Nevertheless She Persisted? Gender Peer Effects in Doctoral STEM Programs

Valerie K. Bostwick*
Bruce A. Weinberg†

Abstract

We study the effects of peer gender composition in STEM doctoral programs on persistence and degree completion. Leveraging unique new data and quasi-random variation in gender composition across cohorts within programs, we show that women entering cohorts with no female peers are 11.7pp less likely to graduate within 6 years than their male counterparts. A 1 sd increase in the percentage of female students differentially increases women's probability of on-time graduation by 4.4pp. These gender peer effects function primarily through changes in the probability of dropping out in the first year of a Ph.D. program.

*Department of Economics, Kansas State University, email: vkbostwick@gmail.com.

†Department of Economics, Ohio State University; National Bureau of Economic Research; IZA Institute of Labor Economics

We thank Joshua Hawley, Brendan Price, and Christopher Severen for valuable feedback on this project. We also thank the participants in the OSU Applied Micro Seminar, the UC Davis Labor Seminar, the Broom Center for Demography at UCSB, the IZA 2nd Workshop on Gender and Family Economics, the Population Association of American 2018 Meeting, the Association for Education Finance and Policy 2018 Meeting, and the Society of Labor Economists 2017 Meeting. We gratefully acknowledge support from NIA, OBSSR, and NSF SciSIP through P01 AG039347; R01 GM140281, UL1 TR002733, NSF EHR DGE 1348691, 1535399, 1760544; and the Ewing Marion Kauffman and Alfred P. Sloan Foundations. Weinberg was supported on P01 AG039347 by the NBER directly and on a subaward from NBER to Ohio State. We are also grateful to Academic Analytics for providing data on faculty rosters for this project.

1 Introduction

The underrepresentation of women in science, technology, engineering, and mathematics (STEM) fields starts as early as grade school and intensifies at each successive career stage so that men greatly outnumber women as scientists and engineers at senior levels. A female-unfriendly climate is one cause of underrepresentation in STEM that resonates for many female scientists. In a report on the lack of women in engineering, Corbett and Hill (2015) summarize: "Stereotypes and biases lie at the core of the challenges facing women in engineering and computing. Educational and workplace environments are dissuading women who might otherwise succeed in these fields." Unfortunately, the climate in these fields has been difficult to quantify empirically and researchers have consequently struggled to estimate the impact of environment on the gender gap in STEM.

This paper studies one aspect of the environment in STEM doctoral programs[1] and its effects on gender differences in persistence and completion. Specifically, we estimate the effects of the gender composition of one's entering cohort peers within a STEM Ph.D. program on the gender gap in persistence and on-time graduation rates. To the best of our knowledge, we are the first paper to study gender peer effects in the context of doctoral programs and the production of STEM Ph.D.s. This context is noteworthy as STEM professionals are an important and growing segment of the workforce, and many of the graduate students we study will train the next generation of scientific researchers.

In general, doctoral students are an understudied group and they may be of particular interest when considering the gender gap in STEM fields. These students have made substantial investments and demonstrated commitment to pursuing a STEM career and yet are still highly likely to drop out (more than 30% of our sample drop out in the first 6 years of enrollment). Furthermore, there is reason to believe that incoming cohorts within doctoral programs provide a particularly salient peer group that may function quite differently than, say, classmates in an undergraduate STEM course.

We provide the first causal evidence of gender peer effects among Ph.D. students using a quasi-experimental identification strategy. Specifically, within a particular doctoral program at a given university, year-to-year fluctuations in the gender composition of each cohort are plausibly and, to some extent, testably exogenous.[2] This identification strategy leverages the fact that there is uncertainty, both on the part of admissions and on the part of potential doctoral students, as to the gender composition of each incoming cohort. While a doctoral program's admissions committee might target a specific gender mix and an incoming student

1. This includes doctoral programs in economics and psychology. For a full list of programs included in our estimation sample, see Table 1.

2. Hoxby (2000) employs a similar strategy for grade-school students.

might know the average gender mix of past cohorts in a program, neither party can fully anticipate the realized gender composition of an incoming cohort of students. This provides a source of plausibly exogenous variation in students' peer gender composition that allows us to identify a causal effect on the gender gap in STEM Ph.D. outcomes. Note that this approach might be applied to a variety of contexts wherein individuals are grouped into well-defined cohorts of peers (e.g. firms that hire a new cohort of recent college graduates every summer) and to other dimensions of underrepresentation.

In order to implement this strategy, we introduce a new dataset that links administrative transcript records from all public universities in the state of Ohio[3] to data from the UMETRICS project, which provides information on the research environment (e.g. source, timing, and duration of funding) for students who are supported by federal research grants. In contrast to other datasets commonly-used to study doctoral students (e.g. the Survey of Doctorate Recipients), which are restricted to Ph.D. completers, this administrative data includes all incoming students in doctoral programs and allows for longitudinal observation of program environments and drop-out behavior. Using this novel dataset, we are able to identify a relevant peer group (incoming cohorts within programs)[4] and analyze the effects of peer gender composition on both Ph.D. persistence and completion. Furthermore, the detailed transcript data and the links to UMETRICS allow us to investigate several important mechanisms for our main findings.

Utilizing the approach outlined above, we find that: (1) in cohorts with no female peers, women are 11.7pp less likely to complete a Ph.D. within 6 years than their male counterparts; (2) a 1 standard deviation increase in the share of female students in a cohort differentially increases the probability of on-time graduation for women by 4.4pp; and (3) these effects function almost completely through changes in the probability of dropping out in the first year of a Ph.D. program.

We attempt to investigate the mechanisms underlying these gender peer effects by analyzing some of the most prominent potential channels: gender differences in learning/competing or financial support. We find no evidence of any differences in financial support due to peer gender composition. Although we do find a small effect of cohort gender

3. The Ohio Longitudinal Data Archive is a project of the Ohio Education Research Center (oerc.osu.edu) and provides researchers with centralized access to administrative data. The OLDA is managed by The Ohio State University's Center for Human Resource Research (chrr.osu.edu) in collaboration with Ohio's state workforce and education agencies (ohioanalytics.gov), with those agencies providing oversight and funding. For information on OLDA sponsors, see http://chrr.osu.edu/projects/ohio-longitudinal- data-archive.

4. Even in the more "lab-based" STEM fields, doctoral students are predominantly focused on course work in the first year of these programs, making the incoming cohort a salient peer group. This is evident in the detailed transcript data and the results in Section 5 are robust to the exclusion of those few students who do not take a full-time, first-year course load.

composition on grades, these effects account for a relatively small portion of the full gender peer effect on degree completion. While we recognize that other unobserved mechanisms may be at play, our findings are consist with climate being an important mechanism driving the observed gender peer effects.

2 Literature Review

This paper offers several significant contributions to the broad literature on the gender gap in STEM fields. In the literature studying doctoral student outcomes, we are one of the first papers to test for the presence of any type of peer effects.[5] Our primary outcome variable, Ph.D. completion, is also a notably understudied outcome and the literature on the gender gap in STEM doctoral degrees is quite sparse, likely due to a lack of longitudinal data on students entering doctoral programs (Ceci et al. 2014).

The majority of the research on the outcomes of Ph.D. students is in the education literature (Gardner et al. 2009; Nettles 1990; Golde 2005) with a smaller line in the economics literature.[6] Much of the education research focuses on mentoring (Clark, Harden, and Johnson 2000; Hall and Burns 2009; Bell-Ellison and Dedrick 2008; Main 2014) and professional skills development and socialization (Nerad 2004; Golde and Dore 2001). Early work in the economics literature focuses on the relationship between financial support and Ph.D. completion. The literature finds that financial support, and especially fellowships and research assistantships, is highly correlated with Ph.D. completion (Abedi and Benkin 1987; Ehrenberg and Mavros 1995).

Previous work that investigates the interaction between gender and doctoral success has primarily focused on the impact of same-gender mentors. The findings on same-gender mentorship are mixed. Both Neumark and Gardecki (1998) and Hilmer and Hilmer (2007) find that female advisors have no effect on labor market outcomes for female doctoral students in economics. However, Neumark and Gardecki (1998) do find evidence that female advisors reduce the time spent in graduate school for female students and both Pezzoni et al. (2016) and Gaule and Piacentini (2017) find that female doctoral students in STEM fields who have female advisors are more successful in terms of publishing than female students with

5. One exception is a recent paper by Pezzoni et al. (2016) that finds no evidence of a correlation between graduate students' publication levels and the gender composition of research teams.

6. There is also some literature from reserachers in psychology providing evidence of gender bias in STEM graduate program *admissions*. Moss-Racusin et al. (2012) and Milkman, Akinola, and Chugh (2015) each employ audit studies to reveal that STEM faculty members rate applicants to graduate programs as significantly more competent and are more likely to respond to email correspondence when a prospective student is assigned to a male name.

male advisors. Pezzoni et al. (2016) also use the UMETRICS dataset to measure the gender composition of research teams at the California Institute of Technology and find no evidence of a relationship between team gender composition and publication outcomes for graduate students. While these studies provide very interesting descriptive results, our paper focuses on identifying a causal relationship between the doctoral environment and the gender gap in STEM Ph.D. persistence.

Our approach is more similar to papers in the literature on undergraduate students, which have focused on the effects of peer characteristics on the gender gap in STEM major choice.[7] A number of papers have shown that peer gender composition in high school can have an effect on college major choice (Anelli and Peri 2016; Brenøe and Zölitz 2018; Mouganie and Wang 2019). Both Kugler, Tinsley, and Ukhaneva (2017) and Astorne-Figari and Speer (2017) find that the gender composition of majors is correlated with female students' major choice and that women are more likely to switch out of male-dominated fields. Zölitz and Feld (2017) exploit random assignment of students to teaching sections and find that women exposed to more female peers are less likely to choose a male-dominated major. On the other hand, Hill (2017) finds that the gender composition of entire university cohorts can impact major choice; specifically that women exposed to more female peers are likely to choose female-dominated majors. Our paper builds upon these findings by studying a population of students who have surpassed the many hurdles of completing a STEM bachelor's degree and who are demonstrably committed to pursuing a STEM career. The doctoral setting also allows us to study exogenous variation in the gender composition of a peer group that is more comprehensive, meaningful, and salient to these students than groups that others have studied such as classmates in a large undergraduate course or roommates.

This paper is also related to the emerging literature on climate and female underrepresentation in STEM, which has typically relied on descriptive survey results. However, those survey findings clearly point to a negative impact of the workplace environment on female scientists persistence in STEM fields.[8] Fouad and Singh (2011) report that "One-in-three women left [engineering jobs] because they did not like the workplace climate, their boss or the culture." Other studies find that, after controlling for both individual and occupation characteristics, women were more likely to report being unsatisfied with their jobs (Lordan and Pischke 2016) and are more likely to leave the field entirely (Hunt 2016) when the share of men in an occupation/field is higher.

7. In another related paper that does *not* focus on STEM fields, Huntington-Klein and Rose (2018) study West Point undergraduates who are randomly assigned to peer groups and find that women with more female peers are more likely to persist at that institution.

8. Corbett and Hill (2015) provide a comprehensive report on the underrepresentation of women in engineering and the need for changes to college and workplace environments.

These issues have been particularly salient in economics recently. New papers and reports reveal that a toxic workplace culture may be contributing to female underrepresentation (Dupas et al. 2021; Lundberg et al. 2018; Wu 2017) and that female economists face many systematic barriers to success (Hengel 2017; Mengel, Sauermann, and Zölitz 2017; Sarsons 2017).[9] Wu (2017) analyzes comments from a well-known and anonymous online forum of economists and finds evidence of negative gender stereotyping towards female economists and their research. This is one of the few papers in the literature to quantify and provide concrete evidence of the negative climate towards women in academia. Our paper builds upon this work by providing a measure of one aspect of the environment within STEM fields and, importantly, by identifying and estimating the effects of this measure on the educational outcomes of female doctoral students.

3 Data

Our primary data include two linked administrative sources. The Ohio Longitudinal Data Archive (OLDA) provides administrative transcript records for all students attending public colleges and universities in Ohio between Summer/Fall 2005 and Spring 2016. This data includes student demographics, a doctoral program identifier, degree completions, and course-level data on enrollment and grades. The second source of data is provided by the UMETRICS project, which contains the university payroll records on all individuals employed under federal research grants at one university in Ohio. This data provides month-level information on research grant employment over the period of June 2009 to June 2015 for all graduate students at this university.

Using the OLDA enrollment data, we construct a panel of students that encompasses all individuals who first enrolled in a doctoral program at the main campus of any public 4-year university in Ohio[10] in the Summer/Fall terms of 2005-2015.[11] The enrollment data combined with degree completions allow us to identify students who drop out and to measure persistence to year 2, 3, etc. of the doctoral program.

Each student's doctoral program identifier code in OLDA is linked to a Classification of Instructional Programs (CIP) code[12]. We define a doctoral "program" to include all

9. For a comprehensive overview of the current state of underrepresentation in economics and the exisiting research on the causes of that underrepresentation, see Bayer and Rouse (2016).

10. We exclude students enrolled at the Medical University of Ohio and Youngstown State University due to very small sample sizes.

11. We exclude students who first enroll in a Winter or Spring terms and treat students who first enroll in the Summer term as having enrolled in the following Fall.

12. https://nces.ed.gov/ipeds/cipcode

students attending the same institution with the same enrollment CIP code[13] and define a "cohort" to be all students who first enrolled in a given program in the same year. Note that CIP codes are more specific than broad fields such that, within a given field at the same institution there may be multiple doctoral programs. For example, within the field of Chemistry the same university may have three separate doctoral programs in General Chemistry, Polymer Chemistry, and Chemical Physics. As our primary variable of interest is cohort gender composition, we limit the sample by dropping those students who first enroll in a non-doctoral graduate program and then transfer into a doctoral program at the same institution, as it is not clear to which cohort they belong. If these dropped transfer students encompass more than 20% of the enrollment for a particular program (over all years), then we also drop that program from the sample.[14]

We impose three key restrictions in order to create the final estimation sample. First, we restrict the data to those cohorts for whom we can observe 6 complete years of transcript data: cohorts starting in 2005-2009. This is because our primary dependent variable for this analysis is the probability of completing a Ph.D. within 6 years of initial enrollment. Second, we exclude programs with very small cohort sizes from the sample. For each cohort in each program we calculate the cohort size (# of students) and the percent of the cohort that is female. For each program, we also calculate the average over all years (2005-2015) for both of these variables. We then exclude from the sample all programs with an average cohort size less than or equal to 9 students because very small programs may be structured in such a way that a student's incoming cohort is not the relevant peer group (i.e. in very small programs students are more likely to interact heavily with students from other cohorts).[15][16] Finally, we restrict the sample to STEM programs.[17]

The final "estimation sample" includes 2,513 student observations, grouped into an unbalanced panel of 31 doctoral programs, representing 6 public universities (University of Akron, Bowling Green State University, University of Cincinnati, Ohio State University, Ohio University, and the University of Toledo). Table 1 provides a full list of these 31

13. We aggregate to the CIP code level because the university-provided program identifiers are not consistently defined across school-years. However, our main results are robust to using the university program codes to identify individual Ph.D. programs.

14. The main results reported in Section 5 are robust to including/excluding programs with more/fewer transfer students. See Table 18.

15. The main results reported in Section 5 are robust to including/excluding programs with a smaller/larger average cohort size. See Table 17.

16. Measures of cohort gender composition also tend to be much more volatile in smaller programs. Among the smaller programs in the data, the average of the standard deviation of the proportion female is 0.27 (vs. 0.13 in the larger programs). The probability of observing a 0 or 1 value for the proportion female in a cohort is also very high in smaller programs (32% of program-year observations). In the larger programs we observe these extreme values in only 4% of program-year observations.

17. As designated by the Ohio Department of Higher Education: https://www.ohiohighered.org/node/2104

programs and their corresponding CIP codes, CIP fields, and summary statistics. The top panel of Table 2 shows the calculated cohort characteristics for the main estimation sample. For reference, we also calculate these summary statistics for the estimation sample plus all non-STEM programs and small programs in the middle panel of Table 2, and for the full sample (including all years of the data 2005-2015, non-STEM programs, and small programs) in the bottom panel. In the estimation sample, the average cohort size is approximately 17 students and the average cohort is 38% female. Unsurprisingly, the full sample is comprised of smaller cohorts, on average, than the estimation sample and is somewhat more female (due to including non-STEM programs). Table 2 also reveals a large amount of variation in cohort gender composition across programs. While the average cohort is 38% female, the standard deviation is nearly 21pp. Within programs, the standard deviation of % cohort female is nearly 13pp and the largest deviations from a program mean are -41pp and 30pp.

In order to examine the relationships between cohort gender composition, the probability of receiving financial support via research funding, and Ph.D. persistence, we link the estimation sample (expanded to include all cohorts 2005-2015) to the UMETRICS data for the subset of students who attend one UMETRICS university. This allows us to observe month-by-month employment for students paid by federal researching grants and to construct indicator variables for obtaining research funding in each year of enrollment for each student (i.e. employed for at least 28 days of the school year).

Summary statistics for the main estimation sample and for the linked UMETRICS sample are shown in Table 3. Note that the different time spans for the two linked data sources mean that each of the funding indicator variables is available for a different set of cohorts. For example, research employment in year 1 of enrollment is observed only for cohorts starting in 2009-2014, whereas funding in year 2 is observed for cohorts starting in 2008-2013. Table 3 shows that men and women in the estimation sample are quite similar in terms of demographics, grades, and graduation rates. However, male students do seem to have a somewhat higher probability of obtaining research employment in the first four years of enrollment.

4 Empirical Strategy

The primary empirical strategy is essentially a difference-in-differences approach, comparing women to men between cohorts with a high fraction of female students and cohorts with very few female students within a given doctoral program. We estimate the following model:

$$\mathbb{P}\left(Y_{ipt}=1\right) = \beta_1 Female_i + \beta_2 \% PeersFemale_{ipt} + \beta_3 Female_i * \% PeersFemale_{ipt} + \gamma' X_{ipt} + \delta Z_{pt} + D_p + D_t + \epsilon_{ipt}, \quad \text{(S1)}$$

where $Y_{ipt} = 1$ if student i, who first enrolled in program p in year t, completes a Ph.D. within 6 years. The model includes individual-level covariates, X_{ipt}, which are: age, age^2, race/ethnicity indicators, and a foreign student indicator variable. The variable Z_{pt} measures cohort size, while D_t and D_p are year and program fixed-effects, respectively.

The primary variables of interest are: $Female_i$, an indicator for own gender; $\% PeersFemale_{ipt}$, which measures the percent of student i's peers entering into program p in year t (cohort pt) who are female; and the interaction term of those 2 variables. We test the robustness of our main specification using several alternative measures of "highly female" cohorts including the number of female peers in the cohort and the ratio of women to men in the cohort.[18]

The coefficient β_1 can be interpreted as the percentage point difference in on-time graduation probabilities for a woman with no female peers versus men in a cohort with no female peers. The coefficient β_2 indicates the difference in graduation probabilities for men in a cohort with all female peers versus men in cohorts with no female peers. Finally, the coefficient β_3 is the differential effect on women versus men of being in a cohort with all female peers. The model is estimated using a Probit maximum likelihood estimator (Probit MLE),[19] thus all tables in Section 5 report the marginal effects corresponding to the descriptions above and are evaluated at the mean of all covariates. Standard errors are clustered at the program level.

Identification of the model hinges on the assumption that, within a particular doctoral program, year-to-year variation in cohort gender composition is quasi-random and not correlated with other unobservables influencing gender differences in the graduation rates for that cohort. An example violation of this assumption might be the appointment of a new director of graduate studies who simultaneously puts an emphasis on recruiting more female doctoral students while also enacting other policy changes that improve those new female students' outcomes (but not those of previously enrolled female students). Similarly, the hiring of a new female faculty member who attracts more female graduate students and improves the outcomes of those students might confound the interpretation of our estimates. A telling signal of this type of endogeneity would be any evidence of time trends in the

18. The ratio of women to men is defined as the number of women in the cohort divided by the number of men in the cohort. This measure does not exclude the treated individual. For example, if a female student is the only woman in a cohort of 5 students, the ratio variable will be 1/4.

19. Results estimated using a linear probability model are qualitatively very similar. See Table 15.

cohort gender composition within programs. In the absence of time trends, the influence of the new graduate studies director or faculty member would have to fluctuate year-to-year along with cohort gender composition in order to pose a threat to identification.[20]

Figure 1 plots the percent female in each cohort by program for the years 2005-2015. Each line in Figure 1 represents a program and the panels group those programs into broader fields. In this figure it is clear that programs in some fields (e.g. Psychology and Biology) tend to have higher percentages of female students, while fields such as Computer Engineering and Physics have very low percentages of women in any given cohort. However, it is also clear that there is considerable idiosyncratic variation in gender composition within programs over time and there do not appear to be any overall or program-specific trends in gender composition. Furthermore, an AR(1) model of gender composition with program and year fixed-effects reveals no evidence of path-dependence in this variable.[21]

Figure 2 shows that cohort gender composition is also not significantly correlated with the covariates included in model S1. In each panel, each point represents a cohort and the x-axis measures the percent of the cohort that is female minus the average percent female in the program over all years of the data. The y-axis in each panel represent a different covariate, also demeaned at the program level. These variables are: cohort size, age, foreign status, and an indicator for white race. Note that there does appear to be a negative relationship between cohort age and percent female, however this is largely driven by one outlier observation.[22]

Finally, we show through a Monte Carlo simulation exercise (following Lavy and Schlosser (2011)) that the observed within-program variation in gender composition in our data closely resembles randomly generated variation from a binomial distribution. For each doctoral program, we randomly generate the gender of the students in each cohort using a binomial distribution function with parameters: n equal to the actual cohort size and p equal to the average proportion of females in that program across all years. We then compute the within-program standard deviation of the proportion female over all years. We repeat this process over 1,000 iterations to obtain an empirical confidence interval for the standard deviation for each doctoral program. We find that our observed within-program standard

20. We further explore year-to-year variation in faculty gender composition and gender of the graduate studies director as potential channels for our main findings in Section 5.2. We find that these changes are not highly correlated with the year-to-year variation in cohort gender composition and that neither faculty gender composition nor the gender of the graduate studies director can explain the estimated impact on the gender gap in Ph.D. completion. See Table 14.

21. The Wald test statistic for the lagged % cohort female variable is -0.63.

22. The outlier observations is a cohort with an average age that is more than 13 standard deviations above the mean. All results shown in Section 5 are also robust to including a control for cohort age and an interaction between cohort age and the female indicator.

deviation lies within the empirical 90% confidence interval for 88% of Ph.D. programs in our sample. This test further supports our assumption that the within-program, year-to-year variation in cohort gender composition is, in fact, quasi-random.

5 Results

Table 4 shows the marginal effects from estimating model S1 described above. In each column we apply a different definition of highly female cohorts. In column (1) we use the preferred definition where cohort gender composition is measured by the $\%PeersFemale_{ipt}$ variable (the percent of student i's cohort peers who are female). These results show that there is a significant gender gap in Ph.D. completion in cohorts with few women. Women in a cohort with no female peers are 11.7pp less likely than their male peers to graduate within 6 years of initial enrollment. However, in highly-female cohorts, that gap closes. For each additional 10pp female students in a cohort, men are 0.86pp less likely to graduate on-time (although this effect is statistically insignificant); the differential effect on women is 2.10pp (and statistically significant at the 5% level). This indicates that the effect of an additional 10pp female students for a woman is a 1.24pp increase in the probability of graduating on-time. Another way to interpret these results is that a 1 standard deviation (20.7pp) increase in the share of female students increases the probability of on-time graduation for women relative to men by 4.35pp.

Columns (2) and (3) of Table 4 experiment with alternative measures of cohort gender composition. Column (2) defines cohort gender composition as the ratio of women to men in the cohort. These estimates are qualitatively very similar to the main findings in column (1). They indicate that there is a gender gap in on-time Ph.D. completions among students in cohorts with a low ratio of women to men and that this gap is significantly diminished in cohorts with more female students. In column (3) of Table 4, we implement a count measure of gender composition and replace the $\%PeersFemale_{ipt}$ variable with the number of female peers in the cohort. Interestingly, these results show no evidence of a linear effect of the number of women peers in a program on the probability of Ph.D. completion for either men or women. However this finding is not inconsistent with the main results and merely indicates that the effect of an additional female peer interacts with the cohort size (e.g. 1 additional female peer has a large effect in a small cohort and little-to-no effect in a very large cohort). This interaction is better captured by the use of the percent female measure in the main specification.[23]

This main finding indicates that a 1 sd increase in the proportion female in a cohort increases the probability of on-time Ph.D. completion for women by 4.4pp, relative to men. It is difficult to draw a direct comparison to other estimates in the literature, as our outcome variable, Ph.D. completion, has rarely been available (most datasets on doctoral students are limited to degree-completers). However, we can compare to previous studies that analyze the effect of peer gender on persistence at earlier stages of the STEM pipeline. For example, Mouganie and Wang (2019) find that a 1 sd increase in the share of high-performing high school peers increases the probability of a woman choosing a science track by 2.1pp, relative to men. In another study conducted using data from West Point, Huntington-Klein and Rose (2018) find that adding one additional female peer to a company increases the probability that a woman will advance to the next year by 2.5pp.

We next explore the timing of the gender composition effect over the course of the first 6 years of Ph.D. enrollment. Figure 3 shows the rates of enrollment, dropout, and graduation for the main estimation sample by year of enrollment. This figure reveals that dropout occurs primarily in the first 3 years of doctoral programs and that roughly 50% of students graduate by the end of the 6th year. We model the effect of cohort gender composition on year-to-year persistence rates in doctoral enrollment as,

$$\begin{aligned}\mathbb{P}\left(Y^r_{ipt}=1\right) = &\beta^r_1 Female_i + \beta^r_2 \%PeersFemale_{ipt} + \beta^r_3 Female_i * \%PeersFemale_{ipt} \\ &+\gamma^{r\prime} X_{ipt} + \delta^r Z_{pt} + D^r_t + D^r_p + \epsilon^r_{ipt},\end{aligned} \tag{S2}$$

where $Y^r_{ipt}=1$ if individual i is still enrolled (or has graduated) in the Fall term of program-year $r \in [2,6]$. All other variables are unchanged from model S1.

Table 5 shows the marginal effects from estimating model S2 using a Probit MLE. In this table (and all further tables) we show results using our preferred specification where cohort gender composition is measured by the fraction of student i's peers in cohort pt who are female. Columns (1)-(5) show the effect of cohort gender composition on the probability of *not* dropping out before program-years 2 through 6, respectively. For example, in column

23. We show this more formally by estimating the following triple-interaction model:

$$\begin{aligned}Y_{ipt} = &\beta_1 Female_i + \beta_2 \#PeersFemale_{ipt} + \beta_3 CohortSize_{pt} + \beta_4 Female_i * \#PeersFemale_{ipt} \\ &+ \beta_5 \#PeersFemale_{ipt} * CohortSize_{pt} + \beta_6 Female_i * CohortSize_{pt} \\ &+ \beta_7 Female_i * \#PeersFemale_{ipt} * CohortSize_{pt} + \gamma' X_{ipt} + D_p + D_t + \epsilon_{ipt}.\end{aligned}$$

We find that the differential effect of one additional female peer for a woman in a small cohort (β_4), is a 2.9pp increase in the probability of degree completion (p-value of 0.002). However, the triple-interaction, β_7, is negative and significant indicating that the effect is diminishing as cohort size increases.

(1) the dependent variable is equal to one if student i, who enrolled in program p in year t, is either still enrolled or has graduated with a Ph.D. at the start of the Fall term of the following year.

The results in column (1) of Table 5 indicate that women in cohorts with no female peers are 7.2pp less likely to make it to the second year of a doctoral program than their male peers. That is equivalent to saying that these women are 7.2pp more likely to dropout in the first year of their Ph.D. program. A 1 sd increase in the share of female students decreases the dropout rate for women relative to men by 2.5pp in the first year of Ph.D. enrollment.[24] These results indicate that the majority of the gender composition effect on Ph.D. completion is accounted for by persistence through the first year of the program. This is unsurprising given that the majority of drop out occurs in the first program-year and that the cohort peer group should be most influential in the first year when these students are primarily involved in coursework, rather than lab research.[25]

As discussed above (and shown in Figure 1), there are some fields within the broad category of STEM that have a much lower average level of female representation than other fields. There is some evidence at the undergraduate level that these very male-dominated programs drive the gender gap in STEM major attrition (Astorne-Figari and Speer 2017). We next explore whether our main findings are primarily driven by these "typically male" programs

We first utilize a data-driven approach to identify which programs are the most male-dominated and calculate for each program the average fraction female across all years of the data (2005-2015) as well as the median value of this average across all 31 programs in the estimation sample (36.7% female). We then categorize programs with an average below this sample median as "typically male" and programs with an average above the sample median as "typically female."

The results of estimating model S1 separately for these two subsamples are shown in columns (1) and (2) of Table 6. These results indicate that the effect of cohort gender composition on Ph.D. completion is evident in both sets of programs. Column (1) shows that, in the typically male programs, the gender gap in Ph.D. completion is quite large. Women are 18.0pp less likely than men to graduate on-time in cohorts with no female peers and a 1 sd increase in the fraction of female peers differentially increases the probability of on-time graduation by 12.5pp for women relative to men. The results in column (2) of Table

24. Note that a small number of observations are dropped from the regression in column (1) of Table 5 due to one of the program fixed effects perfectly predicting the outcome variable (persist to program-year 2).

25. Note that these findings hold even within the subset of programs where students are admitted directly to an advisor/lab and are robust to excluding those few students in the sample who do not enroll in a full first-year course load.

6 for the typically female programs are similar but smaller in magnitude.

In columns (3) and (4) of Table 6, we implement a more generalizable approach to identifying male-dominated programs and categorize programs based on broad fields of study that are commonly known to have low female representation. Using the CIP field associated with each Ph.D. program, we redefine typically male programs to be those in the fields of: Engineering, Mathematics & Statistics, and Physical Sciences. Programs in all other fields (Agricultural Science, Biological and Biomedical Sciences, Chemistry, Economics, Health, and Psychology) are categorized as typically female. These results similarly indicate that the effect of cohort gender composition on Ph.D. completion is present in both groups of STEM fields.[26] While the estimates in column (3) for typically male programs are not statistically significant, this is likely due to the smaller sample size and point estimates across the two groups are similar in magnitude. Taken together, the results in Table 6 indicate that the effects of peer gender on female Ph.D. success rates are evident across the spectrum of STEM fields and programs.

5.1 Potential Mechanisms

There are a number of potential explanations for our finding that women persist longer and are more likely to complete programs when they have more female peers, some of which we are able to explore empirically. An increase in the share of peers who are female might benefit female students through improvements to performance in first year classes. A higher share of female peers might also improve women's chances of obtaining financial support on faculty research grants. Finally, the share of women in a cohort might have an intangible effect on the climate surrounding the students in that cohort, which increases female persistence in a less measurable manner. We test for the first two mechanisms by analyzing the effect of cohort gender composition on first year grades and on the probability of research support and conclude that, while the share of women in a cohort does have a small effect on female students' grades, the majority of the effect on Ph.D. persistence and completions must be attributed to unobserved mechanisms.

We first investigate whether cohort gender composition has an effect on first year performance. This might be an important mechanism for the gender peer effect if women are better able to learn when surrounded by other women or when studying with other women. There is also experimental evidence showing that women are less competitive, especially when competing against men (Gneezy, Niederle, and Rustichini 2003), so that

26. The results in columns (3) and (4) of Table 6 are largely unchanged if Economics is included in the group of typically male fields.

women in cohorts with more women may exert more effort studying and on assignments and exams. Both of these hypotheses suggest that women should have higher grades in cohorts with more women. Symmetrically, in cohorts with very few female peers, women would perform worse in first year courses, leading to a higher probability of dropping out. This mechanism might be heightened by a higher responsiveness to bad grades among women. Rask and Tiefenthaler (2008), Ost (2010), and Kugler, Tinsley, and Ukhaneva (2017) show that undergraduate women are more discouraged by low grades than men when making the choice of undergraduate major. Relatedly, Stinebrickner and Stinebrickner (2012) find that female undergraduates are more likely to update their beliefs about own ability in response to bad grades and to subsequently drop out of college. This issue has not been previously addressed at the doctoral level. If these findings carry over to the graduate level, then women may be more discouraged (and less likely to persist) due to lower first year grades in cohorts with very few female peers.

We test for these learning and competition mechanisms by looking for an effect of cohort gender composition on grades and by looking for a differential response to first year grades across genders. For this analysis, we maximize our potential sample by including additional cohorts of students who start their Ph.D. programs in 2010-2015.[27] The raw distribution of GPA at the end of the first term of enrollment for this expanded sample is shown in Figure 4 for men and women separately in both highly-male (left panel) and highly-female (right panel) cohorts.[28] Based on these unadjusted distributions it appears that there may be some small closing of a gender grade gap at the top of the distribution in highly-female programs but the visual differences are subtle. We estimate this more formally using the model S1, where we redefine Y_{ipt} to be a measure of individual i's first year grades. We measure this alternately as first term GPA or first year GPA. We estimate this model with an Ordinary Least Squares (OLS) estimator.

Column (1) of Table 7 shows the results of estimating model S1 with first term GPA as the dependent variable. These estimates show that women have first term GPAs that are 0.12 grade points lower than their male peers (on a 4-point scale) when they are the sole woman in a cohort. At the sample mean of 3.54, this is equivalent to a 3% gender gap in first term GPA. A 1 sd increase in the share of female students closes this gap by 0.04 grade points. Thus, these estimated effect are statistically significant but somewhat small in magnitude. Column (2) shows a similar effect of gender composition on GPA at the end of

27. Including these additional cohorts for whom we observe less than 6 years of data should not influence our analysis of first year grades. This is particularly relevant because Table 5 shows that the effect of cohort gender composition functions primarily through dropout decisions in the first year of enrollment.

28. In this figure, highly female cohorts are defined to be those cohorts with a fraction of women that is above average for that particular Ph.D. program.

the first year, but column (3) indicates that the effect on first year grades is entirely captured by the first term GPA.[29]

In Table 8, we estimate the models in S1 and S2 while allowing for a differential relationship between GPA and Ph.D. completion and persistence by gender (by interacting GPA with the $Female_i$ indicator). In columns (1)-(2) the dependent variable is Ph.D. completion in 6 years (model S1) and in columns (3)-(4) the dependent variable is an indicator for remaining enrolled into the Fall of the second year of the Ph.D. program (as in column (1) of Table 5). Although these results cannot be interpreted causally, they imply that, while first year grades appear to be largely predictive of both Ph.D. completion and persistence, female students' outcomes are not more strongly related to grades than are men's outcomes. If anything, the direction of the interaction term coefficients would indicate that female students are less responsive to first year grades than male students.

The estimates in Tables 7 and 8 indicate that peer gender composition has a small effect on first term GPA such that women have worse grades than men in highly-male cohorts. However, this effect can explain only a small portion of the overall impact of cohort gender composition on Ph.D. persistence and completion. For example, a 1 sd increase in the share of female peers closes the GPA gender gap by 0.04 grade points in the first term of enrollment. The coefficients in column (1) of Table 8 show that a 1 point increase in first term GPA is associated with an increase in the probability of on-time graduation by 30.1pp for men and 21.5pp for women. Note that these two coefficients are likely biased upwards as unobserved ability is almost surely positively correlated with both first term grades and on-time graduation. We can think of these as providing an upper bound on the causal effect of GPA on Ph.D. completion. Thus, the grade effect of a 1 sd increase in the share of female students is a differential increase in the female probability of on-time graduation of 1.16pp. This upper bound could then account for up to 1/4 of the total differential effect of a 4.35pp increase in Ph.D. completion shown in Table 4. The remaining 3/4 of our main finding remains to be explained.

Another empirically testable mechanism by which cohort gender composition might influence Ph.D. success is through a differential probability of obtaining research support. Previous work has shown that financial support is highly correlated with Ph.D. completion (Abedi and Benkin 1987; Ehrenberg and Mavros 1995).[30] Using the linked sample of UMETRICS data on students supported through research projects, we first verify that this

29. Note that the sample size in columns (2) and (3) of Table 7 is slightly smaller because we do not observe the outcome variable for those students who drop out and do not complete the first program-year.

30. Chang et al. (2017) provide detailed statistics on sources of financial support for STEM graduate students using the UMETRICS data matched to the Survey of Earned Doctorates.

correlation holds in our setting. We model this relationship by,

$$\mathbb{P}\left(Y_{ipt}^{r}=1\right)=\beta^{r} Funding_{i}^{r-1}+\gamma^{r\prime} X_{ipt}+\delta^{r} Z_{pt}+D_{t}^{r}+D_{p}^{r}+\epsilon_{ipt}^{r}, \tag{S3}$$

where $Y_{ipt}^{r}=1$ if individual i remains enrolled (or has graduated) in the Fall of year r of the doctoral program ($r \in [2,5]$) and $Funding_{i}^{r-1}=1$ if individual i receives federally-funded research support during year $r-1$ of the program. For example, when $r=2$, β^{2} measures the correlation between receiving funding in the first year of a STEM doctoral program and persisting to the 2nd year of the program. In this model, the vector X_{ipt} includes gender along with age, age^2, race/ethnicity indicators, and a foreign student indicator variable.

In column (1) of Table 9, we estimate the relationship between being employed on a federally-funded research grant for at least 28 days during the first year of enrollment and the probability of remaining enrolled (or having graduated) in the Fall term of the second program-year. Column (2) shows the relationship between employment in the second year of enrollment, conditional on enrollment in the second year, and the probability of persisting to the third year of the doctoral program. As expected, we find that obtaining research funding is highly correlated with persistence at each year of the doctoral program. Table 10 shows the relationship between obtaining research support and the probability of on-time graduation for STEM doctoral students. These results show a similar, positive relationship.

Given that research support appears to be strongly related to Ph.D. success, we next investigate whether cohort gender composition has an effect on the probability of obtaining research employment. If female students are more likely to obtain research funding in cohorts with more female (and fewer male) peers, then funding could be an important mechanism in explaining our main findings in Table 4. We model the relationship between cohort gender composition and research support using model S2 where we change the dependent variable to be equal to one only if individual i is employed on a federal research grant for at least 28 days during year r of enrollment in the doctoral program. The marginal effects results of estimating this specification are shown in Table 11. These estimates fail to provide evidence that peer gender composition has any effect on research funding in any year for either gender. The marginal effects are statistically insignificant and are inconsistent in sign and magnitude across years. These findings, along with our main results, do not support a research funding mechanism.

In conclusion, we do not find any evidence that peer gender composition has an impact on students' financial support through research funding. However, we do find a small effect of peer gender on first year grades. We estimate an upper bound showing that changes in learning and\or effort (as they are reflected in grades) can account for up to one quarter of the total effect of peer gender composition on Ph.D. completions. In the absence of additional observable mechanisms, we are left to speculate whether our measure of peer gender composition might also capture some unobservable changes in the climate of each cohort. This would imply that when cohort gender composition is particularly high, the intangible climate towards women (in that specific cohort) improves, thereby increasing female students' persistence and on-time graduation.

5.2 Alternate Explanations

Another set of potential explanations for our main findings focus on mentoring and the gender mix of faculty. We have explored mentoring by relating Ph.D. completion to the gender composition of the cohort that entered one year earlier, under the assumption that the older cohort interacts with the younger cohort. If our main results are driven by some trend within programs in both student outcomes and gender composition (such as the hiring of a more female-friendly graduate advisor or department chair), then we would expect to find a spurious "effect" when looking at adjacent cohort gender peer effects.

We investigate this possibility by estimating the following model, which includes the previous cohort's gender composition as an independent variable:

$$\begin{aligned}\mathbb{P}\left(Y_{ipt}=1\right)=&\beta_1 Female_i+\beta_2\%PeersFemale_{ipt}+\beta_3 Female_i*\%PeersFemale_{ipt}\\&+\beta_4\%CohortFemale_{pt-1}+\beta_5 Female_i*\%CohortFemale_{pt-1}\\&+\gamma' X_{ipt}+\delta Z_{pt}+\phi Z_{pt-1}+D_p+D_t+\epsilon_{ipt},\end{aligned}\tag{S4}$$

where $Y_{ipt}=1$ if student i, who first enrolled in program p in year t, completes a Ph.D. within 6 years. The new variable, $\%CohortFemale_{pt-1}$, measures the fraction of women in cohort $t-1$ still enrolled in program p at the beginning of the following year (when cohort t first enrolls). The variable Z_{pt} measures the number of students in cohort pt and Z_{pt-1} measures the number of student in the previous cohort, $pt-1$. All other variables are as in (S1). Note that $\%CohortFemale_{pt-1}$ is not measured for the very first cohort of our main sample, so the sample for this model includes only cohorts entering in 2006-2009.

In column (1) of Table 12, we replicate our main results (column (1) of Table 4) using this abbreviated sample. In column (2), we include the previous cohort's gender composition ($\%CohortFemale_{pt-1}$) as a control variable. In column (3), we allow for this older cohort's gender mix to have a differential impact on male and female students in the next year's cohort by interacting with the $Female_i$ indicator variable. These estimates show no effect of the older cohort on Ph.D. completion of the younger cohort. This may be because interactions with older cohorts are more prevalent in later years of the doctoral program when students are less susceptible to dropping out. These results indicate that the relevant peer group in this setting are students in one's own cohort and allays the concern that our main results are driven by a common trend in student outcomes and peer gender composition.

Another factor that may play an important role with regards to retention of female students is faculty gender composition. Boustan and Langan (2019) show that in economics Ph.D. programs, there is a strong correlation between the faculty gender composition and the gender composition of students graduating from those programs. Carrell, Page, and West (2010) show that at the undergraduate level, professor gender has a large impact on female students' performance in math and science courses and their propensity to major in a STEM field. We next investigate whether our main findings might be explained by a strong correlation between cohort gender composition and changes in the faculty gender composition. However, given that our estimates are identified from year-to-year fluctuations in the composition of cohorts and given the relatively slow turnover of faculty, faculty composition seems to be an unlikely explanation.

We obtain faculty gender data from two separate sources. First, we use the OLDA transcript data to identify the roster of courses for each program in each year that enroll at least 50% of the incoming doctoral cohort. We can then identify the instructors (and their genders) for these courses and calculate the gender composition of those tenure-track faculty who are teaching first-year courses for each Ph.D. program in each year of the main estimation sample. Because we may also care about the gender composition of entire departments (rather than just those faculty who are teaching the first-year courses), we also collect annual faculty rosters by department for Ohio State University thanks to Kim (2019) and for the other 5 Ohio universities from Academic Analytics.[31] We are able to link the program CIP codes from the OLDA transcript data to these department rosters for 26 of the 31 doctoral programs in our main estimation sample and then calculate the year-to-year gender composition of faculty in each program.

Summary statistics for these 26 programs with faculty data are shown in Table 13. On average, the programs in our sample have 8 faculty teaching at least one first-year course in a given year and 34 faculty in the department as a whole. The average gender composition of first-year teaching faculty is 13% female and the department faculty are 17% female overall. The overall correlation between the cohort gender composition of students and the faculty gender composition is positive and fairly large (0.29 for teaching faculty and 0.56 for all department faculty). However, this correlation is driven almost entirely by across-program differences in both variables. Once we account for program averages and demean

31. https://academicanalytics.com

the raw data, the year-to-year within-program variation in cohort gender composition and faculty gender composition are essentially uncorrelated (0.02 for teaching faculty and 0.05 for department faculty).

In Table 14, we investigate whether changes in faculty gender composition are driving our main results by estimating the model in (S4) substituting faculty gender composition ($\%FacultyFemale_{pt}$) in place of the variable $\%CohortFemale_{pt-1}$.[32] In column (1) we replicate our main results (column (1) of Table 4) using this restricted sample. In columns (2) and (4), we include the fraction of faculty who are female ($\%FacultyFemale_{pt}$) as a control variable. In columns (3) and (5), we allow for faculty gender to have a differential impact on male and female students by interacting with the $Female_i$ indicator variable. These estimates show that, if anything, an increase in the fraction of female faculty has a negative impact of the probability of male students' on-time graduation and there is no differential effect for female students. These results indicate that our main findings cannot be accounted for by correlated changes in faculty gender composition, but are truly driven by changes in peer gender composition.

Finally, we seek to verify that changes in the gender of the graduate studies directors for these programs are not driving our main results. Data on graduate studies directors is not available from any administrative source that we are able to identify, however, we were able to collect this information for individual programs at the Ohio State University by contacting each department represented in our sample. In this way, we managed to gather information on the gender of the graduate studies directors for the years 2005-2015 for 16 of the 31 doctoral programs in our main estimation sample.

The gender composition of the student cohorts in this smaller sample are very similar to the main estimation sample. On average, programs in this sample have a female graduate director 25% of the time. The overall correlation between the cohort gender composition of students and the gender of graduate director is 0.25. However, as with the faculty gender composition, this correlation is driven almost entirely by across-program differences. Once we demean the raw data using program-level means, the year-to-year within-program variation in cohort gender composition and the gender of the program director are essentially uncorrelated (corr = 0.02). The gender of the graduate director is similarly uncorrelated with our primary outcome variable. The overall correlation between gender of the graduate director and the probability of completing a Ph.D. within 6 years of initial enrollment is less than -0.01.

Unfortunately, we cannot study the effects of graduate studies director's gender on Ph.D. completion (as we did with faculty gender composition in Table 14). In order to

32. The regressions in Table 14 do not include the variable Z_{pt-1} shown in (S4).

analyze effects on Ph.D. completion, we must restrict the sample to the 2005-2009 cohorts (for whom we can observe 6 years of data). However, during this abbreviated time period, we observe that only 1 program in this sample experienced a change in the gender of the graduate studies director. Thus, the variable capturing gender of the graduate director is essentially collinear with the program fixed effects in our model. This lack of variation, along with the negligible within-program correlation between student gender composition and the gender of the graduate director strongly suggest that our main findings cannot be accounted for by correlated changes in the graduate studies director.

6 Robustness Checks

This section demonstrates the robustness of our main findings on the effects of cohort gender composition on Ph.D. completion to a number of alternate specifications and alternate samples. These robustness results are shown in Tables 15-18. We first show that our main results are not sensitive to the use of a Probit MLE. In column (2) of Table 15 we estimate the main specification in model S1 as a linear probability model using an OLS estimator.[33] Next, we address the concern that 6 years may not be a sufficient time span to accurately capture an effect on Ph.D. completions. In column (3), we re-estimate model S1 by replacing the dependent variable with an indicator for graduating within 7 years of initial enrollment. Despite the diminished sample size, these results are larger than our main results and statistically significant, indicating that women in highly-male cohorts are not merely delaying graduation to program-year 7. These results are consistent with those in Table 5 showing that the majority of the peer gender effect is driven by dropping out in the first year of enrollment. In columns (4)-(7) of Table 15, we verify that the main results are robust to the inclusion of university-by-year fixed effects, linear time trends, university-specific linear time trends, and quadratic trends. All of these estimates are very similar in both magnitude and precision to the main results.

In Table 16, we implement several alternate samples and program definitions. In columns (2)-(3) of Table 16, we implement alternate definitions of doctoral programs.[34] Recall that in the main estimation sample we define a doctoral program to include all students attending the same institution with the same enrollment CIP code. In column (2), we aggregate this definition up to include all students attending the same institution with the same enrollment CIP field. Note that CIP fields are a much broader classification than CIP codes. Under this classification, the effects of peer gender composition are both smaller and

33. Column (1) of Table 15 replicates the main findings in column (1) of Table 4 for reference.
34. Column (1) of Table 16 replicates the main findings in column (1) of Table 4 for reference.

more noisy, which is consistent with an attenuation bias associated with measurement error (likely incurred by aggregating, for example, 5 different Biology CIP codes into 1 very large "program"). However, in column (3) of Table 16, we instead disaggregate the CIP codes into university-specific program identifiers and use these codes to define each program.[35] These results are very similar in both size and precision to the main results shown in column (1).

In columns (4) and (5) of Table 16, we test whether our results hold in non-STEM doctoral programs. In column (4) of Table 16 we include both STEM and non-STEM programs (conditional on an average cohort size of more than 9 students) in the sample. These results are largely similar to the main findings in column (1). In column (5), we limit the sample to non-STEM programs only. Note that this alternate sample is quite small because most non-STEM doctoral programs have average cohort sizes that are less than 10 students. The magnitudes of these results are consistent with the main findings, but the estimates are very noisy (which is unsurprising, given the sample size). This suggests that the effect of peer gender on Ph.D. success may not be limited to STEM programs.[36]

As noted above, in the main estimation sample, we limit the data to include only programs with an average cohort size greater than 9. This was due to ex-ante concern that the structure of very small programs might be quite different such that incoming cohorts are not a relevant peer group. However, in Table 17 we replicate the main specification in model S1 with alternate estimation samples excluding/including programs with higher/lower average cohort sizes. Note that column (3) is a replication of the main findings in column (1) of Table 4. These results show that the main findings are robust to alternate definitions of small and large programs.

Recall that the main estimation sample excludes students who first enroll in a non-doctoral graduate program and then transfer into a doctoral program at the same institution. For those "transfer" students it is difficult to determine which cohort they belong to. We also exclude all programs where these transfer students encompass more than 20% of total enrollment. In the final robustness check shown in Table 18, we allow for programs with a higher/lower percentage of transfer students than in the main estimation sample and show that the main results are largely unchanged.

35. We do not use these program identifiers in the main sample because they are not consistently defined across all years of the sample.

36. For additional analyses including the expanded sample of non-STEM programs with smaller average cohort sizes, see Appendix Table A1.

7 Conclusion

The underrepresentation of women in STEM is a topic of great interest in economics and public policy today. However, the factors affecting persistence in STEM fields are not well-understood and our causal understanding is especially limited at the graduate education level. We investigate one factor in the training process of STEM doctoral degrees, peer gender composition, and find that it has a significant impact on the gap in Ph.D. completion rates between men and women. Using year-to-year variation within doctoral programs in the fraction of each cohort that is female, we find that women who have no female peers in a cohort are 11.7pp less likely to graduate within 6 years of initial enrollment than men. However, a 1 sd increase in the share of female peers in a cohort increases the probability of on-time graduation for women as compared to their male counterparts by 4.4pp.

We find that this effect is largely due to dropout behavior in the first year of enrollment. We investigate several potential mechanisms and find that peer gender composition has a small effect on first term GPA (which explains only a quarter of the overall gender peer effect on retention and completion) and no effect on the probability of obtaining research funding. The small/null findings for these two channels suggest that women learning or competing more successfully in cohorts with more female peers is at best a secondary effect in explaining our results. On the other hand, our findings are consistent with a climate mechanism, through which more female peers create a female-friendly environment that encourages women to persist in doctoral programs. Taken together, these findings indicate that peer gender composition may be a useful proxy for climate and that year-to-year variation in this measure can provide a useful identification strategy for investigating gender gaps in outcomes.

References

Abedi, Jamal, and Ellen Benkin. 1987. "The Effects of Students' Academic, Financial, and Demographic Variables on Time to the Doctorate." *Research in Higher Education* 27 (1): 3–14. http://www.jstor.org/stable/40195800.

Anelli, Massimo, and Giovanni Peri. 2016. "The Effects of High School Peers' Gender on College Major, College Performance and Income."

Astorne-Figari, Carmen, and Jamin D Speer. 2017. "Are Changes of Major Major Changes? The Roles of Grades, Gender, and Preferences in College Major Switching." http://www.sole-jole.org/17322.pdf.

Bayer, Amanda, and Cecilia Elena Rouse. 2016. "Diversity in the Economics Profession: A New Attack on an Old Problem." *Journal of Economic Perspectives* 30, no. 4 (November): 221–242. ISSN: 0895-3309. doi:10.1257/jep.30.4.221. http://pubs.aeaweb.org/doi/10.1257/jep.30.4.221.

Bell-Ellison, Bethany A., and Robert F. Dedrick. 2008. "What do Doctoral Students Value in their Ideal Mentor?" *Research in Higher Education* 49, no. 6 (September): 555–567. ISSN: 0361-0365. doi:10.1007/s11162-008-9085-8. http://link.springer.com/10.1007/s11162-008-9085-8.

Boustan, Leah, and Andrew Langan. 2019. "Variation in Women's Success across PhD Programs in Economics." *Journal of Economic Perspectives* 33 (1): 23–42. doi:10.1257/jep.33.1.23. https://doi.org/10.1257/jep.33.1.23.

Brenøe, Anne Ardila, and Ulf Zölitz. 2018. "Exposure to More Female Peers Widens the Gender Gap in STEM Participation." http://www.econ.uzh.ch/static/wp/econwp285.pdf.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and science: how professor gender perpetuates the gender gap." *The Quarterly Journal of Economics* 125 (3): 1101–1144. ISSN: 00335533. doi:10.1162/qjec.2010.125.3.1101. arXiv: arXiv:1011.1669v3.

Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams. 2014. "Women in Academic Science: A Changing Landscape." *Psychological Science in the Public Interest* 15, no. 3 (December): 75–141. ISSN: 1529-1006. doi:10.1177/1529100614541236. http://www.ncbi.nlm.nih.gov/pubmed/26172066%20http://journals.sagepub.com/doi/10.1177/1529100614541236.

Chang, Wan-Ying, Wei Cheng, Julia Lane, and Bruce Weinberg. 2017. "Federal Funding of Doctoral Recipients: Results from new Linked Survey and Transaction Data." Cambridge, MA, January. doi:10.3386/w23019. http://www.nber.org/papers/w23019.pdf.

Clark, Richard A., Sherry L. Harden, and W. Brad Johnson. 2000. "Mentor Relationships in Clinical Psychology Doctoral Training: Results of a National Survey." *Teaching of Psychology* 27, no. 4 (November): 262–268. ISSN: 0098-6283. doi:10.1207/S15328023TOP2704_04. http://top.sagepub.com/lookup/doi/10.1207/S15328023TOP2704%7B%5C_%7D04.

Corbett, Christianne, and Catherine Hill. 2015. *Solving the Equation-The Variables for Women's Success in Engineering and Computing.* 159. AAUW. ISBN: 9781879922457. doi:10.1103/PhysRevA.75.063427. arXiv: 0612219 [physics]. http://www.aauw.org/resource/get-the-solving-the-equation-report/.

Dupas, Pascaline, Alicia Sasser Modestino, Muriel Niederle, Justin Wolfers, and The Seminar Dynamics Collective. 2021. *Gender and the Dynamics of Economics Seminars.* Technical report. Cambridge, MA: National Bureau of Economic Research, February. doi:10.3386/w28494. http://www.nber.org/papers/w28494.pdf.

Ehrenberg, Ronald G, and Panagiotis G Mavros. 1995. "Do Doctoral Students' Financial Support Patterns Affect Their Times-To-Degree and Completion Probabilities?" *The Journal of Human Resources* 30 (3): 581–609. http://www.jstor.org/stable/146036.

Fouad, Nadya A., and Romila Singh. 2011. *Stemming the Tide: Why Women Leave Engineering.* Technical report. University of Wisconsin-Milwaukee.

Gardner, Susan K., AE Austin, HD Calhoun, J Trevino, B Lisek, and M Hess. 2009. "Conceptualizing Success in Doctoral Education: Perspectives of Faculty in Seven Disciplines." *The Review of Higher Education* 32 (3): 383–406. ISSN: 1090-7009. doi:10.1353/rhe.0.0075. http://muse.jhu.edu/content/crossref/journals/review%7B%5C_%7Dof%7B%5C_%7Dhigher%7B%5C_%7Deducation/v032/32.3.gardner.html.

Gaule, Patrick, and Mario Piacentini. 2017. "An Advisor Like Me? Advisor Gender and Post-graduate Careers in Science." http://www.cerge-ei.cz/pdf/wp/Wp594.pdf.

Gneezy, U., M. Niederle, and A. Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *The Quarterly Journal of Economics* 118, no. 3 (August): 1049–1074. ISSN: 0033-5533. doi:10.1162/00335530360698496. https://academic.oup.com/qje/article-lookup/doi/10.1162/00335530360698496.

Golde, Chris M. 2005. "The Role of the Department and Discipline in Doctoral Student Attrition: Lessons from Four Departments." *The Journal of Higher Education* 76 (6): 669–700.

Golde, Chris M., and Timothy M. Dore. 2001. *At Cross Purposes: Survey of Doctoral Education Report.* Technical report. Philadelphia, PA: A report prepared for the Pew Charitable Trusts. http://www.phd-survey.org/report.htm.

Hall, Leigh, and Leslie Burns. 2009. "Identity Development and Mentoring in Doctoral Education." *Harvard Educational Review* 79, no. 1 (April): 49–70. ISSN: 0017-8055. doi:10.17763/haer.79.1.wr25486891279345. http://hepgjournals.org/doi/10.17763/haer.79.1.wr25486891279345.

Hengel, Erin. 2017. "Publishing while Female. Are women held to higher standards? Evidence from peer review." December. https://econpapers.repec.org/paper/camcamdae/1753.htm.

Hill, Andrew J. 2017. "The positive influence of female college students on their male peers." *Labour Economics* 44 (January): 151–160. ISSN: 0927-5371. doi:10.1016/J.LABECO.2017.01.005. https://www.sciencedirect.com/science/article/pii/S0927537117300581.

Hilmer, Christiana, and Michael Hilmer. 2007. "Women helping women, men helping women? Same-gender mentoring, initial job placements, and early career publishing success for economics PhDs." *American Economic Review* 97 (2): 422–426. ISSN: 00028282. doi:10.1257/aer.97.2.422.

Hoxby, Caroline M. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." doi:10.3386/w7867. http://www.nber.org/papers/w7867.

Hunt, Jennifer. 2016. "Why Do Women Leave ScIence And EngIneering?" *ILR Review* 69 (1): 199–226. doi:10.1177/0019793915594597. http://journals.sagepub.com/doi/pdf/10.1177/0019793915594597.

Huntington-Klein, Nick, and Elaina Rose. 2018. "Gender Peer Effects in a Predominantly Male Environment : Evidence from West Point." *American Economic Review: Papers & Proceedings* 108:1–4.

Kim, MinSub. 2019. "Opening the Door for Others?: Female Leadership and Gender Disparities."

Kugler, Adriana D, Catherine H Tinsley, and Olga Ukhaneva. 2017. "Choice of Majors: Are Women Really Different from Men?" http://www.nber.org/papers/w23735.

Lavy, Victor, and Analía Schlosser. 2011. "Mechanisms and impacts of gender peer effects at school." *American Economic Journal: Applied Economics* 3, no. 2 (April): 1–33. ISSN: 19457782. doi:10.1257/app.3.2.1.

Lordan, Grace, and Jörn-Steffen Pischke. 2016. "Does Rosie Like Riveting? Male and Female Occupational Choices." http://www.nber.org/papers/w22495.

Lundberg, Shelly, Jennifer Bennett Shinall, Jennifer Doleac, Kathryn J. Norlock, and Shereen G. Bingham. 2018. *Reports from the American Economic Association's Committee on the Status of Women in the Economics Profession.* Technical report. https://www.aeaweb.org/content/file?id=6769.

Main, J. 2014. "Gender Homophily, Ph.D. Completion, and Time to Degree in the Humanities and Humanistic Social Sciences." *The Review of Higher Education* 37 (3): 349–375. ISSN: 1090-7009. doi:10.1353/rhe.2014.0019. http://muse.jhu.edu/journals/rhe/summary/v037/37.3.main.html.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2017. "Gender Bias in Teaching Evaluations." www.iza.org.

Milkman, Katherine L., Modupe Akinola, and Dolly Chugh. 2015. "What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations." *Journal of Applied Psychology* 100, no. 6 (November): 1678–1712. ISSN: 1939-1854. doi:10.1037/apl0000022. http://www.ncbi.nlm.nih.gov/pubmed/25867167%20http://doi.apa.org/getdoi.cfm?doi=10.1037/apl0000022.

Moss-Racusin, Corinne A, John F Dovidio, Victoria L Brescoll, Mark J Graham, and Jo Handelsman. 2012. "Science faculty's subtle gender biases favor male students." *Proceedings of the National Academy of Sciences of the United States of America* 109, no. 41 (October): 16474–9. ISSN: 1091-6490. doi:10.1073/pnas.1211286109. http://www.ncbi.nlm.nih.gov/pubmed/22988126%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3478626.

Mouganie, Pierre, and Yaojing Wang. 2019. "High-Performing Peers and Female STEM Choices in School." http://ftp.iza.org/dp12455.pdf.

Nerad, Maresi. 2004. "The PhD in the US: Criticisms, Facts, and Remedies." *Higher Education Policy* 17, no. 2 (June): 183–199. ISSN: 0952-8733. doi:10.1057/palgrave.hep.8300050. http://link.springer.com/10.1057/palgrave.hep.8300050.

Nettles, Michael T. 1990. "Success in doctoral programs: Experiences of minority and white students." *American Journal of Education* 98, no. 4 (August): 494. ISSN: 0195-6744. doi:10.1086/443974. http://www.journals.uchicago.edu/doi/10.1086/443974.

Neumark, David, and Rosella Gardecki. 1998. "Women Helping Women ? Role Model and Mentoring Effects on Female Ph . D . Students in Economics." *The Journal of Human Resources* 33 (1): 220–246. http://www.jstor.org/stable/146320.

Ost, Ben. 2010. "The role of peers and grades in determining major persistence in the sciences." *Economics of Education Review* 29:923–934. doi:10.1016/j.econedurev.2010.06.011. http://ac.els-cdn.com/S0272775710000762/1-s2.0-S0272775710000762-main.pdf?%7B%5C_%7Dtid=2fe9215a-26ca-11e7-967a-00000aab0f26%7B%5C&%7Dacdnat=1492803728%7B%5C_%7D3720f277ca0608a330d214cf8e260ff3.

Pezzoni, Michele, Jacques Mairesse, Paula Stephan, Julia Lane, G Burnett, and K Burnett. 2016. "Gender and the Publication Output of Graduate Students: A Case Study." Edited by Wolfgang Glanzel. *PLOS ONE* 11, no. 1 (January): e0145146. ISSN: 1932-6203. doi:10.1371/journal.pone.0145146. http://dx.plos.org/10.1371/journal.pone.0145146.

Rask, Kevin, and Jill Tiefenthaler. 2008. "The role of grade sensitivity in explaining the gender imbalance in undergraduate economics." *Economics of Education Review* 27 (6): 676–687. ISSN: 02727757. doi:10.1016/j.econedurev.2007.09.010. http://www.sciencedirect.com/science/article/pii/S0272775707001185.

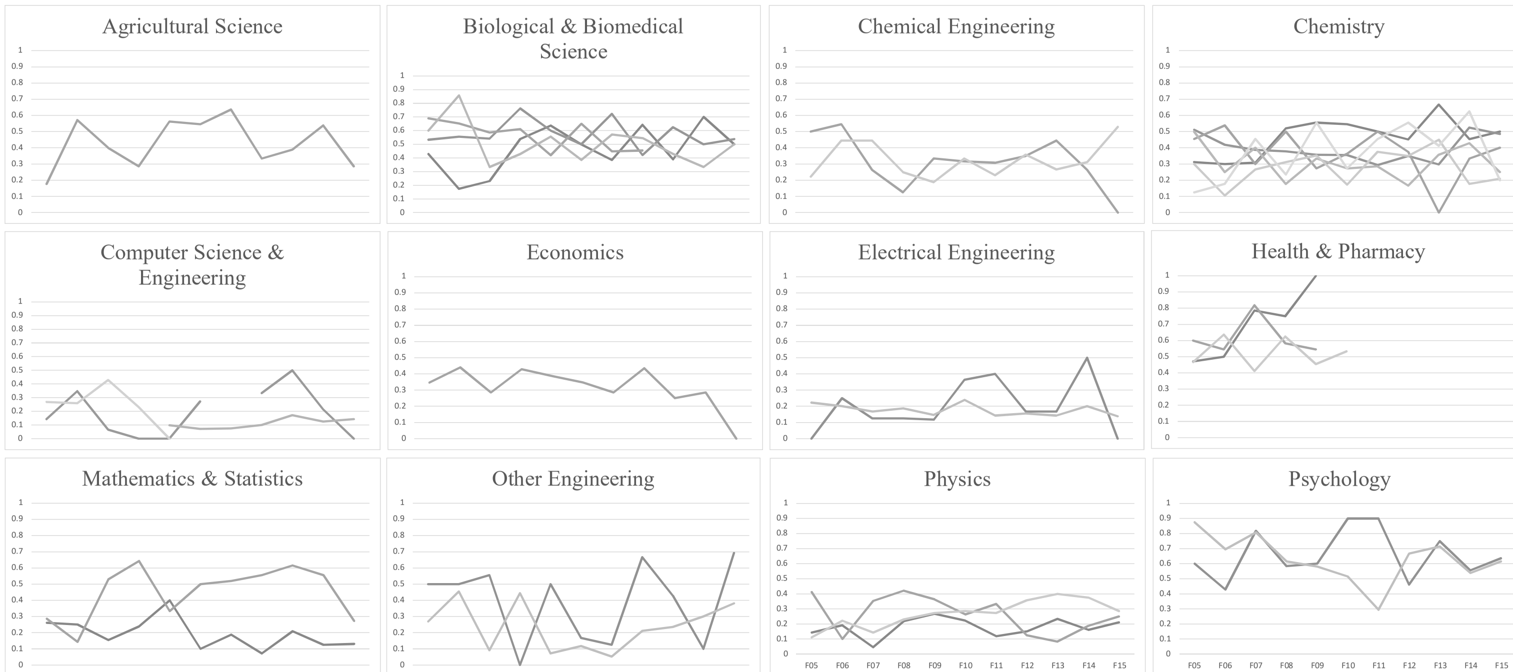
Sarsons, Heather. 2017. "Recognition for Group Work: Gender Differences in Academia." *American Economic Review: Papers and Proceedings* 107 (5): 141–145.

Stinebrickner, Todd, and Ralph Stinebrickner. 2012. "Learning about Academic Ability and the College Dropout Decision." *Journal of Labor Economics* 30, no. 4 (October): 707–748. ISSN: 0734-306X. doi:10.1086/666525. http://www.journals.uchicago.edu/doi/10.1086/666525.

Wu, Alice H. 2017. "Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum." December. doi:10.2139/ssrn.3051462. https://www.ssrn.com/abstract=3051462.

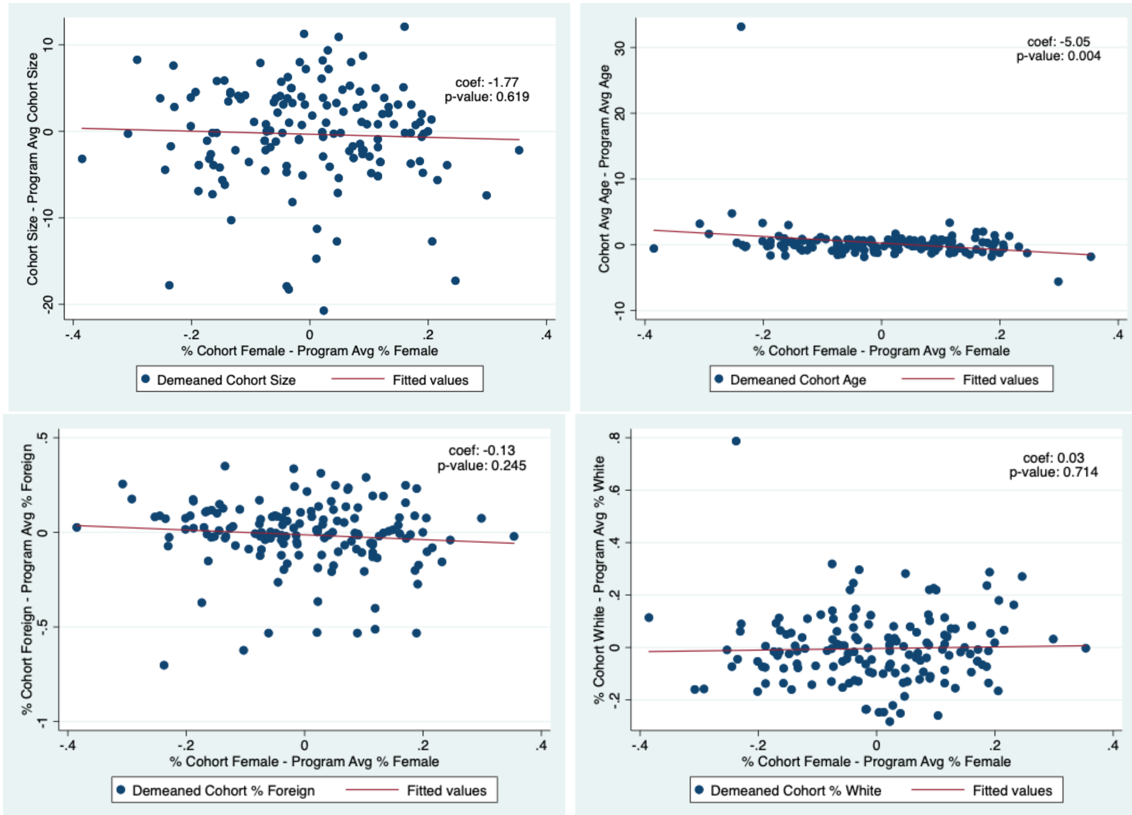
Zölitz, Ulf, and Jan Feld. 2017. "The Effect of Peer Gender on Major Choice." http://www.econ.uzh.ch/static/wp/econwp270.pdf.

Figure 1: Trends in Cohort Gender Composition By Field



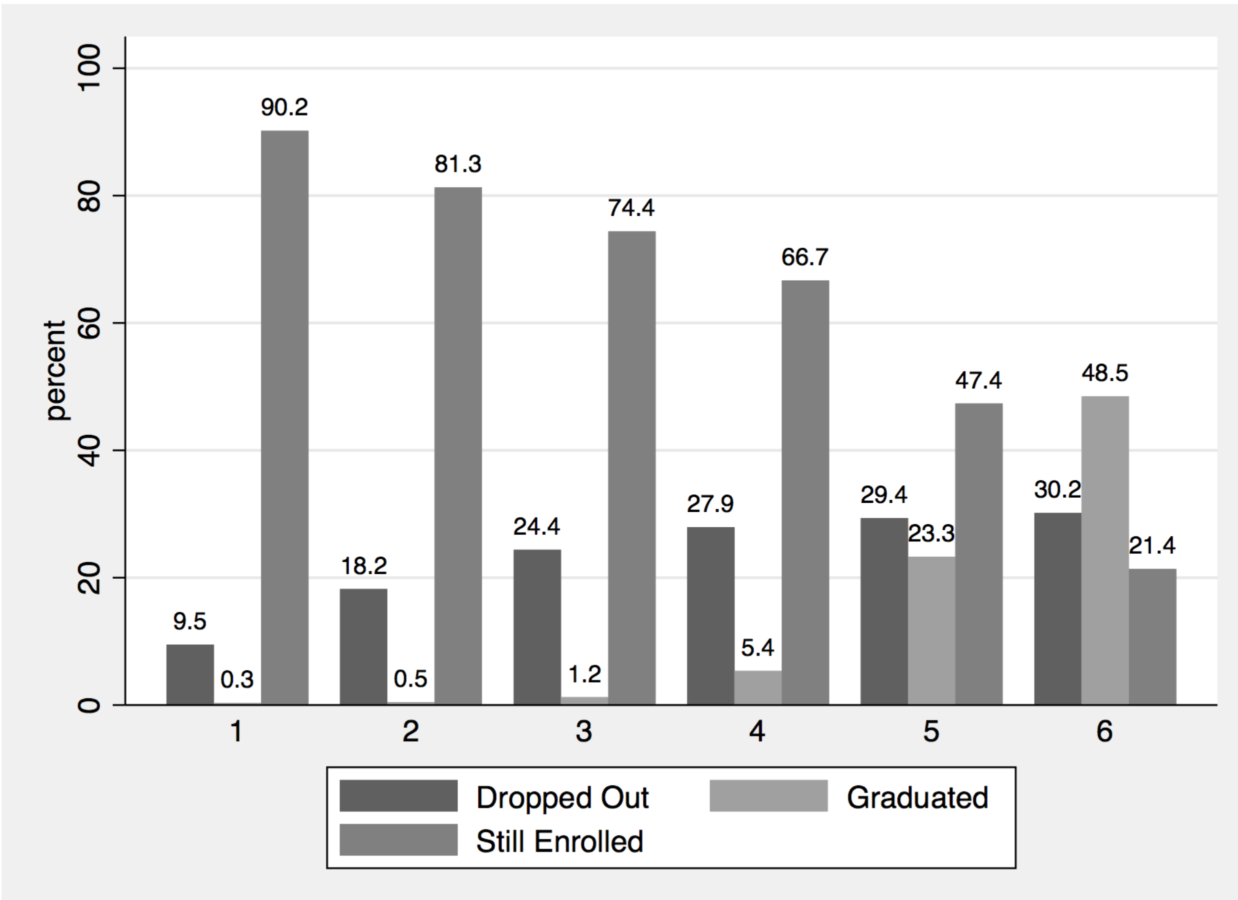
Each panel represents a STEM field and each line within a panel represents a unique doctoral program in the main estimation sample. The x-axis measures incoming cohorts over time and the y-axis measures the share of each incoming cohort that is female.

Figure 2: Correlation Between Cohort Gender Composition and Covariates (Demeaned)



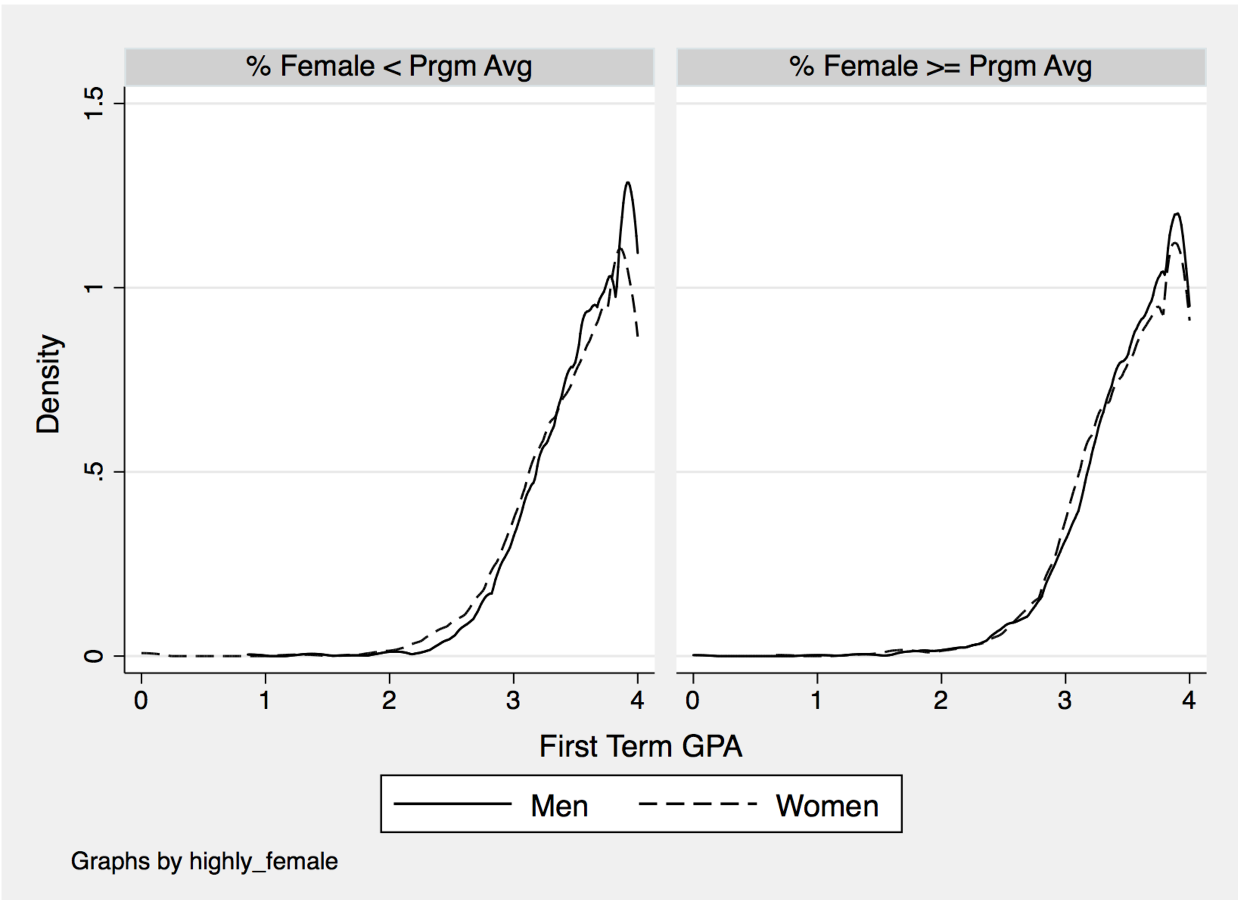
In each panel above, a point represents a cohort within a program. The x-axis measures the percent of each cohort that is female minus the average percent female in the program over all years of the data. The y-axis in each panel represent a different covariate, also demeaned at the program level. Proceeding clockwise from the top-left, these variables are: cohort size, age, an indicator for white race, and an indicator for foreign-born. Each panel includes the estimated slope coefficient and corresponding p-value from a simple linear regression of the demeaned y-variable on the demeaned cohort gender composition.

Figure 3: Dropout and Graduation Rates by Year of Enrollment



The x-axis measures the program-year. The y-axis indicates the fraction of students in the main estimation sample belonging to each category (dropped out, graduated, or still enrolled) measured at the end of each program-year.

Figure 4: Distribution of First Term Grades by Gender



The left panel shows the first term GPA distributions for all students in cohorts where the fraction female is below the average for that program ("highly-male" cohorts). The right panel includes all students in cohorts where the gender composition is above the average for that program ("highly-female" cohorts).

Table 1: Summary Statistics by CIP Code

CIP Field	CIP Code	CIP Code Subject Title	Avg Cohort Size	Avg % Female	# of Institutions
Computer Engineering	140901	Computer Engineering, General.	24.0	15%	2
Electrical, Electronics, and Communications Engineering	141001	Electrical and Electronics Engineering	19.9	19%	2
Mathematics and Statistics	270101	Mathematics, General.	17.7	19%	1
Physics	400801	Physics, General.	18.5	24%	3
Computer Science	110101	Computer and Information Sciences, General.	18.8	24%	1
Materials Engineering	141801	Materials Engineering.	16.6	24%	1
Chemistry	400507	Polymer Chemistry.	20.1	28%	1
Chemical Engineering	140701	Chemical Engineering.	14.9	31%	1
Economics (Social Science)	450601	Economics, General.	23.0	32%	1
Chemical Engineering	143201	Polymer/Plastics Engineering.	12.5	33%	1
Chemistry	400599	Chemistry, Other.	12.5	37%	1
Chemistry	400501	Chemistry, General.	22.2	38%	4
Other Engineering	140501	Bioengineering and Biomedical Engineering.	9.2	38%	1
Agricultural Science	10103	Agricultural Economics.	13.2	43%	1
Mathematics and Statistics	270501	Statistics, General.	14.3	45%	1
Biological and Biomedical Sciences	260202	Biochemistry.	14.7	47%	1
Biological and Biomedical Sciences	260502	Microbiology, General.	9.2	50%	1
Pharmacy	512001	Pharmacy.	12.8	52%	1
Biological and Biomedical Sciences	269999	Biological and Biomedical Sciences, Other.	25.1	56%	1
Biological and Biomedical Sciences	260499	Cell/Cellular Biology and Anatomical Sciences, Other.	19.0	57%	1
Other Health	511401	Medical Scientist.	11.0	62%	1
Psychology	420101	Psychology, General.	18.2	64%	2
General Health/Public Health	512202	Environmental Health.	9.4	70%	1

Table 2: Cohort Characteristics

	Mean	Std Dev	Min	Max
Panel A: Estimation Sample (2005-2009)				
STEM Field	1	0	1	1
Cohort Size	16.64	9.19	1	49
# Female in Cohort	6.29	4.68	0	23
% Female in Cohort	0.38	0.21	0	1
Ratio Female/Male	0.89	1.04	0	7
Obs		151 cohorts		
Panel B: Estimation Sample + Non-STEM + Small Programs				
STEM Field	0.69	0.46	0	1
Cohort Size	7.81	7.27	1	49
# Female in Cohort	3.52	3.42	0	23
% Female in Cohort	0.50	0.31	0	1
Ratio Female/Male	1.15	1.35	0	7
Obs		689 cohorts		
Panel C: Full Sample (All Years, 2005-2015)				
STEM Field	0.70	0.46	0	1
Cohort Size	7.60	7.68	1	80
# Female in Cohort	3.31	3.25	0	28
% Female in Cohort	0.49	0.31	0	1
Ratio Female/Male	1.10	1.27	0	11
Obs		1,529 cohorts		

Table 3: Summary Statistics

	Male		Female	
	Mean	Std Dev	Mean	Std Dev
PhD in 6 Yrs	0.48	0.50	0.49	0.50
Yrs to Graduate	5.52	1.16	5.53	1.07
Drop Out (by end of 6 yrs)	0.30	0.46	0.31	0.46
Still Enrolled (by end of 6 yrs)	0.22	0.42	0.20	0.40
# Yrs Enrolled	4.45	2.06	4.44	2.02
Age (Yr Enrolled - Birth Yr)	25.19	3.87	24.62	3.64
Foreign	0.51	0.50	0.49	0.50
First Term GPA	3.52	0.42	3.52	0.44
First Year GPA	3.56	0.35	3.57	0.33
UMETRICS Variables:				
Ever Research Funded Yrs 2-4	0.66	0.47	0.61	0.49
Research Funded Yr 1	0.29	0.46	0.26	0.44
Research Funded Yr 2	0.43	0.50	0.40	0.49
Research Funded Yr 3	0.50	0.50	0.47	0.50
Research Funded Yr 4	0.46	0.50	0.43	0.49

In main estimation sample: N = 1,563 for men and N = 950 for women.

Table 4: Effect of Cohort Gender Composition on Ph.D. Completion Within 6 Years

	Cohort Gender Comp Measured by:		
	% Female Peers (1)	Ratio F/M (2)	# Female Peers (3)
Female (β_1)	-0.117***	-0.096***	0.023
	(.0421)	(.0364)	(.0625)
Cohort Gender Composition (β_2)	-0.086	-0.030	0.001
	(.0923)	(.0187)	(.0066)
Female*Cohort Gender Composition (β_3)	0.210**	0.067***	-0.007
	(.0941)	(.0209)	(.0081)
Effect of +1 sd in Treatment	0.044	0.070	-0.033
Total Gender Peer Effect on Women ($\beta_2 + \beta_3$)	0.124	0.038	-.005
	[0.25]	[0.03]	[0.34]
Obs	2,512	2,511	2,513

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. The dependent variable in all columns is an indicator for completing the Ph.D. degree within 6 years of initial enrollment. All specifications include: cohort size, age, age-squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs. The reported effect of a +1 sd in treatment corresponds to the marginal effect of a 1 standard deviation increase in the measure of cohort gender composition ($\beta_3 * StdDev$(Cohort Gender Composition)). Total gender peer effect on women corresponds to the marginal effect of cohort gender composition conditional on $Female_i = 1$ ($\beta_2 + \beta_3$) with corresponding p-values shown in brackets.

Table 5: Effect of Cohort Gender Composition on Ph.D. Persistence

	Graduated or Still Enrolled in:				
	Year 2	Year 3	Year 4	Year 5	Year 6
	(1)	(2)	(3)	(4)	(5)
Female	-0.072*	-0.069	-0.077	-0.086*	-0.082
	(.0383)	(.0429)	(.0485)	(.0492)	(.0517)
% Peers Female	0.018	0.018	0.038	-0.010	-0.019
	(.0668)	(.0722)	(.0743)	(.0917)	(.0985)
% Peers Female*Female	0.122*	0.099	0.092	0.092	0.086
	(.0725)	(.0868)	(.1082)	(.1039)	(.1101)
Total Gender Peer Effect on Women	0.139	0.117	0.130	0.082	0.067
	[0.01]	[0.14]	[0.15]	[0.34]	[0.48]
Obs	2,450	2,512	2,512	2,512	2,512

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age-squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs. Total gender peer effect on women corresponds to the marginal effect of cohort gender composition conditional on $Female_i = 1$ $(\beta_2 + \beta_3)$ with corresponding p-values shown in brackets.

Table 6: Effects of Cohort Gender Composition on Ph.D. Completion By Typically Male/Female Programs

	Complete Ph.D. within 6 Years			
	Typically Male (avg % cohort female $\leq$ 36.7%)	Typically Female (avg % cohort female $>$ 36.7%)	Typically Male (Engin., Math, & Physics)	Typically Female (All Other Sciences)
	(1)	(2)	(3)	(4)
Female	-0.180***	-0.212**	-0.043	-0.216***
	(.0592)	(.0854)	(.0758)	(.0775)
% Peers Female	-0.096	-0.135	-0.064	-0.138
	(.1357)	(.1381)	(.1530)	(.1374)
% Peers Female*Female	0.604**	0.336**	0.102	0.386***
	(.2568)	(.1548)	(.2448)	(.1259)
Total Gender Peer Effect on Women	0.508	0.201	0.038	0.248
	[0.12]	[0.17]	[0.88]	[0.03]
Obs	1,275	1,232	936	1,571

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age-squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs. Total gender peer effect on women corresponds to the marginal effect of cohort gender composition conditional on $Female_i = 1$ $(\beta_2 + \beta_3)$ with corresponding p-values shown in brackets. In columns 1-2, typically male programs are those that have an average cohort gender composition $\leq$ 36.7% female. In columns 3-4, typically male programs include all programs in the fields of: Engineering, Mathematics & Statistics, and Physics.

Table 7: Effect of Cohort Gender Composition on Grades

	First Quarter GPA	First Year GPA	
	(1)	(2)	(3)
Female	-0.120***	-0.069***	0.008
	(.0295)	(.0223)	(.0102)
% Peers Female	-0.111	-0.090	-0.031
	(.0840)	(.1152)	(.0660)
% Peers Female*Female	0.217***	0.129**	-0.019
	(.0654)	(.0490)	(.0290)
First Quarter GPA			0.704***
			(.0217)
Total Gender Peer Effect on Women	0.105	0.039	-0.050
	[0.23]	[0.73]	[0.44]
Obs	5,429	5,206	5,206

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age-squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs. Total gender peer effect on women corresponds to the marginal effect of cohort gender composition conditional on $Female_i = 1$ ($\beta_2+\beta_3$) with corresponding p-values shown in brackets.

Table 8: Differential Response to Grades By Gender

	PhD in 6 Yrs		Persist to Yr 2	
	(1)	(2)	(3)	(4)
First Q GPA	0.301***		0.095***	
	(.0489)		(.0157)	
First Q GPA*Female	-0.086		-0.004	
	(.0530)		(.0194)	
First Yr GPA		0.452***		0.105***
		(.0524)		(.0142)
First Yr GPA*Female		-0.085		-0.010
		(.0623)		(.0186)
Female	0.014	0.000	-0.037	0.004
	(.0181)	(.0006)	(.1320)	(.0227)
% Peers Female	-0.094	-0.079	-0.022	-0.008
	(.0984)	(.0956)	(.0361)	(.0293)
% Peers Female*Female	0.173*	0.198**	0.086**	0.081***
	(.0953)	(.0982)	(.0405)	(.0268)
Obs	2,512	2,371	5,032	4,816

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age- squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs.

Table 9: Correlation Between Research Funding and Ph.D. Persistence

	Graduated or Still Enrolled in: Year 2	Year 3	Year 4	Year 5
	(1)	(2)	(3)	(4)
Research Funded in Year 1	0.045***			
	(.0125)			
Research Funded in Year 2		0.142***		
		(.0182)		
Research Funded in Year 3			0.285***	
			(.0190)	
Research Funded in Year 4				0.406***
				(.0182)
Obs	2,044	2,080	2,079	2,045
Cohorts in Sample	09-14	08-13	07-12	06-11

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age- squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs.

Table 10: Correlation Between Research Funding and Ph.D. Completion

	Complete Ph.D. within 6 Years			
	(1)	(2)	(3)	(4)
Ever Research Funded Years 2-4	0.208***			
	(.0563)			
Research Funded in Year 2		0.134		
		(.0864)		
Research Funded in Year 3			0.215***	
			(.0373)	
Research Funded in Year 4				0.269***
				(.0470)
Obs	633	633	969	1,281
Cohorts in Sample	08-09	08-09	07-09	06-09

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age- squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs.

Table 11: Effect of Cohort Gender Composition on Receiving Funding

	Receive Research Funding in: Year 1	Year 2	Year 3	Year 4
	(1)	(2)	(3)	(4)
Female	0.006	-0.073*	-0.022	-0.100
	(.0359)	(.0419)	(.0621)	(.0731)
% Peers Female	0.068	0.005	0.107	0.042
	(.1071)	(.1183)	(.1013)	(.1243)
% Peers Female*Female	-0.065	0.145	0.017	0.215
	(.0860)	(.1012)	(.1346)	(.1809)
Total Gender Peer Effect on Women	0.003	0.150	0.124	0.257
	[0.97]	[0.29]	[0.27]	[0.10]
Obs	2,050	2,079	2,078	2,044
Cohorts in Sample	09-14	08-13	07-12	06-11

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age- squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs. Total gender peer effect on women corresponds to the marginal effect of cohort gender composition conditional on $Female_i = 1$ ($\beta_2 + \beta_3$) with corresponding p-values shown in brackets.

Table 12: Effect of Previous Cohort's Gender Composition on Ph.D. Completion

	Complete PhD in 6 Yrs (1)	(2)	(3)
Female	-0.138***	-0.134***	-0.117**
	(.0504)	(.0521)	(.0570)
% Peers Female	-0.105	-0.086	-0.107
	(.1111)	(.1184)	(.1260)
% Peers Female*Female	0.282***	0.275***	0.335***
	(.0993)	(.1001)	(.1227)
% Previous Cohort Female		-0.005	0.034
		(.0888)	(.0933)
% Previous Cohort Female*Female			-0.105
			(.1359)
Obs	1,982	1,982	1,982

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age- squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs. Regressions in columns (2) and (3) also include previous cohort's size.

Table 13: Summary Statistics on Faculty Gender

	Mean	Std Dev	Min	Max
Faculty Teaching 1st-yr Courses:				
# Faculty	8.00	3.56	1	21
% Faculty Female	0.13	0.13	0	0.5
All Department Faculty:				
# Faculty	34.45	24.32	10	132
% Faculty Female	0.17	0.10	0.03	0.51

Sample includes an unbalanced panel of 26 STEM doctoral programs covering the 2006-2009 school-years for a total of 98 program-year cohorts.

Table 14: Effect of Faculty Gender Composition on Ph.D. Completion

	Complete PhD in 6 Yrs:				
	Main Specification	Faculty Teaching 1st-Yr Courses		All Department Faculty	
	(1)	(2)	(3)	(4)	(5)
Female	-0.103**	-0.103**	-0.101*	-0.103**	-0.083*
	(.0491)	(.0490)	(.0548)	(.0488)	(.0444)
% Peers Female	-0.111	-0.113	-0.113	-0.106	-0.106
	(.1286)	(.1302)	(.1303)	(.1202)	(.1310)
% Peers Female*Female	0.191*	0.191*	0.193*	0.194*	0.195
	(.1081)	(.1086)	(.1046)	(.1104)	(.1568)
% Faculty Female		-0.027	-0.021	-1.517*	-1.515*
		(.2287)	(.2134)	(.8678)	(.8477)
% Faculty Female*Female			-0.017		-0.003
			(.1917)		(.3402)
Obs	1,685	1,685	1,685	1,685	1,685

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age- squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs.

Table 15: Specification Robustness Checks

	Main Specification	LPM	Ph.D. in 7 Yrs	Ph.D. in 6 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.117***	-0.108**	-0.140***	-0.103**	-0.126***	-0.109***	-0.121***
	(.0421)	(.0395)	(.0524)	(.0444)	(.0413)	(.0425)	(.0409)
% Peers Female	-0.086	-0.073	-0.083	-0.099	-0.094	-0.101	-0.090
	(.0923)	(.0874)	(.0962)	(.0954)	(.0918)	(.0921)	(.0911)
% Peers Female*Female	0.210**	0.190**	0.272**	0.172*	0.231**	0.187*	0.219**
	(.0941)	(.0902)	(.1069)	(.1000)	(.0960)	(.0959)	(.0920)
Program Fes	X	X	X	X	X	X	X
Year Fes	X	X	X	X		X	X
University*Year FEs				X			
Linear Time Trend					X		
University-Specific Linear Time Trend						X	
Quadratic Time Trend							X
Linear Model or Probit MLE	Probit	Linear	Probit	Probit	Probit	Probit	Probit
Obs	2,512	2,512	1,983	2,512	2,512	2,512	2,512

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Reported estimates in columns (1) and (3)-(7) are marginal effects derived from a Probit MLE. Estimates in column (2) are obtained via OLS. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age-squared, gender, foreign status, and race/ethnicity indicators.

Table 16: Sample Definition Robustness Checks

	Main Specification	Define Programs Using: CIP Field	Prgm Code	STEM & Non-STEM	Non-STEM Only
	(1)	(2)	(3)	(4)	(5)
Female	-0.117***	-0.036	-0.119**	-0.120***	-0.118
	(.0421)	(.0370)	(.0486)	(.0399)	(.1260)
% Peers Female	-0.086	-0.018	-0.199*	-0.109	-0.252
	(.0923)	(.0610)	(.1135)	(.0873)	(.2403)
% Peers Female*Female	0.210**	0.051	0.200*	0.220**	0.239
	(.0941)	(.0697)	(.1090)	(.0882)	(.3004)
Obs	2,512	3,330	2,432	2,904	391
# Programs	31	32	51	38	7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age-squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs.

Table 17: Robustness Checks - Drop Small Programs

	Drop Programs with Avg Cohort Size ≤ 7	≤ 8	≤ 9	≤ 10	≤ 11
	(1)	(2)	(3)	(4)	(5)
Female	-0.064	-0.109***	-0.117***	-0.135***	-0.139***
	(.0427)	(.0416)	(.0421)	(.0423)	(.0447)
% Peers Female	-0.056	-0.125	-0.086	-0.161*	-0.163*
	(.0782)	(.0861)	(.0923)	(.0837)	(.0896)
% Peers Female*Female	0.116	0.199**	0.210**	0.202**	0.199*
	(.0833)	(.0855)	(.0941)	(.1016)	(.1106)
Obs	3,099	2,803	2,512	2,221	2,100

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age-squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs.

Table 18: Robustness Checks - Drop High-Transfer Programs

	Drop Programs with % Transfer Students				
	≥10%	≥15%	≥20%	≥25%	≥30%
	(1)	(2)	(3)	(4)	(5)
Female	-0.120**	-0.112**	-0.117***	-0.108***	-0.121***
	(.0512)	(.0472)	(.0421)	(.0410)	(.0402)
% Peers Female	-0.112	-0.083	-0.086	-0.085	-0.108
	(.1006)	(.0947)	(.0923)	(.0916)	(.0928)
% Peers Female*Female	0.211*	0.200**	0.210**	0.195**	0.217**
	(.1091)	(.1019)	(.0941)	(.0943)	(.0922)
Obs	2,132	2,388	2,512	2,594	2,689

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age-squared, gender, foreign status, race/ethnicity indicators, year FEs, and program FEs.

Appendix

Table A1: Comparison of Effects in Large and Small Programs

	STEM Fields Only				Non-STEM Fields Only		
	Large Programs (1)	Small Programs (2)	All Programs (3)	Large Cohorts Only (4)	Large Programs (5)	Small Programs (6)	All Programs (7)
Female	-0.117***	-0.003	-0.000	-0.084*	-0.118	0.043	0.045
	(.0421)	(.0591)	(.0584)	(.0462)	(.1260)	(.1237)	(.1206)
% Peers Female	-0.086	-0.037	-0.025	-0.155	-0.252	0.005	0.008
	(.0923)	(.0845)	(.0825)	(.1035)	(.2403)	(.1455)	(.1383)
% Peers Female*Female	0.210**	0.059	0.032	0.145*	0.239	-0.152	-0.138
	(.0941)	(.1126)	(.1099)	(.0849)	(.3004)	(.1786)	(.1740)
Female*Large Program			-0.119***				-0.142
			(.0420)				(.1209)
% Peers Female*Large Program			-0.083				-0.264
			(.0900)				(.2793)
% Peers Female*Fem*Large Pgrm			0.222**				0.249
			(.0949)				(.2538)
Obs	2,512	1,582	4,094	2,705	391	787	1,179

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects derived from a Probit MLE. Standard errors in parentheses are clustered by program. All specifications include: cohort size, age, age-squared, gender, foreign status, and race/ethnicity indicators. Columns (1) and (5) exclude any program with an average cohort size ≤ 9. Columns (2) and (6) exclude any program with an average cohort size > 9. Column (4) includes all STEM programs, but excludes any individual cohort that is ≤ 9.