NBER WORKING PAPER SERIES

NEVERTHELESS SHE PERSISTED? GENDER PEER EFFECTS IN DOCTORAL STEM PROGRAMS

Valerie K. Bostwick
Bruce A. Weinberg

Working Paper 25028
http://www.nber.org/papers/w25028

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2018

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; 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Nevertheless She Persisted? Gender Peer Effects in Doctoral STEM Programs
Valerie K. Bostwick and Bruce A. Weinberg
NBER Working Paper No. 25028
September 2018
JEL No. I23,J16,O3

ABSTRACT

We study the effects of peer gender composition, a proxy for female-friendliness of environment, 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 11.9pp less likely to graduate within 6 years than their male counterparts. A 1 sd increase in the percentage of female students differentially increases the probability of on-time graduation for women by 4.6pp. These gender peer effects function primarily through changes in the probability of dropping out in the first year of a Ph.D. program and are largest in programs that are typically male-dominated.

Valerie K. Bostwick
Department of Economics
Ohio State University
410 Arps Hall
1945 North High Street
Columbus, OH 43210
vkbostwick@gmail.com

Bruce A. Weinberg
Department of Economics
Ohio State University
410 Arps Hall
1945 North High Street
Columbus, OH 43210
and NBER
weinberg.27@osu.edu

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 step 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 the environment in STEM doctoral programs and the effect on gender differences in Ph.D. persistence and completion. We provide a generalizable, quantitative proxy for the climate towards women using peer gender composition. Specifically, we propose that the gender composition of one's peers within an incoming cohort of a particular doctoral program functions as a measure of the female-friendliness of that cohort's environment. We then estimate the effect of within-program variation in cohort gender composition on the gender gap in Ph.D. persistence and on-time graduation rates.

This proxy has the advantage of implying a natural identification strategy. Specifically, within a particular doctoral program at a given university, year-to-year changes in the gender composition of each cohort are plausibly and, to some extent, testably exogenous (Hoxby (2000) employs a similar strategy for grade-school students). 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 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).

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[1] to data from the

[1]The Ohio Longitudinal Data Archive is a project of the Ohio Education Research Center (oerc.osu.edu)

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)[2] and analyze the effects of peer gender composition on both Ph.D. persistence and completion. Furthermore, the detailed trascript 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 11.9pp 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 4.6pp; (3) these effects function almost completely through changes in the probability of dropping out in the first year of a Ph.D. program; and (4) these gender peer effects tend to be largest in programs that are typically male-dominated. We further investigate whether these gender peer effects might be due to differences in learning, competing, or securing financial support. We find no evidence of any differences in financial support due to peer gender composition and, although we find evidence of a small effect of cohort gender composition on grades, the results largely indicate that climate is the mechanism driving the observed gender peer effects.

2 Contribution & Literature Review

This paper offers several significant contributions and adds to two areas of existing research. Our findings contribute to the broad literature on the gender gap in STEM fields. Importantly, we are one of the first papers to provide a measureable proxy for the female-friendliness of a particular environment. In the literature studying doctoral student out-

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.

[2]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 first-year course load.

comes, we are one of the first papers to test for the presence of any type of peer effects.[3] Doctoral students are an understudied group that is of particular interest in the context of investigating the gender gap in STEM fields. They 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). We provide the first causal evidence of gender peer effects among Ph.D. students.

The existing literature on climate and female underrepresentation in STEM is limited and has typically relied on descriptive survey results. However, these survey findings clearly point to a negative impact of the workplace environment on female scientists persistence in STEM fields.[4] 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.

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 (Lundberg et al., 2018; Wu, 2017) and that female economists face many systematic barriers to success (Hengel, 2017; Mengel et al., 2017; Sarsons, 2017).[5] 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 another measure of climate and, importantly, by identifying and estimating the effect of climate on the educational outcomes of female doctoral students.

Ph.D. completion is a notably understudied outcome variable and very few papers have investigated the gender gap in STEM doctoral degrees, likely due to a lack of longitudinal data on students entering doctoral programs (Ceci et al., 2014). However, several studies have found evidence of gender bias in graduate program admissions in STEM fields. Moss-Racusin et al. (2012) and Milkman et al. (2015) each employ audit studies to reveal that STEM faculty members rate applicants to graduate programs as significantly more competent and are more

[3]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.

[4]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.

[5]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).

likely to respond to email correspondence when a prospective student is assigned to a male name.

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. Much of the education research focuses on mentoring (Clark et al., 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. 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 with 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.[6] Fischer (2017) studies classroom peers at a large public university and finds that the presence of higher ability peers in an introductory STEM course has negative effect on STEM major persistence for female students only. Anelli and Peri (2016) study high school students in Italy and find that cohort gender composition has an effect on initial college major choice for male students only. Specifically, the authors find that men are more likely to choose "predominantly male" majors when they are exposed to a higher share of male peers in high school. Similarly, both Kugler et al. (2017) and Astorne-Figari and Speer

[6]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.

(2017) find that the gender composition of majors is correlated with female students major choice and that women are likely to switch out of male-dominated fields.

3 Data

The data include two linked administrative sources. The Ohio Longitudinal Data Archive (OLDA) provides administrative transcript records for all students attending public colleges 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[7] in the Summer/Fall terms of 2005-2015.[8] 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[9]. We define a doctoral "program" to include all students attending the same institution with the same enrollment CIP code[10] 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

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

[8]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.

[9]https://nces.ed.gov/ipeds/cipcode

[10]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.

than 20% of the enrollment for a particular program (over all years), then we also drop that program from the sample.[11]

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 10 students because very small programs may be structured in such a way that a student's incoming cohort is not the relevant peer group.[12] Finally, we restrict the sample to STEM programs.[13]

The final "estimation sample" includes 2,541 student observations, grouped into an unbalanced panel of 33 doctoral programs, representing 6 public universities. Table 1 provides a full list of these 33 programs and their corresponding CIP codes, CIP fields, and summary statistics. Table 2 shows the calculated cohort characteristics for the estimation sample in the top 2 panels, and for the full sample (including all years of the data 2005-2015, non-STEM programs, and small programs) in the bottom panel for reference. 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 the 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 Section 5, we investigate whether our main findings are more salient in programs that typically have a relatively high or low shares of female students. To do this, we calculate for each program the average percent female across all years of the data (2005-2015) as well as the median value of this average across all 33 programs in the estimation sample (38.5% 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

[11]The main results reported in Section 5 are robust to including/excluding programs with more/fewer transfer students. See Table 15.

[12]The main results reported in Section 5 are robust to including/excluding programs with a smaller/larger average cohort size. See Table 14.

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

female." Student and cohort-level summary statistics for these two subsamples are provided in Table 3. By dividing the sample in this way, we see that students in typically female programs are more likely to graduate on-time and are less likely to be foreign-born. Cohorts in typically female programs and typically male programs are very similar in size.

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 4. Note that the different time spans for the two linked data sources mean that each of the funding indictor variables have a different set of cohorts as the support. 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 4 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 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 highly-female cohorts and highly-male cohorts within a given doctoral program. We estimate the following specification:

$$\begin{aligned}\mathbb{P}\left(Y_{ipt}=1\right) =& \beta_1 Female_i + \beta_2 HighlyFemale_{pt} + \beta_3 Female_i * HighlyFemale_{pt} \\ &+ \gamma' X_{ipt} + \delta Z_{pt} + D_p + D_t + \epsilon_{ipt},\end{aligned} \tag{S1}$$

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; $HighlyFemale_{pt}$, which equals the percent of students 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 women in the cohort; the ratio of women to men in the cohort; and an indicator variable that equals 1 if the cohort has a fraction of female students that is above the mean for that program (over all years 2005-2015).

The coefficient β_1 can be interpreted as the percentage point difference in on-time graduation probabilities for women versus men in highly-male cohorts (cohorts where the percent of female students is zero). The coefficient β_2 indicates the difference in graduation probabilities for men in highly-female cohorts versus men in highly-male cohorts. Finally, the coefficient β_3 is the differential effect on women versus men of being in a highly-female cohort. The model is estimated using a Probit maximum likelihood estimator (Probit MLE),[14] 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 graduation rates for that cohort. An example violation of this assumption might be the appointment of a new department chair 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 cohort gender composition within programs. In the absence of time trends, the influence of the new department chair or faculty member would have to fluctuate year-to-year along with cohort gender composition in order to pose a threat to identification.

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

[14] Results estimated using a linear probability model are qualitatively very similar. See Table 13.

fixed-effects reveals no evidence of path-dependence in this variable.[15]

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 include: 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.[16]

5 Results

Table 5 shows the marginal effects results of 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 the $HighlyFemale_{pt}$ variable is measured as the percent of students in the cohort who are female. These results show that there is a significant gender gap in Ph.D. completion in cohorts with few women. Women in cohorts with no female peers are 11.9pp 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 10% female students in a cohort, men are 1.10pp less likely to graduate on-time (although this effect is statistically insignificant in most specifications) and the differential effect on women is 2.24pp (and statistically significant at the 5% level). This indicates that the effect of an additional 10% female students for a woman is a 1.14pp 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 4.63pp.

Columns (2)-(4) of Table 5 experiment with alternative definitions for highly female cohorts. Column (2) measures highly female cohorts as the ratio of women to men in the cohort. In column (3), we calculate the average over all years (2005-2015) of the percent of women in each program and define the $HighlyFemale_{pt}$ variable to be an indicator that equals one if the percent of female students in year t is above the overall average for program p. This measure incorporates the notion that program norms may be important in the

[15] The Wald test statistic for the lagged % cohort female variable is -1.08.

[16] 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.

salience of gender composition effects. That is, a cohort with 40% women might seem "highly female" in a typically male program (such as Physics) but that same level might feel "highly male" in a program with a higher average gender balance (such as Psychology). The results in both columns (2) and (3) 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 highly male cohorts and that this gap is significantly diminished in cohorts with more female peers.

In column (4) of Table 5, we implement a count measure of gender composition and set $HighlyFemale_{pt}$ equal to the number of women in the cohort. Interestingly, these results show no evidence of a linear effect of the number of women in a program on the probability of Ph.D. completion for either gender. 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.

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 majors 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 with especially low fractions of female students. As described in Section 3, we divide the main estimation sample into two subsamples and categorize programs with an average percent female that is below the sample median (38.5%) 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 Table 6. These results indicate that the effect of cohort gender composition on Ph.D. completion is driven largely by typically male programs. In these programs, the gender gap in Ph.D. completion is even larger. Women are 15.8pp 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 8.82pp for women relative to men. The results for typically female programs are similar, but looking across the specifications in columns (1)-(4), the estimates for this sample are smaller in magnitude and not consistently significant. Thus, it appears that the effect of peer gender on female Ph.D. success rates is largely driven by those programs that have the highest rates of female underrepresentation within the realm of STEM doctoral programs.

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 nearly 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,

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

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 7 shows the marginal effects results of estimating model S2 using a Probit MLE. The top panel is estimated using the full estimation sample. In Panels B and C, we estimate model S2 using the typically male and typically female subsamples, respectively. In this table (and all further tables) we show results using our preferred specification where $HighlyFemale_{pt}$ is defined as the percent of students 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 2-6, respectively. For example, in column (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.

These results indicate that nearly all of the gender composition effect is present by the beginning of program-year 2. 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.[17] Women in cohorts with no female peers are 10.2pp less likely to make it to the second year of a doctoral program than their male peers. That is equivalent to saying that women in cohorts with no female peers are 10.2pp 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 3.68pp in the first year of Ph.D. enrollment. It is clear from panels B and C that these persistence results are again being driven by the subsample of typically male programs.

[17]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.

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 is driven by an improvement to how female-friendly the climate is in a given cohort.

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 et al., 2003), so that 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); Kugler et al. (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.[18] The raw distribution of

[18]Including these additional cohorts for whom we observe less than 6 years of data should not influence

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.[19] 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 following model,

$$
\begin{aligned}
Y_{ipt} = &\beta_1 Female_i + \beta_2 HighlyFemale_{pt} + \beta_3 Female_i * HighlyFemale_{pt} \\
&+ \gamma' X_{ipt} + \delta Z_{pt} + D_t + D_p + \epsilon_{ipt},
\end{aligned}
\tag{S3}
$$

where Y_{ipt} is a measure of individual i's first year grades. We measure this alternately as first term GPA or first year GPA. All other variables are unchanged from model S1. We estimate this model with an Ordinary Least Squares (OLS) estimator.

Column (1) of Table 8 shows the results of estimating model S3 with first term GPA as the dependent variable. These estimates show that women in cohorts with no female peers have first term GPAs that are 0.11 grade points lower than their male peers (on a 4-point scale). At the sample mean of 3.53, 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. Column (2) shows a similar effect of gender composition on GPA at the end of the first year, but column (3) indicates that the effect on first year grades is entirely captured by the first term GPA.

In Table 9, 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 7). 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 men's. 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 8 and 9 indicate that peer gender composition has a small effect on first term GPA such that women have worse grades than men in highly-male

our analysis of first year grades. This is particularly relevant because Table 7 shows that the effect of cohort gender composition functions primarily through dropout decisions in the first year of enrollment.

[19] In this figure, highly female cohorts are defined using the same indicator variable as applied in column (3) of Tables 5 and 6.

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 9 show that a 1 point increase in GPA is associated with an increase in the probability of on-time graduation by 29.4pp for men and 22.7pp for women.[20] 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 1.15pp. This accounts for, at most, a quarter of the total differential effect of peer gender composition shown in Table 5.

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). Using the linked sample of UMETRICS data on students supported through research projects, we first verify that this correlation holds in our setting. We model this relationship by,

$$\mathbb{P}\left(Y_{ipt}^{r}=1\right)=\beta^{r}Funding_{i}^{r-1}+\gamma^{r\prime}X_{ipt}+\delta^{r}Z_{pt}+D_{t}^{r}+D_{p}^{r}+\epsilon_{ipt}^{r}, \tag{S4}$$

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 10, 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 11 shows the relationship between obtaining research support and the probability of on-time graduation for STEM doctoral students. These results are less precise as there are very few cohorts for whom we can observe both UMETRICS employment and 6-year graduation

[20]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.

rates.

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 5. 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 12. These estimates provide no evidence that peer gender composition has any effect on research funding in any year for either gender. The marginal effects are small, inconsistent in sign, and statistically insignificant. Clearly, these findings along with our main results do not support a research funding mechanism.

Another set of potential explanations for our main findings focus on mentoring and the gender mix of faculty. We have explored mentoring by relating completion and retention to the gender composition of the cohort that entered one year earlier, under the assumption that the older cohort interacts with the younger cohort. Our data show no effect of the older cohort on completion or retention of the younger cohort.[21] The gender mix of faculty is an important factor and one that we plan to explore in future work, but 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.

The analysis in this section indicates that peer gender composition does not impact students' financial support through research funding and that there is only 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 at most one quarter of the total effect of peer gender composition on Ph.D. completions. Having ruled out any observable mechanisms, we are left to conclude that our measure of peer gender composition is capturing the unobservable: changes in the climate of each cohort. This implies that when cohort gender composition is particularly high, the intangible climate towards women improves, thereby increasing female students' persistence and on-time graduation. This persistence occurs despite the fact that these women experience no change in the prospect of financial support and only a marginal improvement in first year grades.

[21]This may be because the interaction with older cohorts is more prevalent in later years of the program when students are less likely to drop out.

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 13-15. We first show that our main results are not sensitive to the use of a Probit MLE. In column (2) of Table 13 we estimate the main specification in model S1 as a linear probability model using an OLS estimator.[22] These estimates are very similar in both magnitude and precision to the main results.

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 7 showing that the majority of the peer gender effect is driven by dropping out in the first year of enrollment.

In columns (4)-(5) of Table 13, we implement alternate definitions of doctoral programs. 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 (4), 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 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 (5) of Table 13, we instead disaggregate the CIP codes into university-specific program identifiers and use these codes to define each program.[23] These results are very similar in both size and precision to the main results shown in column (1).

In columns (6) and (7) of Table 13, we test whether our results hold in non-STEM doctoral programs. In column (6) of Table 13 we include both STEM and non-STEM programs (conditional on an average cohort size of more than 10 students) in the sample. These results are largely similar to the main findings in column (1). In column (6), 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

[22] Column (1) of Table 13 replicates the main findings in column (1) of Table 5 for reference.

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

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.

As noted above, in the main estimation sample, we limit the data to include only programs with an average cohort size greater than 10. 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 14 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 5. These results show that the main findings are robust to the inclusion/exclusion of smaller/larger 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 15, 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.

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 in cohorts with no female peers are 11.9pp 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 4.63pp.

We find that this effect is largely driven by students in typically-male programs (with less than 38.5% female student in the average cohort) and by 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 our results largely cannot be explained by women learning or competing more successfully in cohorts with more female peers. However, 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, despite having no significant effect on learning or financial support. Taken together, these findings indicate that peer gender composition can 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. This strategy might be applied to a number of different contexts where individuals are consistently grouped into defined cohorts.

References

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

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

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

Amanda Bayer and Cecilia Elena Rouse. Diversity in the Economics Profession: A New Attack on an Old Problem. *Journal of Economic Perspectives*, 30(4):221–242, nov 2016. ISSN 0895-3309. doi: 10.1257/jep.30.4.221. URL http://pubs.aeaweb.org/doi/10.1257/jep.30.4.221.

Bethany A. Bell-Ellison and Robert F. Dedrick. What do Doctoral Students Value in their Ideal Mentor? *Research in Higher Education*, 49(6):555–567, sep 2008. ISSN 0361-0365. doi: 10.1007/s11162-008-9085-8. URL http://link.springer.com/10.1007/s11162-008-9085-8.

Stephen J. Ceci, Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams. Women in Academic Science: A Changing Landscape. *Psychological Science in the Public Interest*, 15(3):75–141, dec 2014. ISSN 1529-1006. doi: 10.1177/1529100614541236. URL http://www.ncbi.nlm.nih.gov/pubmed/26172066http://journals.sagepub.com/doi/10.1177/1529100614541236.

Richard A. Clark, Sherry L. Harden, and W. Brad Johnson. Mentor Relationships in Clinical Psychology Doctoral Training: Results of a National Survey. *Teaching of Psychology*, 27(4):262–268, nov 2000. ISSN 0098-6283. doi: 10.1207/S15328023TOP2704_04. URL http://top.sagepub.com/lookup/doi/10.1207/S15328023TOP2704{_}04.

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

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

Stefanie Fischer. The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence. *Labour Economics*, 46:211–226, jun 2017. ISSN 09275371. doi: 10.1016/j.labeco.2017.02.003. URL `http://linkinghub.elsevier.com/retrieve/pii/S0927537116303852`.

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

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

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

U. Gneezy, M. Niederle, and A. Rustichini. Performance in Competitive Environments: Gender Differences. *The Quarterly Journal of Economics*, 118(3):1049–1074, aug 2003. ISSN 0033-5533. doi: 10.1162/00335530360698496. URL `https://academic.oup.com/qje/article-lookup/doi/10.1162/00335530360698496`.

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

Chris M. Golde and Timothy M. Dore. At cross purposes: what the experiences of today's doctoral students reveal about doctoral education. Technical report, 2001. URL `https://eric.ed.gov/?id=ED450628`.

Leigh Hall and Leslie Burns. Identity Development and Mentoring in Doctoral Education. *Harvard Educational Review*, 79(1):49–70, apr 2009. ISSN 0017-8055. doi: 10.17763/haer.79.1.wr25486891279345. URL `http://hepgjournals.org/doi/10.17763/haer.79.1.wr25486891279345`.

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

Christiana Hilmer and Michael Hilmer. 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, 2007. ISSN 00028282. doi: 10.1257/aer.97.2.422.

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

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

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

Adriana D Kugler, Catherine H Tinsley, Olga Ukhaneva, George Akerlof, David Autor, Kirsten Butcher, David Figlio, Nicole Fortin, Donna Ginther, Claudia Goldin, Harry Holzer, Lisa Kahn, Shulamit Kahn, Victor Lavy, John Mayo, and Bruce Sacerdote. CHOICE OF MAJORS: ARE WOMEN REALLY DIFFERENT FROM MEN? Choice of Majors: Are Women Really Different from Men? 2017. URL http://www.nber.org/papers/w23735.

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

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

J Main. 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, 2014. ISSN 1090-7009. doi: 10.1353/rhe.2014.0019. URL http://muse.jhu.edu/journals/rhe/summary/v037/37.3.main.html.

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

Katherine L. Milkman, Modupe Akinola, and Dolly Chugh. What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology*, 100(6):1678–1712, nov 2015. ISSN 1939-1854. doi: 10.1037/apl0000022. URL `http://www.ncbi.nlm.nih.gov/pubmed/25867167http://doi.apa.org/getdoi.cfm?doi=10.1037/apl0000022`.

Corinne A Moss-Racusin, John F Dovidio, Victoria L Brescoll, Mark J Graham, and Jo Handelsman. Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America*, 109(41):16474–9, oct 2012. ISSN 1091-6490. doi: 10.1073/pnas.1211286109. URL `http://www.ncbi.nlm.nih.gov/pubmed/22988126http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3478626`.

Maresi Nerad. The PhD in the US: Criticisms, Facts, and Remedies. *Higher Education Policy*, 17(2):183–199, jun 2004. ISSN 0952-8733. doi: 10.1057/palgrave.hep.8300050. URL `http://link.springer.com/10.1057/palgrave.hep.8300050`.

Michael T. Nettles. Success in doctoral programs: Experiences of minority and white students. *American Journal of Education*, 98(4):494, aug 1990. ISSN 0195-6744. doi: 10.1086/443974. URL `http://www.journals.uchicago.edu/doi/10.1086/443974`.

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

Ben Ost. The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review*, 29:923–934, 2010. doi: 10.1016/j.econedurev.2010.06.011. URL `http://ac.els-cdn.com/S0272775710000762/1-s2.0-S0272775710000762-main.pdf?{_}tid=2fe9215a-26ca-11e7-967a-00000aab0f26{&}acdnat=1492803728{_}3720f277ca0608a330d214cf8e260ff3`.

Michele Pezzoni, Jacques Mairesse, Paula Stephan, Julia Lane, G Burnett, and K Burnett. Gender and the Publication Output of Graduate Students: A Case Study. *PLOS ONE*, 11(1):e0145146, jan 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0145146. URL `http://dx.plos.org/10.1371/journal.pone.0145146`.

Kevin Rask and Jill Tiefenthaler. The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review*, 27(6):

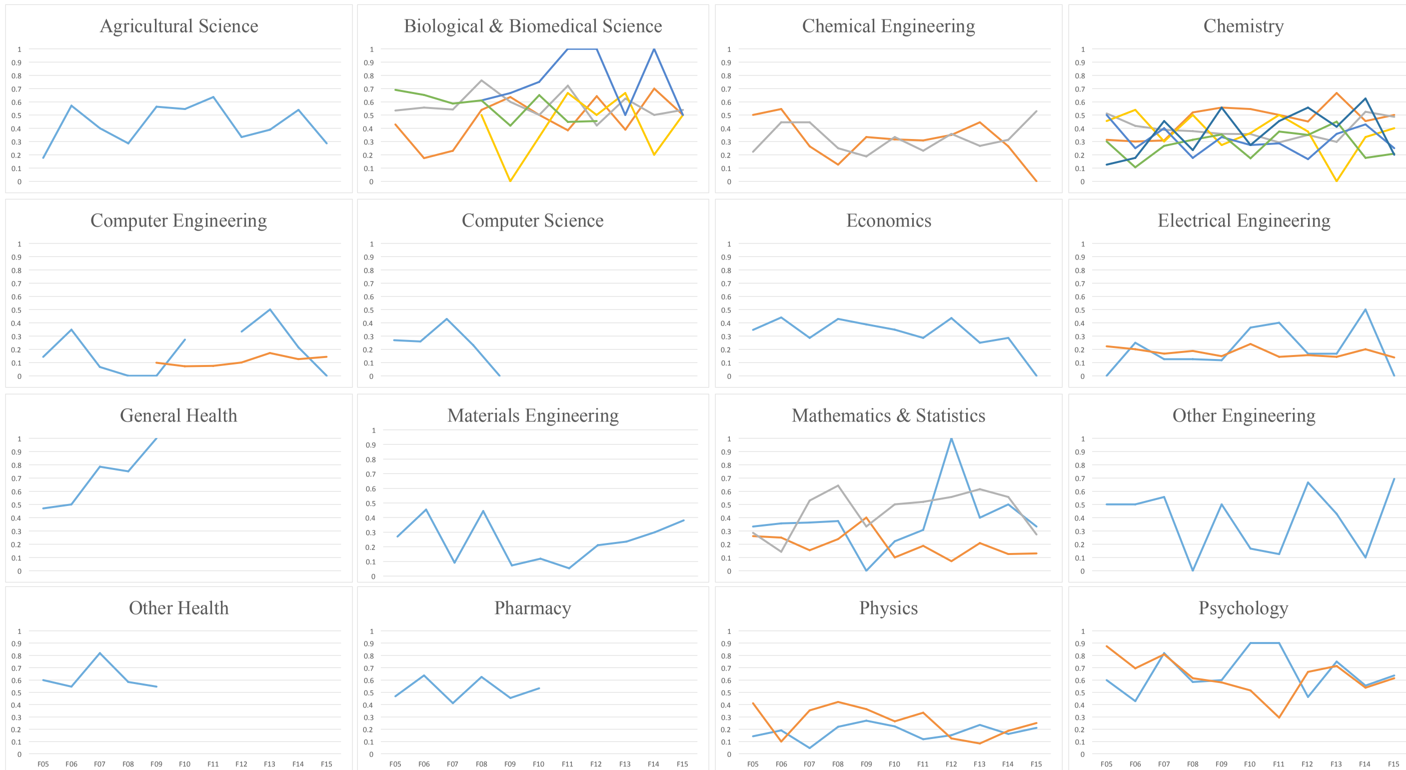
676–687, 2008. ISSN 02727757. doi: 10.1016/j.econedurev.2007.09.010. URL http://www.sciencedirect.com/science/article/pii/S0272775707001185.

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

Todd Stinebrickner and Ralph Stinebrickner. Learning about Academic Ability and the College Dropout Decision. *Journal of Labor Economics*, 30(4):707–748, oct 2012. ISSN 0734-306X. doi: 10.1086/666525. URL http://www.journals.uchicago.edu/doi/10.1086/666525.

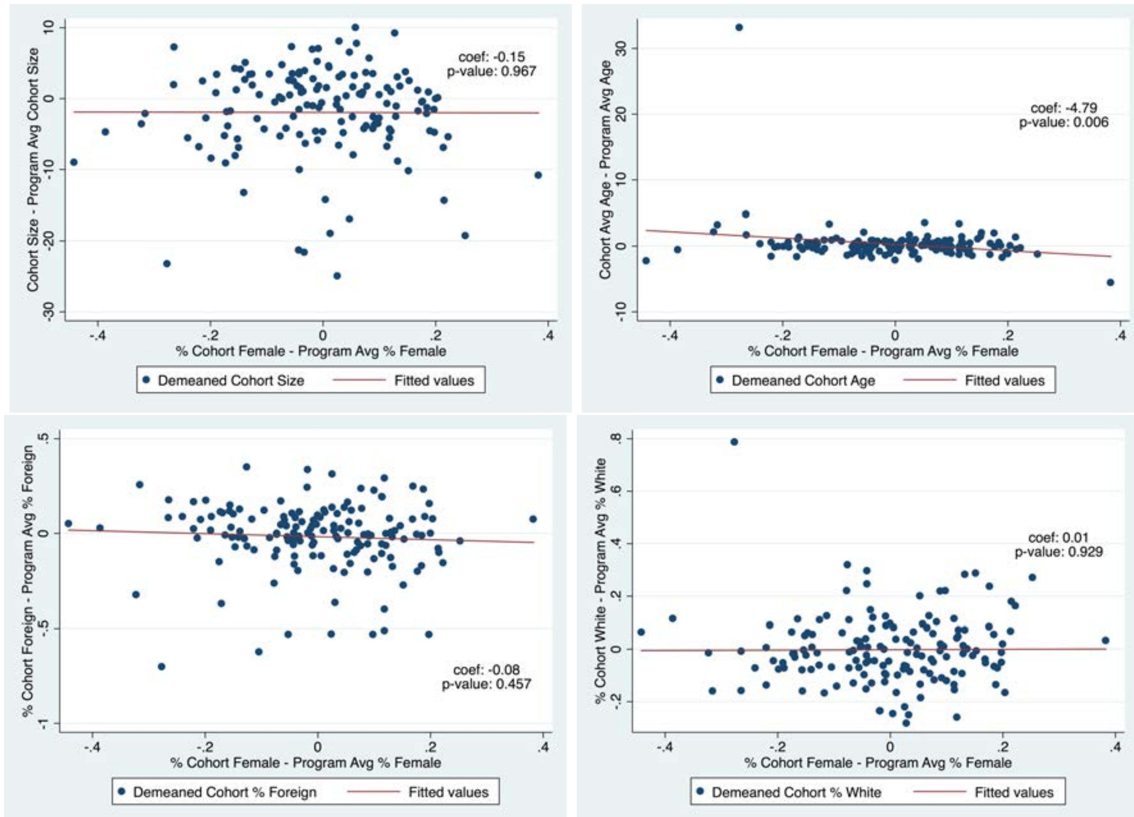
Alice H. Wu. Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum. dec 2017. URL https://www.ssrn.com/abstract=3051462.

Figure 1: Trends in Cohort Gender Composition By Field



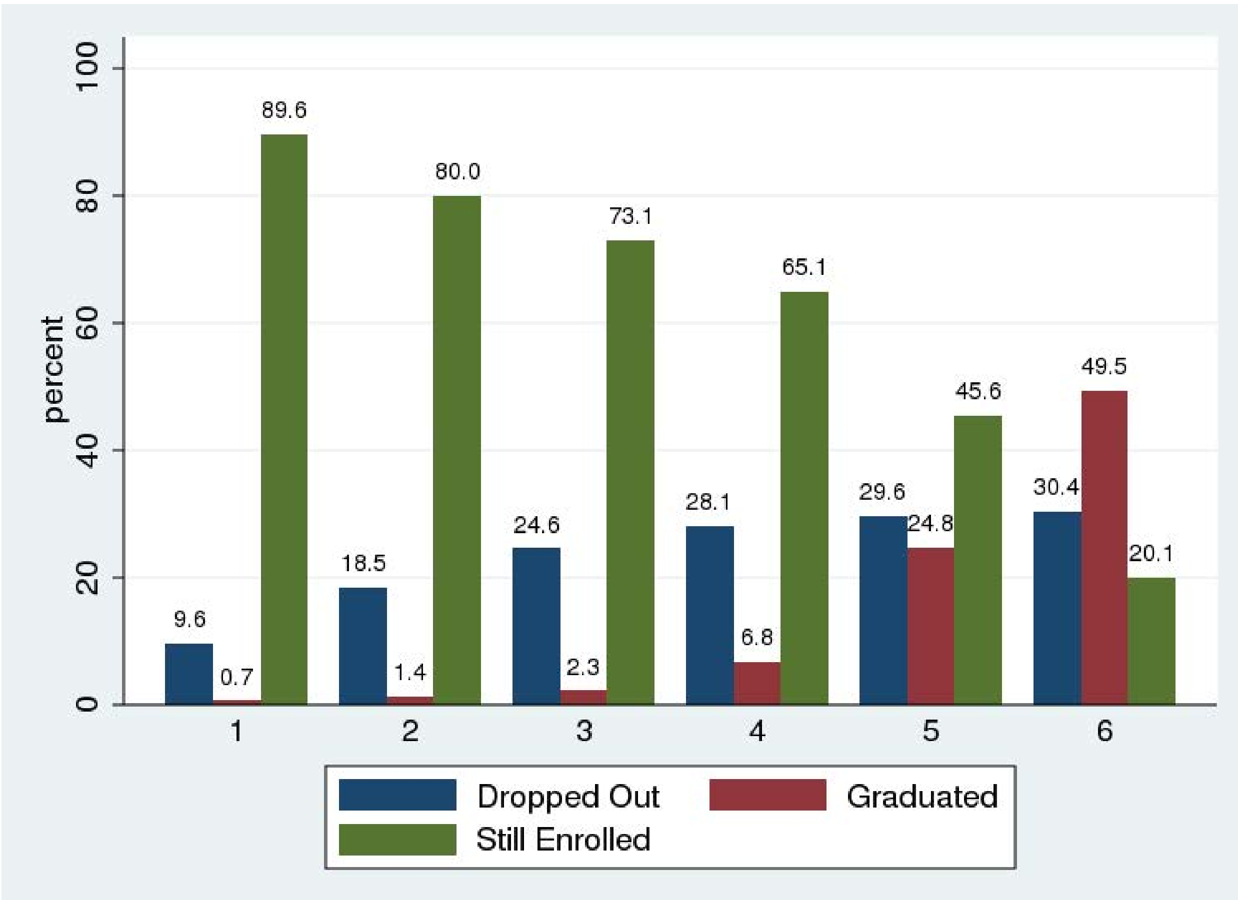
Each panel represents a CIP Field and each line within a panel represents a unique doctoral program in the main estimation sample. The x-axis measures incoming cohorts over time and the y-axis measures the share of each incoming cohort that is female.

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



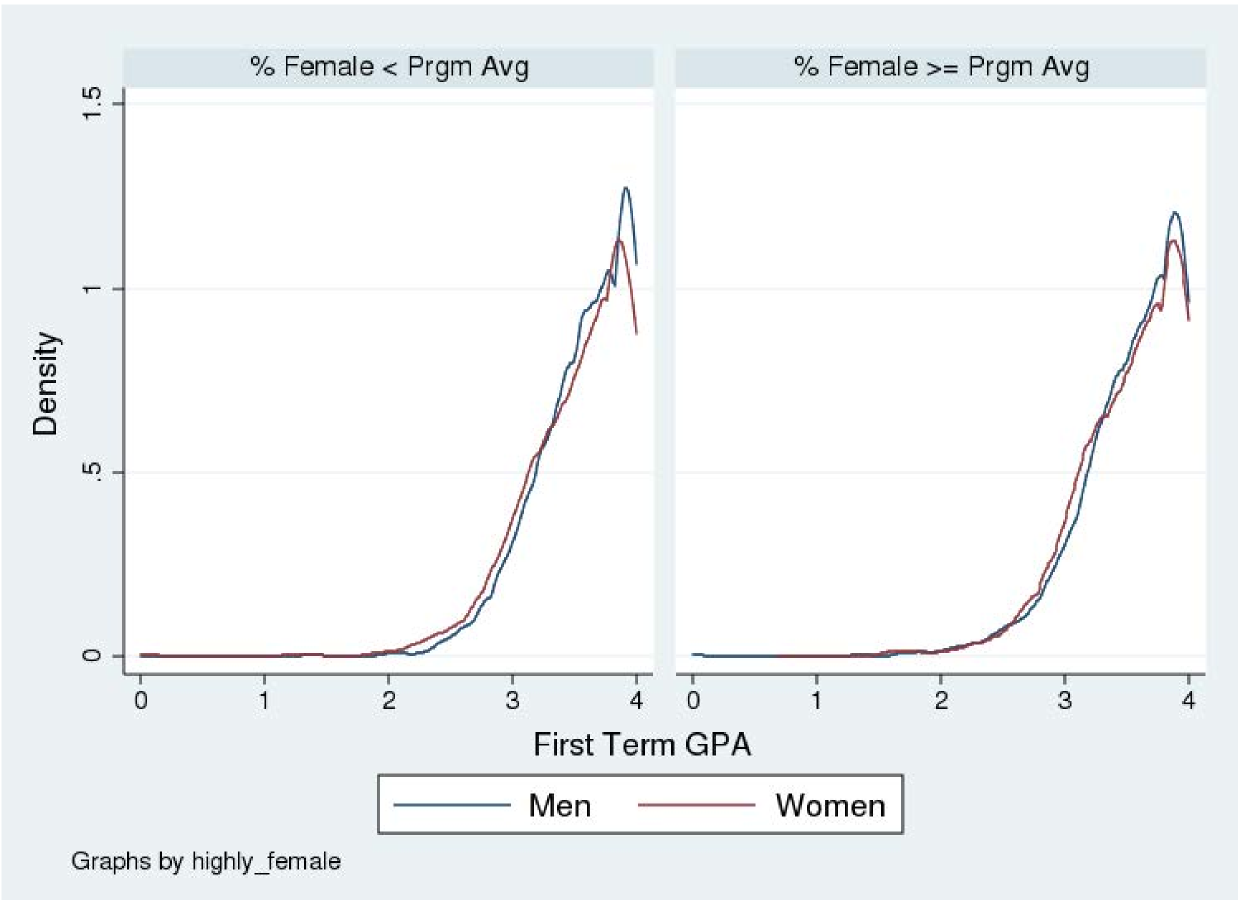
In each panel above, a point represents a cohort within a program. The x-axis measures the percent of each cohort that is female minus the average percent female in the program over all years of the data. The y-axis in each panel represent a different covariate, also demeaned at the program level. Proceeding 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 compositon.

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
Agricultural Science	10103	Agricultural Economics	15.1	44%	1
Biological and Biomedical Sciences	260202	Biochemistry	15.8	44%	1
Biological and Biomedical Sciences	260499	Cell/Cellular Biology and Anatomical Sciences, Other	20.1	57%	1
Biological and Biomedical Sciences	260907	Cardiovascular Science	12.0	44%	1
Biological and Biomedical Sciences	260911	Oncology and Cancer Biology	10.7	67%	1
Biological and Biomedical Sciences	269999	Biological and Biomedical Sciences, Other	18.6	55%	2
Chemical Engineering	140701	Chemical Engineering	16.4	32%	1
Chemical Engineering	143201	Polymer/Plastics Engineering	13.3	33%	1
Chemistry	400501	Chemistry, General	23.9	39%	4
Chemistry	400507	Polymer Chemistry	20.9	28%	1
Chemistry	400599	Chemistry, Other	13.6	36%	1
Computer Engineering	140901	Computer Engineering, General	26.8	16%	2
Computer Science	110101	Computer and Information Sciences, General	24.2	28%	1
Economics (Social Science)	450601	Economics, General	24.5	34%	1
Electrical, Electronics, and Communications Engineering	141001	Electrical and Electronics Engineering	23.6	18%	2
General Health/Public Health	512202	Environmental Health	12.9	62%	1
Materials Engineering	141801	Materials Engineering	17.9	24%	1
Mathematics and Statistics	270101	Mathematics, General	15.0	25%	2
Mathematics and Statistics	270501	Statistics, General	16.1	46%	1
Other Engineering	140501	Bioengineering and Biomedical Engineering	10.8	39%	1
Other Health	511401	Medical Scientist	11.0	62%	1
Pharmacy	512001	Pharmacy	13.6	51%	1
Physics	400801	Physics, General	25.3	23%	2
Psychology	420101	Psychology, General	19.4	63%	2

CIP Codes highlighted in gray represent programs that are typically male (i.e. have an average % cohort female <= 38.5%)

Table 2: Cohort Characteristics

	Mean	Std Dev	Min	Max
Estimation Sample (weighted by # students) $N = 2,541$ students				
STEM Field	1	0	1	1
Cohort Size	21.78	10.57	1	49
# Female in Cohort	8.17	5.38	0	23
% Female in Cohort	.381	.187	0	1
Ratio Female/Male	.831	.848	0	7
Estimation Sample (unweighted) $N = 151$ cohorts				
STEM Field	1	0	1	1
Cohort Size	16.83	9.16	1	49
# Female in Cohort	6.40	4.69	0	23
% Female in Cohort	.383	.207	0	1
Ratio Female/Male	.866	.955	0	7
Full Sample (unweighted) $N = 1,529$ cohorts				
STEM Field	.699	.459	0	1
Cohort Size	7.60	7.68	1	80
# Female in Cohort	3.31	3.25	0	28
% Female in Cohort	.489	.305	0	1
Ratio Female/Male	1.10	1.27	0	11

Table 3: Summary Statistics By Typically Male/Female Programs

	Typically Male Programs		Typically Female Programs	
	Mean	Std Dev	Mean	Std Dev
Student-Level Characteristics				
Ph.D. in 6 Yrs	0.46	0.50	0.54	0.50
Age	24.92	3.64	25.18	4.10
Foreign	0.60	0.49	0.40	0.49
First Term GPA	3.56	0.41	3.49	0.45
Obs	1,288 students		1,253 students	
Cohort-Level Characteristics				
Cohort Size	16.95	8.45	16.71	9.88
% Female in Cohort	0.26	0.14	0.51	0.19
# Female in Cohort	4.43	2.89	8.40	5.31
Obs	76 cohorts		75 cohorts	

Typically male programs are those that have an average cohort gender composition <= 38.5% female. Typically female programs are those that have an average cohort gender composition > 38.5% female.

Table 4: Summary Statistics

	Male		Female	
	Mean	Std Dev	Mean	Std Dev
PhD in 6 Yrs	0.49	0.500	0.50	0.500
Yrs to Graduate	5.45	1.232	5.39	1.261
Drop Out (by end of 6 yrs)	0.30	0.460	0.31	0.461
Still Enrolled (by end of 6 yrs)	0.21	0.406	0.19	0.392
# Yrs Enrolled	4.38	2.078	4.35	2.031
Age (Yr Enrolled - Birth Yr)	25.24	3.977	24.73	3.685
Foreign	0.51	0.500	0.49	0.500
First Term GPA	3.53	0.429	3.53	0.440
First Year GPA	3.56	0.353	3.58	0.336
UMETRICS Variables:				
Ever Research Funded Yrs 2-4	0.66	0.475	0.59	0.493
Research Funded Yr 1	0.29	0.455	0.26	0.437
Research Funded Yr 2	0.44	0.497	0.41	0.491
Research Funded Yr 3	0.57	0.495	0.55	0.498
Research Funded Yr 4	0.57	0.496	0.54	0.499

In main estimation sample: N = 1,574 for men and N = 967 for women.

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

	Complete Ph.D. within 6 Years			
	% Female	Ratio F/M	% Female > program avg	# Female
	(1)	(2)	(3)	(4)
Female	-0.119 ***	-0.081 **	-0.071 **	0.035
	(.0449)	(.0353)	(.0290)	(.0673)
Highly Female Cohort	-0.110	-0.040 **	-0.041	0.001
	(.1067)	(.0195)	(.0269)	(.0068)
Female*Highly Female	0.224 **	0.061 ***	0.082 *	-0.007
	(.1002)	(.0206)	(.0443)	(.0083)
Obs	2,541	2,541	2,541	2,541

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

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

	Complete Ph.D. within 6 Years			
	% Female	Ratio F/M	% Female > program avg	# Female
	(1)	(2)	(3)	(4)
Panel A: Typically Male Programs (N=1,287)				
Female	-0.158 **	-0.118 **	-0.128 **	-0.112 *
	(.0675)	(.0500)	(.0528)	(.0587)
Highly Female Cohort	-0.162	-0.080	-0.070 **	-0.002
	(.1109)	(.0605)	(.0298)	(.0080)
Highly Female*Female	0.426 **	0.184 ***	0.150 **	0.013 *
	(.1865)	(.0700)	(.0609)	(.0078)
Panel B: Typically Female Programs (N=1,249)				
Female	-0.185 *	-0.112 *	-0.041	0.130
	(.1109)	(.0677)	(.0401)	(.1039)
Highly Female Cohort	-0.081	-0.032	0.009	0.005
	(.1839)	(.0271)	(.0419)	(.0102)
Highly Female*Female	0.314 *	0.070 **	0.028	-0.014
	(.1838)	(.0305)	(.0728)	(.0102)

* 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 7: 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)
Panel A: Full Estimation Sample					
Female	-0.102 **	-0.087 *	-0.105 *	-0.109 **	-0.101 *
	(.0464)	(.0510)	(.0552)	(.0534)	(.0563)
% Cohort Female	-0.023	-0.010	-0.002	-0.062	-0.078
	(.0747)	(.0781)	(.0796)	(.0961)	(.1030)
% Cohort Female*Female	0.178 **	0.136	0.149	0.155	0.148
	(.0779)	(.0991)	(.1202)	(.1123)	(.1200)
Obs	2,467	2,529	2,541	2,541	2,541
Panel B: Typically Male Programs					
Female	-0.183 *	-0.186 **	-0.204 **	-0.141 *	-0.106
	(.0948)	(.0811)	(.0799)	(.0794)	(.0784)
% Cohort Female	-0.218 **	-0.130	-0.158	-0.173	-0.216
	(.1068)	(.1164)	(.1302)	(.1461)	(.1557)
% Cohort Female*Female	0.512 **	0.478 **	0.534 **	0.336	0.234
	(.2001)	(.2173)	(.2501)	(.2441)	(.2334)
Obs	1,287	1,287	1,287	1,287	1,287
Panel C: Typically Female Programs					
Female	-0.041	-0.036	-0.085	-0.153	-0.137
	(.1104)	(.0823)	(.0922)	(.1198)	(.1103)
% Cohort Female	0.103	0.033	0.041	-0.044	-0.018
	(.0754)	(.1018)	(.1174)	(.1474)	(.1434)
% Cohort Female*Female	0.023	0.028	0.078	0.196	0.176
	(.1074)	(.1262)	(.1583)	(.1877)	(.1766)
Obs	1,179	1,241	1,253	1,253	1,253

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

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 8: Effect of Cohort Gender Composition on Grades

	First Quarter GPA	First Year GPA	
	(1)	(2)	(3)
Female	-0.113 ***	-0.059 **	0.011
	(.0337)	(.0256)	(.0104)
% Cohort Female	-0.120	-0.117	-0.046
	(.0904)	(.1229)	(.0694)
% Cohort Female*Female	0.210 **	0.116 *	-0.017
	(.0782)	(.0624)	(.0314)
First Quarter GPA			0.714 ***
			(.0212)
Obs	5,425	5,195	5,195

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

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.

Table 9: Differential Response to Grades By Gender

	PhD in 6 Yrs		Persist to Yr 2	
	(1)	(2)	(3)	(4)
First Q GPA	0.294 ***		0.098 ***	
	(.0483)		(.0158)	
First Q GPA*Female	-0.067		-0.002	
	(.0531)		(.0189)	
First Yr GPA		0.446 ***		0.110 ***
		(.0527)		(.0145)
First Yr GPA*Female		-0.056		-0.014
		(.0649)		(.0190)
Female	0.009	0.000	-0.090	0.004
	(.0160)	(.0003)	(.1162)	(.0241)
% Cohort Female	-0.122	-0.088	-0.043	-0.017
	(.1120)	(.1121)	(.0407)	(.0329)
% Cohort Female*Female	0.189 *	0.180 *	0.115 ***	0.087 ***
	(.1023)	(.1020)	(.0427)	(.0311)
Obs	2,541	2,394	5,021	4,799

* 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. Persistence

	Graduated or Still Enrolled in:			
	Year 2	Year 3	Year 4	Year 5
	(1)	(2)	(3)	(4)
Research Funded in Year 1	0.041 ***			
	(.0122)			
Research Funded in Year 2		0.070 ***		
		(.0132)		
Research Funded in Year 3			0.069 ***	
			(.0153)	
Research Funded in Year 4				0.036 ***
				(.0054)
Obs	1,983	1,918	1,766	1,437
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 11: Correlation Between Research Funding and Ph.D. Completion

	Complete Ph.D. within 6 Years			
	(1)	(2)	(3)	(4)
Ever Research Funded Years 2-4	0.241 ***			
	(.0559)			
Research Funded in Year 2		0.110		
		(.0981)		
Research Funded in Year 3			0.099 **	
			(.0422)	
Research Funded in Year 4				0.017
				(.0387)
Obs	617	587	821	978
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 12: Effect of Cohort Gender Composition on Receiving Funding

	Ever Research Funded Years 2-4	Receive Research Funding in: Year 1	Year 2	Year 3	Year 4
	(1)	(2)	(3)	(4)	(5)
Female	-0.095 *	0.013	-0.052	0.052	-0.036
	(.0541)	(.0371)	(.0518)	(.0847)	(.0977)
% Cohort Female	-0.066	0.070	0.068	0.142	-0.041
	(.1124)	(.1159)	(.1470)	(.1150)	(.1744)
% Cohort Female*Female	0.161	-0.074	0.070	-0.143	0.139
	(.1602)	(.0939)	(.1141)	(.1711)	(.2318)
Obs	1,362	1,989	1,922	1,745	1,554
Cohorts in Sample	08-11	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 13: Robustness Checks

	Main Specification	LPM	Ph.D in 7 Yrs	Define Programs Using: CIP Field	Define Programs Using: Prgm Code	STEM & Non-STEM	Non-STEM Only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.119 ***	-0.107 **	-0.164 ***	-0.064	-0.131 **	-0.121 ***	-0.179
	(.0449)	(.0420)	(.0578)	(.0444)	(.0542)	(.0425)	(.1365)
% Cohort Female	-0.110	-0.089	-0.121	-0.082	-0.173	-0.146	-0.419 *
	(.1067)	(.0986)	(.1079)	(.0906)	(.1243)	(.0991)	(.2457)
% Cohort Female*Female	0.224 **	0.195 **	0.328 ***	0.132 *	0.247 **	0.233 **	0.344
	(.1002)	(.0940)	(.1106)	(.0794)	(.1176)	(.0931)	(.3256)
Obs	2,541	2,541	2,015	3,414	2,448	2,898	357
# Programs	33	33	32	34	52	40	7

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

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, race/ethnicity indicators, year FEs, and program FEs.

Table 14: Robustness Checks - Drop Small Programs

	Drop Programs with Avg Cohort Size				
	< 8	< 9	< 10	< 11	< 12
	(1)	(2)	(3)	(4)	(5)
Female	-0.091 **	-0.115 ***	-0.119 ***	-0.135 ***	-0.134 ***
	(.0431)	(.0440)	(.0449)	(.0458)	(.0477)
% Cohort Female	-0.098	-0.122	-0.110	-0.174 *	-0.159
	(.0925)	(.0991)	(.1067)	(.0978)	(.1061)
% Cohort Female*Female	0.177 **	0.229 ***	0.224 **	0.229 **	0.215 *
	(.0853)	(.0879)	(.1002)	(.1025)	(.1179)
Obs	3,176	2,946	2,541	2,372	2,243

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

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: Robustness Checks - Drop High-Transfer Programs

	Drop Programs with % Transfer Students				
	>= 10%	>= 15%	>= 20%	>= 25%	>= 30%
	(1)	(2)	(3)	(4)	(5)
Female	-0.120 **	-0.116 **	-0.119 ***	-0.109 **	-0.122 ***
	(.0556)	(.0517)	(.0449)	(.0439)	(.0427)
% Cohort Female	-0.135	-0.102	-0.110	-0.109	-0.136
	(.1210)	(.1114)	(.1067)	(.1058)	(.1067)
% Cohort Female*Female	0.222 *	0.214 *	0.224 **	0.209 **	0.233 **
	(.1161)	(.1096)	(.1002)	(.1007)	(.0987)
Obs	2,106	2,362	2,541	2,623	2,718

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

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.