, and missing dummies for all the variables. All controls are interacted with the gender dummy. Errors are clustered at the major by branch by year level.

shown in Table 6 for both outcomes.

In the estimations for STEM male-concentrated majors, we observe that the sign of the effect reflected in the coefficient of the percentage of female peers remains the same and is still significant for both outcomes. On the other hand, for Non-STEM male-concentrated majors, the sign flips for both coefficients, and they are statistically insignificant. To formally compare both models, I test the hypothesis that the coefficients are equal by doing a Wald test. For dropout, the difference between both coefficients is significant to the 5%, whereas that for GPA (SD), the difference is significant to the 11%.

This piece of evidence suggests that having a higher percentage of female peers is associated with higher GPA, and lower dropout is not an artifact of majors being mostly male-concentrated but of other mechanisms that seem to be unique to STEM majors.

Table 6: Estimates of the Effect of Percentage of Female Peers in Male-Concentrated Majors (40%)

| | Dropout | | | GPA (SD) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | STEM | Other | Δ | STEM | Other | Δ |
| Percentage Female | -0.206** | 0.073 | -0.279* | 0.658** | -0.093 | 0.751 |
| | (0.076) | (0.114) | | (0.196) | (0.424) | |
| Male | 0.270* | -0.004 | 0.274 | -0.264 | 0.074 | -0.338 |
| | (0.124) | (0.112) | | (0.418) | (0.437) | |
| Percentage Female:Male | 0.187** | 0.014 | 0.173+ | -0.385** | 0.220 | -0.605* |
| | (0.060) | (0.078) | | (0.128) | (0.278) | |
| Mean outcome for women | 0.186 | 0.146 | | -0.047 | 0.225 | |
| Major by Branch and Year FE | Yes | Yes | | Yes | Yes | |
| Controls | Yes | Yes | | Yes | Yes | |
| Major by Branch Time Trend | Yes | Yes | | Yes | Yes | |
| Observations | 51060 | 11343 | | 49537 | 11089 | |

*Notes:* + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. GPA is measured as a students' cohort percentile (0-100). Only male-concentrated (less than .4 women) majors are included in the estimations. Controls include age, diagnostic scores, mother's education work status, financial aid status, education shift, and missing dummies for all the variables. All controls are interacted with the gender dummy. Errors are clustered at the major by branch by year level. Δ represents the difference between the coefficients of the variable for STEM and non-STEM (other) students. P-values are calculated using a wald test testing the hypothesis that the coefficients are equal.

# 4 Mechanisms

There are several potential explanations of why women improve their academic outcomes when surrounded by more women. One channel that has been studied extensively is the role model effect (Porter & Serra, 2020). If we think of teachers as role models, and if students identify themselves more with same-sex role models, performance may be enhanced when students are assigned to a same gender teacher (Dee, 2007). In this section, I explore the role-model channel, through which having a higher percentage of female instructors may interact with gender peer effects. If that was the case, then the gender peer effect I observe in this context is different depending on the percentage of female instructors of the student. If having more female instructors is a substitute for having a higher percentage of female peers, then the coefficient on the interaction term and the share of female coefficient would have opposite signs. If, on the other hand, these are complements, then both coefficients would have the same sign as having a higher proportion of female instructors would magnify the effect of having a higher share of female peers. To shed light on this idea, I estimate the original model, adding an interaction term between the percentage of female peers and percentage of female instructors:

$$\begin{aligned} Y_{itk} = \beta_0 + \beta_1 \%female_{tk} + \beta_2 Male_{itk} + \beta_3 \%female_{tk} * Male_{itk} + \beta_4 \%femaleinstructors_{itk} + \\ \beta_5 \%female_{tk} * \%femaleinstructors_{itk} + \beta_6 X_{itk} + \gamma_t + \delta_k + \psi_k year_{st} + \varepsilon_{itk} \end{aligned} \tag{5}$$

Here, the coefficient of interest is the marginal effect of the percentage of female peers for women, which will include a component of the percentage of female instructors:

$$\frac{\partial Y_{itk}}{\partial \%female_{tk}} = \beta_1 + \beta_3 * 0 + \beta_5 \%femaleinstructors_{itk} \tag{6}$$

And its standard errors are composed by the variance of percentage of female peers

($\%female_{tk}$), variance of female instructors ($\%femaleinstructors_{itk}$), the covariance between both variables, and the value of the share of female instructors $\%femaleinstructors_{itk}$:

$$\hat{\sigma}_{\frac{\partial Y_{itk}}{\partial \%female_{tk}}} = \sqrt{var(\hat{\beta}_1) + \%femaleinstructors_{itk}^2 * var(\beta_5) + 2 * \%femaleinstructors_{itk} * cov(\hat{\beta}_1, \hat{\beta}_3)} \tag{7}$$

Given the standard error of the marginal effect shown in equation 7, if the covariance between two variables is negative, then it is entirely positive that the linear combination is significant for substantively relevant values of $\%femaleinstructors_{itk}$ even if the model parameters are insignificant [6].

The estimates of model 5 can be found in Table 7. For easier interpretation, I only show the coefficients relevant to female students. When the *%FemaleInstructors* takes the value of the mean percentage of female instructors that students in STEM have (30%), the marginal effect of *%Female* in women's dropout is -17 percentage points ($-0.263 + 0.3 * 0.309$), and it is significant at the 5% level. The opposite signs suggest that as women have a higher percentage of female instructors, the gender peer effect caused by having more female peers decreases: if instead of having 30% female instructors, women had 35% of female instructors, the marginal effect of *%Female* in dropout would drop from -17 to -15.5 percentage points ($-0.263 + 0.35 * 0.309$).

In the case of GPA, we also observe that the percentage of female instructors seems to be a substitute for having a higher share of female peers. Here, when the *%FemaleInstructors* takes the value of the mean percentage of female instructors that students in STEM have (30%), the marginal effects of *%Female* in GPA is 0.55 standard deviations ($0.668 - 0.3 * 0.4$), and it is significant at the 1% level. Similarly to what I observe in dropout, the opposite signs suggest that as students have a higher percentage of female instructors, the

[6] For more information and practical examples of this, see (Brambor, Clark, & Golder, 2006)

gender peer effect caused by having more female peers decreases: if instead of having 30% female instructors, the students had 35% of female instructors, the marginal effect of $\%Female$ in GPA would decrease from 0.55 to 0.53 percentage points $(0.668 - 0.35 * 0.4)$.

Table 7: Percentage of Female Instructor Interaction

| | Dropout | | | GPA (Standardized) | | |
|---|---|---|---|---|---|---|
| | New Model | | Original Model | New Model | | Original Model |
| % Female | -0.312** | -0.263** | -0.179* | 0.783** | 0.668** | 0.555** |
| | (0.084) | (0.081) | (0.069) | (0.243) | (0.245) | (0.192) |
| % Female Instructors | -0.100** | -0.102** | | 0.242* | 0.191* | |
| | (0.035) | (0.036) | | (0.096) | (0.097) | |
| % Female:% Female Instructors | 0.366** | 0.309* | | -0.948* | -0.400 | |
| | (0.122) | (0.135) | | (0.407) | (0.440) | |
| Mean outcome for women | 0.186 | 0.186 | 0.186 | -0.047 | -0.047 | -0.047 |
| Major by Branch and Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Major by Branch Time Trend | No | Yes | Yes | No | Yes | Yes |
| Observations | 51256 | 51256 | 52316 | 49730 | 49730 | 50756 |

*Notes:* + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. GPA is measured as a students' cohort percentile (0-100). Controls include age, diagnostic scores, mother's education work status, financial aid status, education shift, and missing dummies for all the variables. All controls are interacted with the gender dummy. Errors are clustered at the major by branch by year level.

# 5 Discussion

The results presented in this paper support the hypothesis that there are gender peer effects in the vocational education system. In particular, having a higher percentage of female peers positively affects students in STEM majors, decreasing women's dropout rates and GPA in vocational post-secondary education. For men, a higher percentage of female peers decreases dropout and increases GPA on a on a smaller scale, and the coefficients are not statistically significant. The evidence presented in this paper suggests that both women benefit from having a higher percentage of female peers and men are, at least, not harmed by it. This has a two-fold implication: first, it supports the scholarship that has established that a higher share of female peers is associated with better outcomes (Lavy & Schlosser, 2011). Second, it challenges recent evidence that students' outcomes are hindered by having a higher share of opposite gender schoolmates (Hill, 2017).

In terms of policymaking, this paper presents evidence that gender peer effects exist in vocational education and that they might point to an intervention path to "stop the leaking" in the STEM sector. The heterogeneity analysis also suggests that increasing the percentage of female instructors could improve female students' outcomes by substituting the percentage of female peers. Although post-secondary vocational institutions cannot directly control the gender composition of cohorts in this context, they do have discretion in instructor selection. Therefore, these results suggest that increasing the percentage of female instructors is an avenue to improve female students' outcomes when increasing the percentage of female peers is not feasible. Nevertheless, more causal evidence on this mechanism is necessary to affirm that role models will benefit women in this context. As scholars and institutions develop analyses of gender peer composition, collecting data and using causal inference methodologies will allow identifying levers to avoid gender polarization in STEM. From a theory perspective, this paper proposes that in the case of vocational education, having a higher percentage of female students represents a Pareto improvement: both women and men benefit from it, or at the very least are not harmed by it. Although the literature has studied the effect on women extensively, there is little evidence on the impact that gender composition has on men, an important point when thinking about the general welfare of students.

A big question when implementing policies that improve women's outcomes in education is the existence of a trade-off between improvements for women and men. The estimations presented in this paper provide evidence that for STEM majors, an increase of women within major-by-branches would benefit women and would not harm men. This is an essential point for policy-makers and institutions that want to push for policies and strategies to increase female participation in STEM programs.

This paper builds on the gender peer effect literature, providing evidence for a novel context (a middle-income country in Latin America) for an education sector that has not been thoroughly studied: post-secondary vocational education. It provides strong evidence that the international trends of female student achievement and peer effects hold in this context and that actions geared toward improving gender balance within majors can positively affect students' outcomes. To understand better the actions that could improve gender balance in this context, the next step should be to explore the mechanisms at play that create these dynamics experimentally.

# References

Arango, M., Evans, S., & Quadri, Z. (2016, January). *Education Reform in Chile* (Tech. Rep.). Woodrow Wilson School of Public Policy.

Bostwick, V., & Weinberg, B. (2018, September). *Nevertheless She Persisted? Gender Peer Effects in Doctoral STEM Programs* (NBER Working Paper No. 25028).

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis*, *14*(1), 63–82. Retrieved 2019-10-31, from https://www.cambridge.org/core/product/identifier/S1047198700001297/type/journal_article doi: 10.1093/pan/mpi014

Busso, M., & Frisancho, V. (2021, May). Good Peers Have Asymmetric Gendered Effects on Female Educational Outcomes: Experimental Evidence from Mexico. *IDB Working Papers Series*(1220).

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. , 27.

Duryea, S., Galiani, S., Nopo, H., & Piras, C. C. (2007). The Educational Gender Gap in Latin America and the Caribbean. *SSRN Electronic Journal*. Retrieved 2018-12-01, from http://www.ssrn.com/abstract=1820870 doi: 10.2139/ssrn.1820870

Goldin, C. (2002, April). *The Rising (and then Declining) Significance of Gender* (Tech. Rep. No. w8915). Cambridge, MA: National Bureau of Economic Research. Retrieved 2018-11-05, from http://www.nber.org/papers/w8915.pdf doi: 10.3386/w8915

Goldin, C., Katz, L. F., & Kuziemko, I. (2006). The Homecoming of American College Women: The Reversal of the College Gender Gap. *Journal of Economic Perspectives*, *20*(4), 133–156.

Greiner, D. J., & Rubin, D. B. (2011, August). Causal Effects of Perceived Immutable Characteristics. *Review of Economics and Statistics*, *93*(3), 775–785. Retrieved

2018-11-30, from http://www.mitpressjournals.org/doi/10.1162/REST_a_00110 doi: 10.1162/REST_a_00110

Han, L., & Li, T. (2009, February). The gender difference of peer influence in higher education. *Economics of Education Review*, *28*(1), 129–134. Retrieved 2018-11-05, from http://linkinghub.elsevier.com/retrieve/pii/S0272775708000575 doi: 10.1016/j.econedurev.2007.12.002

Hill, A. J. (2017, January). The positive influence of female college students on their male peers. *Labour Economics*, *44*, 151–160. Retrieved 2019-10-14, from https://linkinghub.elsevier.com/retrieve/pii/S0927537117300581 doi: 10.1016/j.labeco.2017.01.005

Hoxby, C. M. (2000, August). Peer Effects in the Classroom: Learning From Gender and Race Variation. *NBER Working Paper Series*, *7867*. Retrieved from https://www.nber.org/papers/w7867.pdf

Lavy, V., & Schlosser, A. (2011, April). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, *3*(2), 1–33. Retrieved 2018-05-04, from http://pubs.aeaweb.org/doi/10.1257/app.3.2.1 doi: 10.1257/app.3.2.1

Lundberg, S. (2017). CSWEP Annual Report to the Ameri- can Economic Association. , 16.

Mouganie, P., & Wang, Y. (2020). High-Performing Peers and Female STEM Choices in School. *Journal of Labor Economics*, *38*(3), 805–841.

Mummolo, J., & Peterson, E. (2018, October). Improving the Interpretation of Fixed Effects Regression Results. *Political Science Research and Methods*, *6*(04), 829–835. Retrieved 2018-12-20, from https://www.cambridge.org/core/product/identifier/S2049847017000449/type/journal_article doi: 10.1017/psrm.2017.44

Paredes, V. (2018, September). Gender Gaps in Single-Sex Classrooms. *Serie de Documentos* de *Trabajo, FEN*, 32.

Porter, C., & Serra, D. (2020, July). Gender Differences in the Choice of Major: The Importance of Female Role Models. *American Economic Journal: Applied Economics*, *12*(3), 226–254. Retrieved 2020-11-29, from https://pubs.aeaweb.org/doi/10.1257/app.20180426 doi: 10.1257/app.20180426

Sacerdote, B. (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In *Handbook of the Economics of Education* (Vol. 3, pp. 249–277). Elsevier. Retrieved 2018-11-05, from http://linkinghub.elsevier.com/retrieve/pii/B9780444534293000041 doi: 10.1016/B978-0-444-53429-3.00004-1

Schneeweis, N., & Zweimüller, M. (2012, August). Girls, girls, girls: Gender composition and female school choice. *Economics of Education Review*, *31*(4), 482–500. Retrieved 2018-11-05, from http://linkinghub.elsevier.com/retrieve/pii/S0272775711001749 doi: 10.1016/j.econedurev.2011.11.002

*Skills beyond school: synthesis report*. (2014). Paris: OECD Publishing. (OCLC: ocn891126521)

Wu, A. H. (2017). Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum. *SSRN Electronic Journal*. Retrieved 2018-11-05, from https://www.ssrn.com/abstract=3051462 doi: 10.2139/ssrn.3051462

Zölitz, U., & Feld, J. (2020, March). The Effect of Peer Gender on Major Choice in Business School. *Management Science*, mnsc.2020.3860. Retrieved 2021-09-29, from http://pubsonline.informs.org/doi/10.1287/mnsc.2020.3860 doi: 10.1287/mnsc.2020.3860

# Appendices

## A Identifying variation plots

Figure 4: Within major-by-branch Yearly percentage female variation

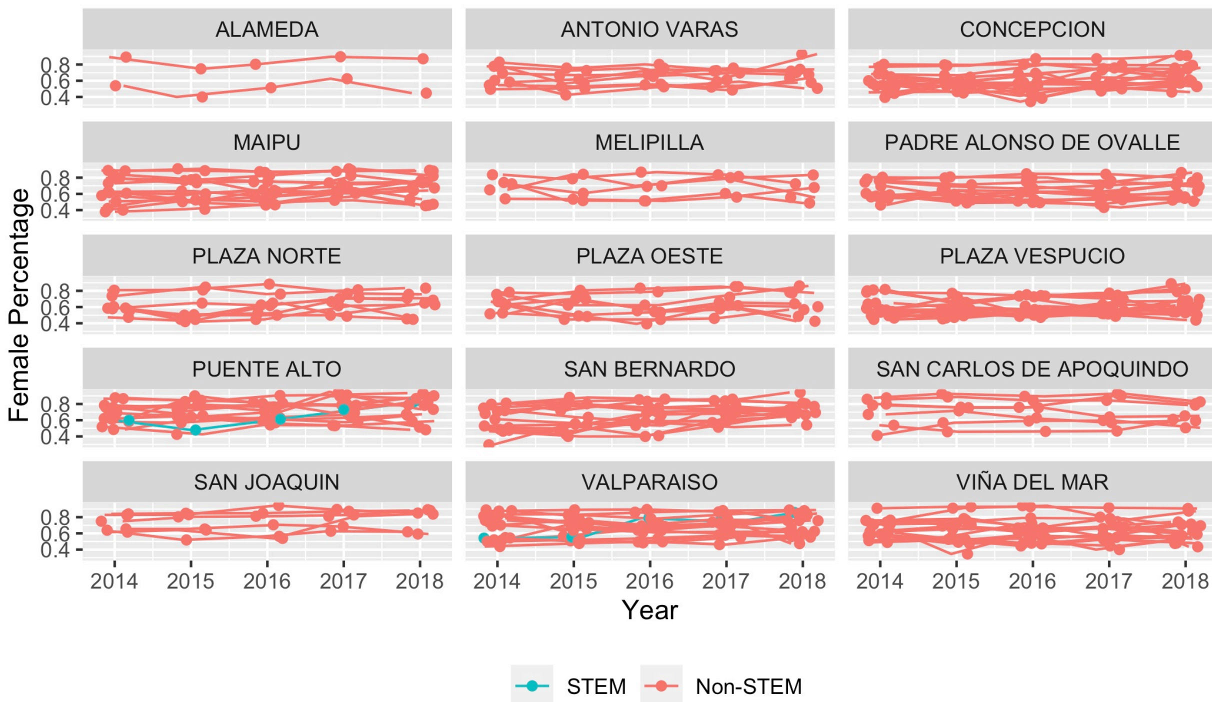


Figure 4: Within major-by-branch Yearly percentage female variation

# B Alternative Male Concentrations Estimations

Table 8: Estimates of the Effect of Percentage Female in Male-Concentrated Majors (35%)

| | Dropout | | | GPA (SD) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | STEM | Other | Δ | STEM | Other | Δ |
| Percentage Female | -0.142 | 0.420* | -0.562** | 0.564* | -0.809 | 1.373* |
| | (0.094) | (0.166) | | (0.251) | (0.648) | |
| Male | 0.253 | 0.026 | 0.227 | -0.135 | -0.188 | 0.053 |
| | (0.161) | (0.170) | | (0.541) | (0.570) | |
| Percentage Female:Male | 0.129 | -0.279* | 0.408** | -0.309 | 0.763 | -1.072* |
| | (0.083) | (0.127) | | (0.199) | (0.466) | |
| Mean outcome for women | 0.186 | 0.146 | | -0.047 | 0.225 | |
| Major by Branch and Year FE | Yes | Yes | | Yes | Yes | |
| Controls | Yes | Yes | | Yes | Yes | |
| Major by Branch Time Trend | Yes | Yes | | Yes | Yes | |
| Observations | 49127 | 8292 | | 47677 | 8092 | |

*Notes:* + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. GPA is measured as a students' cohort percentile (0-100). Only male-concentrated (less than .35 women) majors are included in the estimations. Controls include age, diagnostic scores, mother's education work status, financial aid status, education shift, and missing dummies for all the variables. All controls are interacted with the gender dummy. Errors are clustered at the major by branch by year level. Δ represents the difference between the coefficients of the variable for STEM and non-STEM (other) students. P-values are calculated using a wald test testing the hypothesis that the coefficients are equal.

Table 9: Estimates of the Effect of Percentage Female in Male-Concentrated Majors (30%)

| | Dropout | | | GPA (SD) | | |
|---|---|---|---|---|---|---|
| | (1) STEM | (2) Other | (3) Δ | (4) STEM | (5) Other | (6) Δ |
| Percentage Female | -0.106 | 0.642* | -0.748** | 0.493+ | -1.266 | 1.759 |
| | (0.100) | (0.236) | | (0.282) | (1.099) | |
| Male | 0.259 | -0.002 | 0.261 | -0.146 | -0.188 | 0.042 |
| | (0.161) | (0.178) | | (0.547) | (0.563) | |
| Percentage Female:Male | 0.056 | -0.329 | 0.385+ | -0.170 | 0.891 | -1.061 |
| | (0.086) | (0.214) | | (0.224) | (0.998) | |
| Mean outcome for women | 0.186 | 0.146 | | -0.047 | 0.225 | |
| Major by Branch and Year FE | Yes | Yes | | Yes | Yes | |
| Controls | Yes | Yes | | Yes | Yes | |
| Major by Branch Time Trend | Yes | Yes | | Yes | Yes | |
| Observations | 48754 | 5862 | | 47315 | 5709 | |

*Notes:* + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. GPA is measured as a students' cohort percentile (0-100). Only male-concentrated (less than .3 women) majors are included in the estimations. Controls include age, diagnostic scores, mother's education work status, financial aid status, education shift, and missing dummies for all the variables. All controls are interacted with the gender dummy. Errors are clustered at the major by branch by year level. Δ represents the difference between the coefficients of the variable for STEM and non-STEM (other) students. P-values are calculated using a wald test testing the hypothesis that the coefficients are equal.

Table 10: Estimates of the Effect of Percentage Female in Male-Concentrated Majors (30%)

| | Dropout | | | GPA (SD) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | STEM | Other | Δ | STEM | Other | Δ |
| Percentage Female | -0.106 | 0.642* | -0.748** | 0.493+ | -1.266 | 1.759 |
| | (0.100) | (0.236) | | (0.282) | (1.099) | |
| Male | 0.259 | -0.002 | 0.261 | -0.146 | -0.188 | 0.042 |
| | (0.161) | (0.178) | | (0.547) | (0.563) | |
| Percentage Female:Male | 0.056 | -0.329 | 0.385+ | -0.170 | 0.891 | -1.061 |
| | (0.086) | (0.214) | | (0.224) | (0.998) | |
| Mean outcome for women | 0.186 | 0.146 | | -0.047 | 0.225 | |
| Major by Branch and Year FE | Yes | Yes | | Yes | Yes | |
| Controls | Yes | Yes | | Yes | Yes | |
| Major by Branch Time Trend | Yes | Yes | | Yes | Yes | |
| Observations | 48754 | 5862 | | 47315 | 5709 | |

*Notes:* + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. GPA is measured as a students' cohort percentile (0-100). Only male-concentrated (less than .3 women) majors are included in the estimations. Controls include age, diagnostic scores, mother's education work status, financial aid status, education shift, and missing dummies for all the variables. All controls are interacted with the gender dummy. Errors are clustered at the major by branch by year level. Δ represents the difference between the coefficients of the variable for STEM and non-STEM (other) students. P-values are calculated using a wald test testing the hypothesis that the coefficients are equal.