Gender, Peer Effects, and

College Major and Course Selection

by

Eilidh Geddes

(Under the Direction of David B. Mustard)

Abstract

What role do peer effects play in college math course taking and STEM persistence and major choice? This paper tests the hypothesis that peer effects affect female and male college students differently in their math course taking and major selection, widening the gender gap in math-based fields. I use the National Longitudinal Survey of Freshmen (NLSF) and employ two identification strategies. The first controls for baseline math skills as measured by high school achievement and freshmen course selection to examine how choices change as peer groups develop. A second strategy exploits exogenous variation in the gender composition of the university to examine how course taking and major choice depend on the university's gender composition. I find limited evidence that peer groups affect men and women differently with respect to the outcomes of interest. A higher proportion of female STEM majors increases the probability women are STEM majors and decreases the probability for men.

Index words: Economics of Education, Gender, Peer Effects

Gender, Peer Effects, and College Major and Course Selection

by

Eilidh Geddes

A Thesis Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Master of Arts

Athens, Georgia

2015

Gender, Peer Effects, and

College Major and Course Selection

by

Eilidh Geddes

Approved:

Major Professor: David B. Mustard

Committee: Christopher Cornwell
Meghan Skira

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
May 2015

Gender, Peer Effects, and College Major and Course Selection

Eilidh Geddes

April 27, 2015

Acknowledgments

I thank my thesis committee chair, Dr. David Mustard, for his help and support throughout the process of writing this thesis. I also thank the other members of my committee: Dr Christopher Cornwell and Dr. Meghan Skira. Dr. Cornwell initially helped recruit me to UGA and convinced me that UGA would be an excellent place to pursue my undergraduate education in economics.

Of course, I thank my parents and sister for all of their support over the years. Without them, there is not a doubt in my mind that I wouldn't be where I am today.

Given the topic of this thesis, it is appropriate to thank my peers who have influenced me in positive ways (as hopefully I have influenced them). John B Stroud has been a great source of support throughout. I further thank my roommates, Grace Siemietkowski and Alex Rowell, who now know far more about this topic than they ever wished to.

Much appreciation is also given to the UGA Honors Program and the Foundation Fellowship for their generous support, financial and otherwise, here at UGA. In particular, Jessica Hunt has been an invaluable resource as I have plotted my path through UGA and determined what I am doing after graduation.

Contents

List of Figures

List of Tables

Chapter 1

1.1 Introduction and Motivation

What role do peer effects play in college math course taking and STEM major and persistence? Do these effects differ for male and female students? In what ways? These questions play a role in testing one hypothesis of why there is a gender gap in STEM participation: social factors and peer groups push women away from math and math-based careers. Peers play a large role in determining educational outcomes in math; Fryer and Levitt (2010) find that the gender gaps in math achievement close substantially with single sex education. Furthermore, Hoxby (2000) shows that both male and female students in cohorts with a higher proportion of female students have higher achievement in math classes.

The difference in the number of men and women graduating from college with STEM degrees has important implications for the male-female wage gap; Rendall and Rendall (2014) find that a significant portion of income inequality is driven by the top percentiles of college graduates, many of whom are entering quantitative fields. Lack of representation in highly compensated quantitative fields is one factor driving the gender wage gap. This paper tests the hypothesis that while peer groups play an important role for both genders, peer effects affect female and male students differently in their math course taking and major selection in a college setting, widening the gap between the number of men and women who enter math-based fields.

To answer these questions, I use rich survey data from the National Longitudinal Survey of Freshman (NLSF) from students collected starting their freshman year of college that include information on major selection, course choice, social dynamics, and the institutional characteristics of the colleges that they attend. I combine this data with the data from the Integrated Postsecondary Education Data System (IPEDS) to construct measures of peer groups. I estimate a binary choice model of choosing to major or persist in a STEM field and a linear count model of math course taking controlling for institutional characteristics, individual characteristics, and previous selection into a math or science track. I examine both institutional and campus organization level peer groups and find limited evidence that these peer groups can influence male and female students in different ways. Being at a university with a higher percentage of STEM majors that are female increases the probability that female students choose STEM majors and decreases the probability that male students choose STEM majors. Regardless of gender, a higher percentage of female students, regardless of major, is correlated with taking more math courses.

This paper is organized as follows. I first discuss the relevant literature then describe a model that predicts possibly different behavior by male and female students. I then describe the dataset used in this paper and the econometric methods before providing results.

1.2 Relationship to Literature

This research expands the peer effects literature into a new area of interest: course and major selection. Research on peer effects at the university level examines how students form friendships and how roommates affect social and academic choices (Sacerdote and West 2013; Carrell et al. 2009; Sacerdote and Marmaros 2006). These studies use information from military academy squadrons, random roommate matching, and student email patterns. All three papers find that peer groups have substantial effects on academic performance. Sacerdote and Marmaros (2006) find that proximity in housing plays a large role in peer group formation with roommates having the largest effects. For instance, having a roommate in a fraternity greatly increases the chances that a student joins a fraternity.

Other studies examine how gender or racial composition of a cohort affects student achievement at a younger age (Hoxby 2000; Lavy and Schlosser 2011). These studies find that cohorts that randomly have a larger proportion of female students have higher levels of cooperation, lower levels of teacher fatigue, and better test scores in both math and reading. Lavy and Schlosser (2011) use administrative records collected by the Israeli Ministry of Education and find that the proportion of girls in a class has a significant and non-linear effect on the performance of both boys and girls in the class. Hoxby (2000) uses two empirical methods to study administrative data on elementary school students in Texas from the Texas Schools Microdata project. First, she uses the idiosyncratic differences in the size of racial and gender groups and examines their effect on the academic performance of the whole cohort. Second, she uses the idiosyncratic differences in the achievements of these groups and examines their effect on the whole cohort. Hoxby finds that both male and female students perform better in both math and reading groups that are more female. The exact mechanism that is at play is not determined by either of these two papers. Possible mechanisms are that female students act unenthusiastically about math in settings that are more male, that there is less classroom disruption in higher female-concentrated class, that there is more cooperation in female heavy classes, and that teacher pedagogy changes depending on the gender composition of the classroom.

Other papers on the subject of choice of STEM majors use the dataset that this paper uses. Griffith (2010) estimates a logit model predicting the probability that a student stays in or switches to a STEM major during their undergraduate career. Her model focuses on the impact of the institutional characteristics of the university that the student attends. She finds that while, unsurprisingly, the individual educational characteristics of entering students play a large role in persistence rates, schools that emphasize research have lower persistence rates and schools that have a higher proportion of female graduate students have higher persistence of female undergraduates in STEM majors. Additionally, Griffith (2008) examines how academic and social fit affect success and college major selection for low-income and minority students. She defines academic fit as the difference between a student's SAT score and the mean SAT score of the university that they are attending and social fit as the size of the peer group at the institution in terms of race or income status. She uses both the National Education Longitudinal Study (NELS) of 1988 and the NLSF in conjunction with IPEDS and the College Board's *Annual Survey of Colleges*. Griffith (2010) finds that academic fit matters for persistence and success, while social fit does not matter as much for these outcomes, but does influence college major selection.

This research expands this line of inquiry by analyzing the role that peers may play in determining course selection and by breaking down these peer effects by gender. Unlike Griffith (2010), I examine the impact of the proportion of female undergraduates instead of the proportion of female graduate students and examine social fit in terms of gender composition instead of race or income status. In addition to examining persistence in a major and major selection, I extend this literature to examine individual course selection as well. I exploit variation in gender composition of the cohort in the vein of Hoxby (2000) and Lavy and Schlosser (2011).

1.3 Economic Model

The theoretical prediction that this paper tests is that peer groups influence male and female students differently with regards to their course selection and major choice. The basic theory is that students receive utility in college from both returns from their time spent on academic endeavors and from their time that they spend engaging in social activities. Returns from their time spent on coursework can include the expected return of labor market outcomes post-graduation, benefits derived from the act of learning the subject, and social benefits gained from the students met through the coursework. There are numerous interactions between these two sets of activities as students may socialize with students that they meet through their coursework or study with students met through social activities.

These returns and interactions may differ by gender. For instance, majoring in a male-dominated field or taking a male-dominated class may benefit the social life of a male student and hurt the social life of a female student. The financial returns may vary based on a student's preparedness and thus their success in their major; thus, expectations of success in a given major will be one factor that determines initial freshman choices regarding certain classes or majors. Additionally, the time required to invest to reach a given return from a major may vary based on a student's innate characteristics, whether they be academic or not. For instance, a belief that there exists substantial discrimination in a given field may discourage students from entering that field if they believe that they cannot fully access the returns on their investment of time and money. Furthermore, if students enter college with different levels of preparedness, the level of time commitment necessary for academic success will also differ. This paper tests whether these interactions differ by gender.

1.4 Empirical Model

In this paper, I consider three outcomes: being a STEM major unconditional on being a STEM major in the past, being a STEM major conditional on being an intended STEM major as a freshman (persistence), and the number of math courses taken. I consider two sets of peer groups: macro-level and micro-level peer groups. Macro-level peer groups are the percent of students that are female at the university and the percent of STEM majors who are female at the university. These peer groups vary across schools and time. Micro-level peer groups are measured by campus activities that the student is involved in and the number of female friends that the student has. These peer groups vary across students and schools, but not time due to data constraints which are discussed later.

I estimate a binary choice model given by:

$$Pr[Outcome_{ist} = 1] = \alpha + \beta PeerGroups_{ist} + \theta PeerGroups_{ist} * Female_i + \gamma X_{ist} + \xi G_{st} + \tau_t + \epsilon_{ist} \tag{1.1}$$

where i indexes the individual, s indexes the university i attends, and t indexes the academic year. The outcome of interest is persisting in a STEM major conditional on being a STEM major in $t-1$ or choosing a STEM major unconditional on being a STEM major in the past, X_{ist} is a vector of individual characteristics in school s at time t, G_s is a vector of university characteristics not related to the defined peer group such as whether the university is public or private or if the university is a liberal arts or research institution, τ_t are year fixed effects, and ϵ_{ist} is a mean zero error term. In some specifications, G_s includes fixed effects for each institution in the sample. $PeerGroups_{ist} * Female_i$ is an interaction term between the peer groups and gender. Depending on the specification, $PeerGroups_{ist}$ will either be a vector representing micro-level peer groups or macro-level peer groups as described above.

I also estimate a linear model of the number of courses taken in the math department in an academic year given by:

$$NM_{ist} = \alpha + \beta PeerGroups_{ist} + \theta PeerGroups_{ist} * Female_i + \gamma X_{ist} + \xi G_{st} + \tau_t + \epsilon_{ist} \qquad (1.2)$$

where NM_{ist} denotes the numbers of math courses taken by student i in school s at time t and peer groups are measured either by the percentage of the university that is female and the percentage of students that are taking math classes that are female or by participation in extracurricular activities. $PeerGroups_{ist} * Female_i$ is again an interaction term between the peer groups and gender. I construct various measures of $PeerGroups_{ist}$ as described above. Here, major choice is included in the vector of individual characteristics, X_{ist}. In both estimating equations, θ is the coefficient of interest. While the effects of peer groups are important to determine, this paper's ultimate goal is to determine if they differ by gender and in what ways. The linear combination of the relevant coefficients will determine the full effect of gender and peers.

1.5 Data Description

I use the NLSF, which follows a cohort of freshman throughout their college years from 1999 to 2003. The survey provides comprehensive information to test possible explanations for minority underachievement in college and includes information about academic and social factors that may determine collegiate success. The data were collected in six waves starting in 1999 to follow students throughout their undergraduate degrees and gather preliminary information on their immediate outcomes post-graduation. This dataset includes retrospective information about students' high school characteristics, high school peer groups, and family life, as well as information on course choices in college, time allocation on a weekly basis, campus involvement, grades, friends, and majors. It samples freshmen at twenty-eight elite institutions[1] with a variety of characteristics. The institutions were chosen to mirror those in the College and Beyond Survey with the addition of UC Berkeley due to its abandonment of its affirmative action policies and of historical black college and universities (HBCU). Thirty-five institutions were initially asked, but five schools[2] declined outright and Morehouse and Spelmen could not provide a list of freshmen from which to draw a sample. Three of the institutions are women's colleges (Barnard, Smith, and Byrn Mawr) and one is a HBCU. The survey had a high response rate of 86% at the first time the survey was conducted. I supplement the institutional characteristics with data drawn from the 1999-2003 waves of the Integrated Postsecondary Education Data System (IPEDS) about the gender composition of institutions and major composition within institutions.

There were 4,573 students approached to complete the survey, and, of these, 3,924 completed the survey. I exclude observations that are missing information about standardized test scores as well as observations that do not continue at the same school throughout the sample. With these restrictions, my sample consists of 2,860 students. Of these, 1,656 are female and 1,205 are male. The sample is 26.6% African American, 23.2% Hispanic, and 24.6% Asian.

For the students in the NLSF sample, Table 1.1 provides information about the universities they attend. It lists basic descriptive statistics of the institutions in the first wave of the sample (1999). Generally, students in the sample in 1999 attend institutions

[1]The following institutions were sampled: Howard University, University of Michigan, University of North Carolina, University of California Berkeley, Columbia University, Emory University, Miami University, Northwestern University, Penn State University, Stanford University, Tulane University, University of Pennsylvania, Georgetown University, Oberlin College, Princeton University, Rice University, Tufts University, University of Notre Dame, Washington University (St. Louis), Wesleyan University, Williams College, Yale University, Barnard College, Bryn Mawr College, Denison University, Kenyon College, Smith College, Swarthmore College

[2]Duke, Vanderbilt, Wellesley, Hamilton, and Xavier declined to participate in the survey.

that are more female than male and are predominately white. The racial composition of one institution, Columbia, is missing for 1999. Here, the unit of observation is the student.

Table 1.2 lists basic descriptive statistics for the students in the sample. They are reported unweighted and should be interpreted as descriptive of the sample rather than an estimate of the population characteristics. Using the concordance table reported by the ACT and College Board, SAT scores and ACT scores were imputed when missing if the other test was reported. While ideally these scores could be broken down into math and verbal, this information is only available for students who took the SAT score, as all the ACT reports is the composite score. There is the possibility of measurement error for these test scores since they are self-reported.

I include some information about courses taken in high school. I chose high school calculus and physics as they are generally two of the most advanced and mathematically rigorous courses offered in the math and science departments of high schools and indicate selection into a mathematically challenging track at the high school level. Additionally, all students in the sample report the number of years that they took the subject, whereas recall was less uniform for other subfields of math or science. Over 70% of the sample took these classes in high school, which is unsurprising given the selective nature of the institutions in the sample.

Table 1.1: Institutional Summary Statistics in 1999 by Student Observations

Variable	Observations	Mean	Std Dev	Min	Max
Total Undergraduate Enrollment	2,860	11,921.35	9,062.69	1,316	34,505
Percent Female	2,860	53.9%	9.8%	45.45%	100%
Percent White	2,688	69.16%	16.08%	0%	92%
Percent African American	2,688	8.08%	10.65%	3%	89%
Percent Hispanic	2,688	5.04%	3.38%	0%	13%
Percent Asian	2,688	13.12%	9.76%	1%	39%

Table 1.2: Individual Summary Statistics

Variable	Mean	Std Dev	Min	Max
Female	0.578	0.494	0	1
SAT Score	1306.562	156.51	600	1600
ACT Score	29.068	3.935	12	36
HS Calculus	0.733	0.442	0	1
HS Physics	0.701	0.458	0	1
Freshman STEM Major	0.209	0.406	0	1

N=2,860

Approximately 20% of students in the sample were STEM majors during their freshman year. Table 1.3 reports the percentages of students in STEM majors throughout their college career to determine whether students are generally moving into or out of STEM majors (Table 1.3). Griffith (2010) predicts that the percentage should decrease over time as students attrit from STEM majors.

There is a spike in percentage of students majoring in STEM at the end of their second year in college. This spike reflects an increase in the number of people who have declared a major, since this is the time in college at which point most institutions require students to declare a major. In the first year of college, 63.2% of students in the sample are listed as undeclared. In their second year, this percentage drops to 25.8% and further drops to 0.6% in their third year, supporting the idea that this increase in STEM majors is due to an increase in major declaration rather than a switch towards STEM majors away from non-STEM majors. However, the main takeaway from this table is that, at every point in time, a substantially higher proportion of men choose a STEM major.

One possible problem with only looking at those students who have a declared major is that there may be many students who are planning on eventually declaring a STEM major but who are uncertain about what STEM field in which to major. I do not

Table 1.3: STEM Major By Gender

Variable	STEM Major	Intended STEM Major
Women		
Entering Major	19.5%	
End of Year 1	13.1%	34.7%
End of Year 2	21.6%	29.3%
End of Year 3	24.2%	24.2%
End of Year 4	18.6%	18.6%
N= 1,655		
Men		
Entering Major	22.7%	
End of Year 1	18.6%	43.7%
End of Year 2	30.9%	39.4%
End of Year 3	30.9%	31.7%
End of Year 4	26.2%	26.2%
N= 1,205		

label these students in the data as STEM majors because I am principally interested in students who have a firm and strong declared interest in a STEM field. These students are likely to be different than undecided students with less clear interests. Unfortunately, I cannot tell the difference in the data with any known level of accuracy between a student who is truly undeclared and those who do not know which STEM field to enter. I include the list of STEM majors in Appendix A.

Because there are likely many students who are undeclared in their first year who intend to choose a STEM major, I construct a new measure of being a STEM major to analyze persistence in a STEM major. I consider students an intended STEM major if they either are a STEM major or are taking more than three classes in a STEM field and are undeclared. This measure is missing for their entering semester since at this point, there have been no courses taken yet. With this definition of an intended STEM major, there is a downward trend in the percentage of students in STEM majors throughout college as seen in Column 2 of Table 1.3

Table 1.4: STEM Persistence By Gender

Year	Intended STEM Persistence	
	Women	Men
Second Year	60.75%	69.39%
Third Year	46.54%	52.83%
Fourth Year	36.77%	39.96%

Table 1.4 shows persistence in a STEM major or in an intended STEM major over time. A student is considered to have persisted if he or she was classified as an intended STEM major (STEM major or taking more than three classes in a STEM field while undeclared) in time $t-1$ and is classified as an intended STEM major in time t. I only consider students who were initially classified as intended STEM majors in the first time period when calculating persistence rates; therefore, the denominator for the calculations is the same across time periods. As expected and following the trends in STEM major choice, persistence falls over time. There is also a gender differential in persistence: over time, men persist more in STEM fields in addition to choosing them in higher numbers to begin with. These STEM persistence numbers are slightly lower than existing estimates in the literature; Bettinger (2010) finds that just under half of STEM majors persist to graduation, although this measure of STEM persistence is calculated only on a sample of students who declare a STEM major when they begin college. However, the gender differential between persistence rates matches the pattern found in the literature.

Eccles (2006) find that there may be an even greater gap in specific subsets of STEM majors as there are more women than men in the biological sciences. I create a dummy for majoring in the biological sciences and for majoring in engineering (listed in Appendix A) where biological sciences also includess health-related majors and bio-engineering majors are double-listed. I compare the gender breakdown for biological sciences to that of other STEM majors where research suggests there may be a larger gender gap. The gender breakdown for subcategories of STEM majors is shown in Table 1.5. The numbers shown in the first two rows are the percentage of students of that gender who are in each major. The remainder of the sample are non-STEM majors. The second two rows show, conditional on being one of these majors, what the gender composition is.

Table 1.5: Gender Breakdown of Subcategories of STEM majors

Gender	Biological Sciences	Engineering	Computer Science	Mathematics
In Sample				
Male	7.84%	7.72%	6.18%	1.41%
Female	10.29%	4.06%	1.56%	0.82%
In Major				
Male	35.79%	58.03%	74.31%	55.74%
Female	64.31%	41.97%	25.69%	44.26%

The differences in gender composition across majors in the STEM field suggest that there are possibly preferences across the types of work that are involved in the majors. One possible explanation is that female students avoid fields that are more quantitative. Because of this, the number of math courses taken is an outcome of interest in this paper. Another possible explanation for the larger gender gaps in engineering is that there is more support given to male students from parents, teachers, and peers in entering those fields than to female students (Eccles 2006).

Additionally, I examine the number of freshmen students taking math courses in their first year (either semester). Table 1.6 shows what we would expect from the literature regarding the number of math courses being taken by male and female students. Male students take on average more math classes in their first year, taking on average 0.573 classes to women's 0.484. Additionally, the proportion of male students who are taking a

Table 1.6: Freshman Math Course Taking By Gender

Variable	Observations	Mean	Std Dev	Min	Max
Both Genders					
Number of courses	2,860	0.514	0.532	0	3
Percent taking at least one class	2,862	0.498	0.500	0	1
Women					
Number of courses	1,655	0.484	0.526	0	2
Percent taking at least one class	1,655	0.470	0.499	0	1
Men					
Number of courses	1,205	0.573	0.548	0	3
Percent taking at least one class	1,205	0.551	0.497	0	1

math course in their first year is 17.02% higher than the proportion of women doing so. It is important to examine course choices in the first year of college since there is evidence that courses taken first year are highly important in determining eventual major choice (Bettinger 2010).

1.6 Research Method

I use two identification strategies to estimate my effect of interest. The first strategy exploits exogenous variation in gender composition in colleges across years and determines if math course taking and major selection depends on the proportion of students who are female or the proportion of female students in STEM majors. The group of students in STEM majors proxies for the students' peer groups. There are two macro-level peer groups defined in this strategy. The first is the university wide peer group, measured by overall percentage of the class that is female. The second is STEM major peer groups, measured by the percentage of STEM majors who are female.

Over the time period that the survey was conducted, the percent of female students at these institutions, calculated using IPEDS, fluctuated. When the percent female across all institutions is averaged, the percentage of female students fluctuated between 53.85% and 54.10% (see Figure 1.1). However, there was considerably greater variation over time within some institutions. At the high end, there was as much as a 3.93 percentage point change in gender composition of a university between 1999 and 2003. This small variation is plausibly exogenous with respect to selection into a college given that this is a relatively short time period. It is unlikely that students were accurately able to predict while selecting their college how the gender composition of the university as a whole or of STEM fields would change during their college career.

Figure 1.1: Percent of Female Students

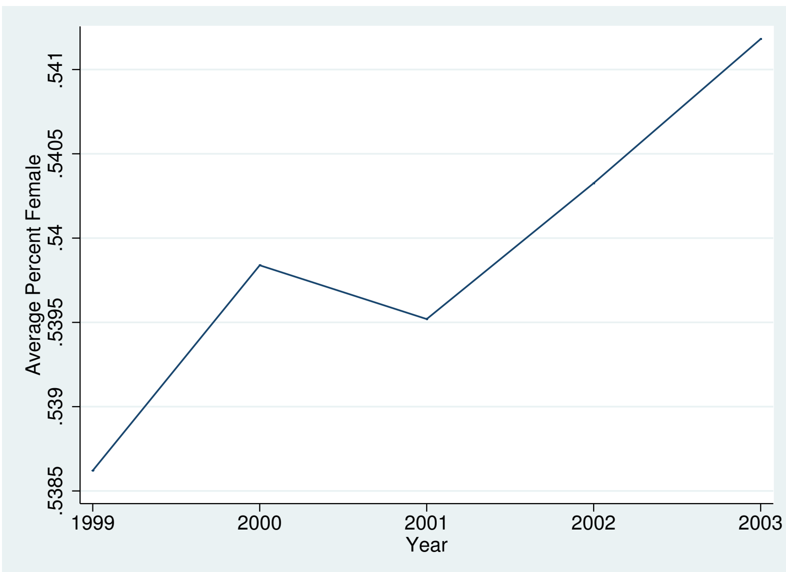
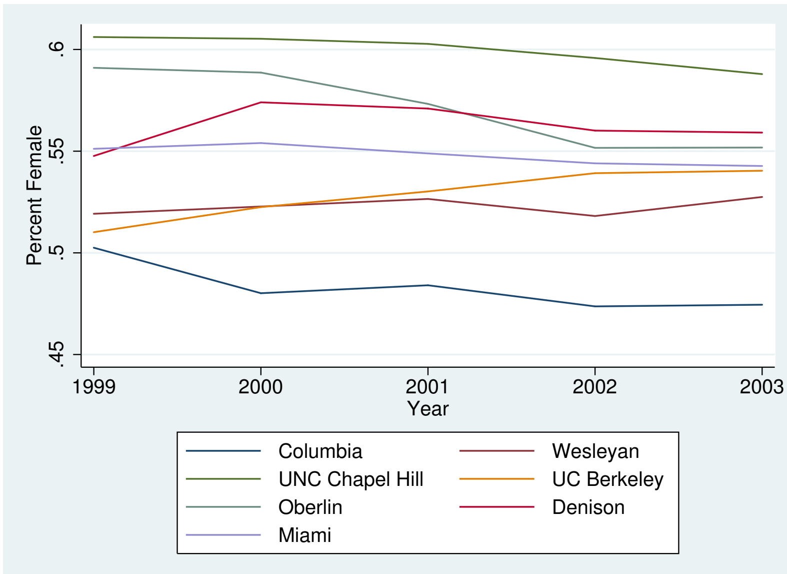


Figure 1.2 shows the variation in gender composition over the time of the sample for seven of the institutions sampled. While some of the institutions show relatively little variation (i.e. women's colleges remain constant at 100% female over time), other

Figure 1.2: Variation within Schools Over Time



institutions experience considerably more variation, which allows me to identify my effect of interest.

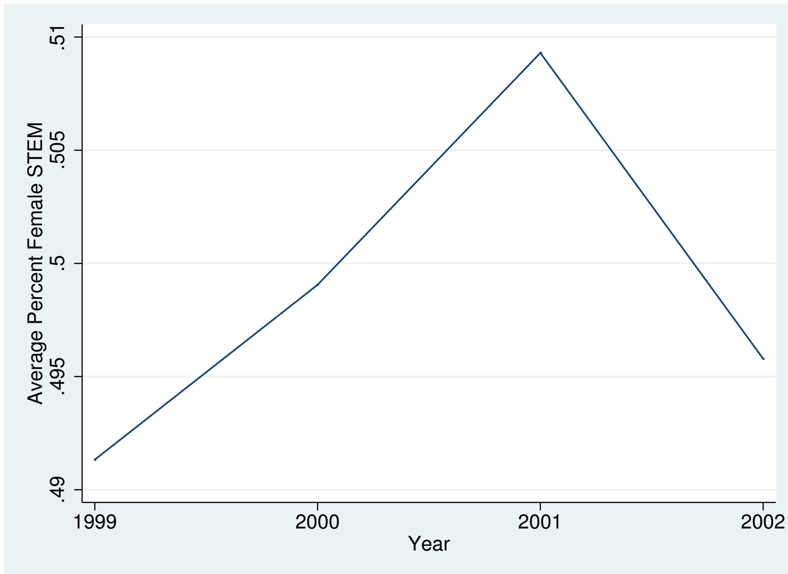
Table 1.7 shows the breakdown in where the variation in gender composition comes from over time. While there is considerable variation between institutions, there is also significant variation within institutions, which I exploit in this paper.

Table 1.7: Variation in Percent Female

Variable	Mean	Std Dev	Min	Max
Percent Female Overall	0.539	0.098	0.454	1.000
Between		0.097	0.463	1.000
Within		0.007	0.519	0.563

Figure 1.3 shows how the percentage of STEM majors who are female changes on average over the time of the sample. The percentage of STEM majors who are female is

Figure 1.3: Variation in Percent Female in STEM



calculated from the NLSF. This additional variation helps identify the model. The assumption made with the computation of the percentage of STEM majors who are female at a school is that when constructed with probability weights from the NLSF, the sample is representative of gender composition in majors at the university.

To check how potentially problematic this assumption is, I compare the NLSF sample constructed measures of the percentage of STEM majors who are female with data from IPEDS. Using IPEDS, I construct another measure of the percentage of STEM majors who are female by summing over the students in engineering, the biological sciences, the physical sciences, and mathematics. Unfortunately, these data are only available for even-numbered years and only for 18 out of 28 institutions[3]. Even within these 18 institutions, the data are not available every year this information was collected

[3]These data are available for Columbia, Howard, Miami, Penn State, Princeton, Rice, Smith, Stanford, Swarthmore, Tufts, Tulane, Berkeley, Michigan, Notre Dame, Washington University, and Yale.

for all of the institutions. Based on a comparison with these data, my constructed measure tends to over-state the percent of STEM majors who are female by on average 8.9 percentage points. I conduct further robustness checks of this assumption by reducing my sample to only student-year observations where I have this institution-level data and rerunning my specifications using the institution-level data from IPEDS.

The second strategy is to control for baseline math skills as determined by high school math achievement, freshmen course selection, and selection into a STEM major to examine how choices change over time as peer groups develop in the university setting. I assess students' baseline characteristics, grades, math course choices, major choice, and peer groups, and then estimate a binary choice model of remaining in a STEM major given intended major was STEM as a freshman, a binary choice model of being a STEM major, and a linear model of number of math courses taken given earlier selection into a STEM major. I include interaction terms between gender and various peer groups at both the macro and micro levels to determine the differential effect that these groups play on male and female students.

The main threat to identification is endogeneity, specifically selection bias. The students who select into a given set of classes, university, or set of activities are likely systematically different than those who do not select into those groups. Because of this, it is difficult to separate the effect of being in the peer group from the effect of being the kind of person who would join that group. Additionally, there may be omitted variable bias as I cannot accurately measure scientific or mathematical ability which may not be fully revealed by success in high school. This ability may be correlated both with the outcomes of interest and selecting in to certain social groups or universities. I address the issue of ability by including high school course selection and test scores, but these measures do not fully reflect the variability in ability or academic preparation, particularly given the selective nature of the schools in my sample. Typically, students have high test scores and took the most rigorous track of classes in high school, including math and science classes.

The assumption that is needed for identification in the first strategy is that with the exception of women's colleges, students do not select into specific institutions based on small variations in the gender composition of the institution, but may select into majors within the institution based on gender composition. Selection into majors based on gender compositions would be the peer effect of interest. In the second strategy, the assumptions are that participation in different social groups is random within the group of students who have already selected in to a STEM major and that men and women select into social groups in similar ways. Then the variation in social groups can be used to explain varying degrees of persistence in a STEM major.

Of these two strategies, identification in the first strategy is more plausible. The variation in gender composition is more clearly defined and more likely to be exogenous than the variation that exists in students' choices into various extracurricular organizations; therefore, the assumptions under which the selection bias will be addressed are more plausible. Another way selection can be addressed is by controlling for unobservable differences in institutions that remain constant over time by using fixed effects, taking advantage of being able to see many students at the same school over time. Controlling for these fixed effects will control for most of the selection into a particular institution based on time-invariant characteristics.

1.7 Results

I first present my results for the effect of macro-level peer groups on three outcomes of interest: selection of a STEM major, persisting in a STEM major if in intended STEM major freshman year and in $t-1$, and number of math courses taken. I then present my results for the effects of micro-level peer groups on the same three outcomes.

1.7.1 Macro Peer Groups

STEM Major Choice

I first examine the effects that large, institution level peer groups have on the decision to select a STEM major. I estimate a binary choice model with the outcome of interest being a declared STEM major (list of STEM majors included in Appendix A). I run both probit and logit models and find that probit is a better fit (psuedo-R^2 = 0.1536 versus pseudo-R^2 = 0.1530) , although the results are robust to the choice of model. I thus make the assumption that the error terms are normally distributed. Table 1.8 reports the marginal effects from the probit model. The coefficient estimates are shown in Table 1.14 in Appendix B. Specification 1 includes basic characteristics of the individual student as well as year fixed effects. Specification 2 adds institution fixed effects. Specification 3 includes information about the type of institution (public research, private research, or liberal arts, with public research being the baseline) and the gender composition of the institution, but does not include school level fixed effects in order to allow me to estimate the effects of being one of these various kinds of institutions. In specifications 1 and 3, there is variation both across and within institutions over time. Specification 4 has institution fixed effects as well as variables for gender composition of both the school and STEM majors that vary over time. Specification 5 includes the interaction terms. In specifications 2, 4, and 5, identification comes from the within school variation in overall gender composition and STEM major gender composition over time. Through specifications 1-5, standard errors are clustered at the individual-student level.

As expected, in all specifications except specification 5, the coefficient on female is negative and statically significant at the 5% level. The predicted probability that female students are STEM majors is 3-4 percentage points less than male students across specifications. This effect persists even when controlling for other academic characteristics and institution type. In specifications 1-4, this effect can be interpreted as the main effect of being female. In specification 5, I estimate the effect of being female at the average percent female and the average percent female STEM and find that it is -3 percentage points(standard error 0.020) and -2 percentage points (standard error of 0.018), respectively. The results for race variables are as predicted by the literature, with Asian students being more likely to major in a STEM field and black and Hispanic students being less likely to major in a STEM field relative to white students.

Ultimately, the variables of interest are the interactions between the peer groups and gender. In the case of macro peer groups, the interaction of interest is the interaction between the percentage of students at the institution who are female and being female and the percentage of STEM majors who are female and being female. The assumption made with the computation of the percentage of STEM majors who are female at a school is that when constructed with probability weights, the sample is representative of gender composition in majors at the university.

In specification 5, the coefficient of the percentage of STEM majors who are female is not statistically significant, but the interaction term is positive and statistically significant. This result supports the hypothesis that peer groups have different effects on male and female students. For female students, the marginal effect of the percentage of STEM majors who are female for men is -0.262 and is statistically significant at the 1% level (standard error 0.101) while the marginal effect for women is 0.256 and is statistically significant at the 1% level (standard error 0.093). The interpretation of this is that a 1 percentage point increase in the percentage of STEM majors who are female leads to a

0.262 percentage point decrease in the probability of being a STEM major for men and a 0.256 percentage point increase in the probability of being a STEM major for women. Since these marginal effects are both statistically different from zero in opposite directions, I can conclude that the marginal effects are different for men and women. Interestingly, the effects were symmetric in opposite directions.

When year effects are included, students are 12.9% more likely to be a STEM major in year 2, 13.3% more likely to be a STEM major in year 3, and 8.6% more likely to be a STEM major in year 4, relative to year 1. All year effects were statistically significant at the 1% significance level.

I include only one of ACT or SAT scores since the two variables, particularly in their imputed form, are so highly correlated with one another. SAT is more significant both statistically and economically in the regressions with number of math courses a the outcomes and is relatively close to ACT in terms of both statistical and economic significance in the model with choice of a STEM major as an outcome. Additionally, I have substantially more reported SAT scores relative to ACT scores; thus, I include SAT over ACT since there are fewer imputed values. Unfortunately, there is some loss of precision when converting ACT scores to SAT scores as each ACT score corresponds to a range of SAT scores. I assign the middle of the SAT range to the corresponding ACT score, with the exception of a 36, which exactly corresponds to an SAT score of 1600.

In specification 2, while it is unsurprising that there are fewer STEM majors at liberal arts institutions, it is interesting that there is such a large, negative, and statistically significant relationship between studying at a private research institution and the probability of selecting a STEM majors, relative to studying at a public research institution. This result may be because of selection into institutions and is absorbed when college-specific fixed effects are included.

Including institutional fixed effects controls for a large amount of the selection into a college, making the gender variation at the university plausibly exogenous. Additionally, including these fixed effects controls for the differing curricula at different universities. My preferred specification is specification 5, which includes institution fixed effects but allows the gender composition both of the institution and of STEM majors at the institution to vary over time. I also estimated the effect of these peer group variables without including the institution fixed effects and found that percent female was significant. However, without including the institutional fixed effects, gender composition is largely endogenous as students likely select their college taking into consideration the gender composition.

As a robustness check, I drop all students who are attending either of the two women's colleges that are included in the dataset since there likely is selection into an all female environment. The coefficients (not reported) remain largely the same, indicating that any selection into a women's college is absorbed by the institution level fixed effects that are included.

Table 1.8: Effects of Macro-Level Peer Groups on STEM Major Choice: Average Marginal Effects from a Probit Estimation

	Specification				
Variables	(1)	(2)	(3)	(4)	(5)
Female	-0.037**	-0.037**	-0.041**	-0.036**	-0.025
	(0.018)	(0.018)	(0.018)	(0.018)	(0.017)
Percent Female			0.025	0.681	0.0471
			(0.129)	(0.774)	(0.779)
For Men			0.026	0.715	0.371
			(0.137)	(0.811)	(0.825)
For Women			0.024	0.655	0.540
			(0.124)	(0.745)	(0.720)
STEM Percent Female			0.045	-0.017	0.012
			(0.045)	(0.072)	(0.744)
For Men			0.048	-0.177	-0.262***
			(0.074)	(0.764)	(0.101)
For Women			0.043	-0.162	0.256***
			(0.067)	(0.070)	(0.093)
SAT Score	0.0000	0.0001*	0.0001*	0.0001*	0.0001*
	(0.0000)	(0.0007)	(0.0001)	(0.0001)	(0.0001)
Freshman STEM Major	0.379***	0.362***	0.367***	0.363***	0.356***
	(0.027)	(0.028)	(0.028)	(0.028)	(0.028)
Liberal Arts			-0.046		
			(0.031)		
Private Research			-0.049**		
			(0.021)		
Includes race dummies	X	X	X	X	X
Year fixed effects	X	X	X	X	X
Institution fixed effects		X		X	X
Includes high school courses	X	X	X	X	X
Interaction terms included					X
N	11,440	11,440	11,364	11,364	11,364
$Pseudo-R^2$	0.119	0.143	0.120	0.139	0.146

Note: Standard errors in parenthesis.
Standard errors are clustered at the student level.
* indicates statistical significance at the 10% level.
** indicates statistical significance at the 5% level.
*** indicates statistical significance at the 1% level.

I additionally estimate models with fixed effects and random effects at the individual level. The coefficient estimates from the fixed effects specification using a logit model are shown in Table 1.14 in Appendix B, along with the coefficients from the probit model that I estimate above. Here, both the coefficient on the percent female STEM and the interaction term are statistically significant. The sample size becomes much smaller because of the number of student-observations with all positive or all negative outcomes. The results are qualitatively the same in a random effects model.

I run the same regressions on two further restricted samples: the sample excluding women's colleges and the sample for which I have data on the number of female STEM majors at an institution level from IPEDS. My results when I exclude women's colleges are not qualitatively different from my results with women's colleges included, supporting the idea that selection into women's colleges is not driving my results. I do not find statistical significance at the same levels when I use the institution level STEM gender data instead of my NLSF sample-constructed measure for any specifications; however, my sample size is substantially smaller at 4,119 for the non-fixed effects specifications and 792 for the fixed effects specification. The difference in the results means that the assumption that the sample is representative of gender and major choice may be problematic and could be one factor driving the results.

STEM Major Persistence

I conduct similar analysis on the probability of persisting in a STEM major. Average marginal effects are shown in Table 1.9. Coefficient estimates including the estimates from the model with individual level fixed effects are shown in Table 1.15 in Appendix B. The specifications proceed as for the model of STEM major choice. Here, I only find effects in a individual level fixed effects model. They are qualitatively very similar to the results from the fixed effects specification for STEM major choice with a positive coefficient on the interaction term between the percentage of STEM majors who are female and being female and a negative coefficient on the percentage of STEM majors who are female. In specifications 1-5, SAT score is positively related with STEM persistence. There is a different pattern with race (marginal effects and coefficients not reported) from STEM choice. Hispanic students are statistically significantly less likely to persist in STEM majors, but the marginal effects for black and Asian students are not significantly different from zero.

Table 1.9: Effects of Macro-Level Peer Groups on STEM Persistence: Average Marginal Effects from a Probit Estimation

	Specification				
Variables	(1)	(2)	(3)	(4)	(5)
Female	-0.019	-0.042	-0.045	-0.042	-0.019
	(0.044)	(0.044)	(0.044)	(0.044)	(0.046)
Percent Female			0.155	0.563	0.293
			(0.336)	(1.804)	(1.819)
For Men			0.154	.559	0.118
			(0.334)	(1.795)	(1.924)
For Women			0.155	0.566	0.469
			(0.338)	(1.815)	(1.938)
STEM Percent Female			0.219	-0.127	-0.114
			(0.219)	(0.213)	(0.216)
For Men			0.218	-0.127	-0.449
			(0.218)	(0.212)	(0.292)
For Women			0.220	-0.128	0.222
			(0.220)	(0.214)	(0.309)
SAT Score	0.0002	0.0003*	0.0003**	0.0003*	0.0003*
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Liberal Arts			-0.011		
			(0.088)		
Private Research			-0.061		
			(0.046)		
Includes race dummies	X	X	X	X	X
Year fixed effects	X	X	X	X	X
Institution fixed effects		X		X	X
Includes high school courses	X	X	X	X	X
N	3,303	3,303	3,303	3,303	3,303
$Pseudo-R^2$	0.035	0.068	0.046	0.068	0.073

Note: Standard errors in parenthesis.
Standard errors are clustered at the student level.
* indicates statistical significance at the 10% level.
** indicates statistical significance at the 5% level.
*** indicates statistical significance at the 1% level.

Number of Math Courses Taken

In addition to looking at the relationship between choice of and persistence in a STEM major and institutional characteristics, I examine the relationship between institution level peer groups and number of math courses taken by the student. I estimate a linear model of the number of math courses taken by a student in an academic year. The general pattern of results for taking math courses are similar to those from the specification for choice of STEM major. As before, Asian students take more math classes and Hispanic and black students take fewer math classes relative to white students. Generally throughout the specifications, the coefficient on female is negative and statistically significant.

When I include fixed effects for the institution so that the percent female is capturing the variation across years in the gender composition of the university, the coefficient on percent female is negative and statistically significant. When the percentage of STEM majors who are female was included, I find that neither that variable or the interaction with the gender dummy are statistically significant. From this model, I find that the gender composition of the university matters for math course taking, but that the effects don't vary by gender.

I run this model on the restricted sample of only non-women's colleges and find estimates that are very similar, indicating that the results are robust to the inclusion of women's colleges. The results are not being driven by selection of students into women's colleges versus co-ed institutions.

Table 1.10: Effects of Macro-Level Peer Groups on Math Course Taking: Linear Estimation

	Specification					
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.158***	-0.149***	-0.143***	-0.149**	-0.637	
	(0.032)	(0.032)	(0.032)	(0.319)	(0.569)	
Percent Female			-0.149	-3.354*	-3.871**	-2.139
			(0.207)	(1.862)	(1.938)	(1.642)
Percent Female * Female					1.009	0.092
					(1.235)	(2.167)
STEM Percent Female			-0.065	-0.153	-0.115	-0.116
			(0.115)	(0.195)	(0.241)	(0.146)
STEM Percent Female * Female					-0.035	-0.219
					(0.210)	(0.196)
SAT Score	0.000	0.0002	0.0003	0.0003	0.0002	
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	
Freshman STEM Major	0.312***	0.264***	0.289***	0.263***	0.262***	
	(0.049)	(0.051)	(0.051)	(0.051)	(0.051)	
Liberal Arts			-0.139**			
			(0.056)			
Private Research			-0.106**			
			(0.043)			
Includes race dummies	X	X	X	X	X	X
Year fixed effects	X	X	X	X	X	X
Institution fixed effects	X	X	X	X	X	X
Includes high school courses	X	X	X	X	X	X
Individual fixed effects						X
N	11,440	11,440	11,364	11,364	11,364	11,364
R^2	0.201	0.226	0.211	0.229	0.217	0.156

Note: Standard errors in parenthesis.
Standard errors are clustered at the student level.
* indicates statistical significance at the 10% level.
** indicates statistical significance at the 5% level.
*** indicates statistical significance at the 1% level.

There is a decrease in sample size across all outcomes I consider when the variable for the percent of STEM majors that are female is added because there are no declared STEM majors in the sample in 1999 for two of the institutions.

Test of Linear Combination of Variables

In my preferred specification (specification 5), I test various linear combinations of variables and coefficients to determine various effects. I first examine the total effect of being female (Female+Percent Female*Female x Average Percent Female + Percent Female STEM*Female x Average Percent Female STEM). The average percent female in the sample is 53.96%. The average percent of STEM majors who are female is 48.89%. I find that the total effect of being female is -0.132 and is significant at the 0.01% level (standard error 0.034). This can be interpreted to mean that female students take on average 0.132 fewer math course.

I now test the full effect of Percent Female and Percent Female STEM on women at the average of these two measures. The full effect of Percent Female is negative, large (-4.963) and statistically significant at the 1% level. As the percentage of students who are female at the institution increases by 1 percentage point, the number of math courses taken by a female student decrease by 0.04. The full effect of Percent Female STEM is -0.538 but is not statistically significant (standard error 0.581)

I finally test if these effects are different for men and women by testing the linear combination of being female and the interaction terms at the average level. Neither of these is statistically different from zero which means I cannot reject that the effects are the same for male and female students.

I also examine the linear combinations for the fixed effects model (specification 6). While the individual coefficient estimates are not statistically significant, I find that the linear combination of STEM percent female and its gender interaction term is -0.335 and is statistically significant at the 5% level (standard error 0.131). Additionally, the linear combination of percent female and its interaction term evaluated at its average level is -2.46 with standard error of 1.063 and is also significant at the 5% level.

1.7.2 Micro Peer Groups

I additionally define peer groups as the number of female friends that a student reports having and choice of campus activities. I again estimate a binary choice model of being a STEM major and persisting in a STEM major using these peer groups as well as a linear model of the number of math courses taken. Average marginal effects are reported for major choice in Table 1.11 and for STEM persistence in Table 1.12. Coefficient estimates for these two outcomes are reported in Tables 1.16 and 1.17 in Appendix B. Coefficient estimates for math course taking are reported in Table 1.13. For each of the three outcomes, I report two specifications. The first includes only the peer groups in addition to a vector of controls while the second also includes the interaction terms between being female and the various peer groups. In the first specification, identification rests on the assumption that there are no unobservable characteristics driving selection into peer groups, which is highly implausible. In specification 2, identification of the interaction terms rests on the assumption that there are no unobservable characteristics driving differential selection of male and female students into peer groups based on gender, which is more plausible.

I include high school and race characteristics, though impacts of those variables are not reported. High school characteristics consist of SAT score and whether the student took high school calculus or physics. For determining whether a student is a STEM major, the relationship with these high school characteristics is small and statistically insignificant. In these specifications, observations from the first year of college are dropped because friends and campus involvement are not asked until Wave 3 which corresponds to the 2nd year of college. The coefficients on these variables should be interpreted as the effect of friends or involvement in the 2nd year of college on course selection and major choice in subsequent years. Ideally, I would have data on all peers for all waves, which would strengthen this identification strategy. Throughout these specifications, I include all schools in the sample, including women's colleges. As before, the standard errors in the linear model are clustered at the student level in both the binary choice and linear models.

STEM Major Choice

The marginal effects reported in specification 1 are the total marginal effect for that peer group, regardless of gender. In specification 2, I break it down into three marginal effects: the overall effect, the effect on women, and the effect of women. Interaction terms are included in specification 2, but not in specification 1. The marginal effects of many of the campus life activities are what we would expect. Being involved in a political organization or an arts group is highly negatively correlated with being a STEM major as individuals in these groups will likely major in fields that are more closely related to their interests outside of class.

Female friends is a variable that ranges from 0 to 4. Individuals were asked to list their four closest friends. This variable counts the number of these friends that are female. The number of female friends has a slight relationship with the probability of being a STEM major for men but not for women. Men are less likely to be a STEM major if they have more female friends. This result could be due to the fact male STEM majors have fewer female classmates and thus are less likely to have female friends. The effect on women is not statistically significantly different from zero.

There is some evidence that either men and women select into peer groups differently or that peer groups affect men and women differently. Certain peer groups

Table 1.11: Effects of Micro-Level Peer Groups on STEM Major Choice: Average Marginal Effects from a Probit Estimation

Variables	Specification (1)	(2)
Female	-0.042 (0.030)	-0.037 (0.029)
Female Friends	-0.006 (0.013)	-0.007 (0.013)
For Men		-0.027* (0.017)
For Women		0.011 (0.182)
Varsity Sports	-0.096** (0.035)	-0.083** (0.036)
For Men		-0.164** (0.037)
For Women		-0.004 (0.057)
Intramural	-0.001 (0.026)	0.004 (0.027)
For Men		-0.022 (0.034)
For Women		0.027 (0.039)
Foreign Language Group	-0.026 (0.045)	-0.030 (0.040)
For Men		-0.064 (0.051)
For Women		0.001 (0.061)
Greek life	-0.025 (0.027)	-0.024 (0.026)
For Men		-0.043 (0.036)
For Women		-0.006 (0.037)
Political Organization	-0.142*** (0.023)	-0.141*** (0.023)
For Men		-0.149*** (0.035)
For Women		-0.128*** (0.032)
Environmental Group	0.162*** (0.048)	0.176*** (0.050)
For Men		0.221*** (0.085)
For Women		0.135** (0.057)
Career Development	-0.025 (0.032)	-0.030 (0.033)
For Men		-0.028 (0.054)
For Women		-0.031(0.036)
Religious Organization	0.062** (0.028)	0.057* (0.028)
For Men		0.037 (0.043)
For Women		0.073** (0.036)
Arts	-0.063** (0.027)	-0.064** (0.027)
For Men		-0.046 (0.044)
For Women		-0.077** (0.031)
Volunteer Organization	-0.030 (0.023)	-0.035 (0.022)
For Men		-0.052 (0.034)
For Women		-0.019 (0.028)
N	7,821	7,821
Pseudo-R^2	0.142	0.149

Note: Standard errors in parenthesis.
Standard errors are clustered at the student level.
* indicates statistical significance at the 10% level.
** indicates statistical significance at the 5% level.
*** indicates statistical significance at the 1% level.

appear to have different impacts on the probability of being a STEM major for men and women (or at least, men and women select into both majors and these activities differently). For instance, male athletes are less likely to be STEM majors relative to female athletes. In certain activities, it seems like one gender is driving the overall effect of the activity; for example, the effect of being in a religious activity is larger and more significant for women and the effect of being involved in an environmental group is larger for men. However, for many activities, the marginal effect for men and women is fairly similar (i.e. participation in a political organization). It could be that these activities are selected into pretty uniformly for individuals with certain interests; for instance, students involved in political science are likely to major in political science as opposed to STEM regardless of gender.

The marginal effect of being involved in a religious activity is somewhat surprising, as being religious would appear to be unassociated with any particular academic area.

STEM Persistence

Table 1.12 shows the results from a probit estimation with micro-level peer groups and STEM persistence as the outcome of interest.

I examine the marginal effects of the various peer groups for both men and women to discover whether there is evidence that suggest that peer groups affect men and women differently. For female friends, the marginal effect for men is -0.037 (standard error 0.032) while it is 0.075 (standard error 0.032) for women, which is statistically significant at the 5% level, providing evidence that this measure of peer groups is related to STEM persistence differently for men and for women. The marginal effects are also different for varsity athletics where the marginal effect for men is -0.352 (standard error 0.091) which is statistically significant at the 0.1% significance level while the marginal effect for women is

Table 1.12: Effects of Micro-Level Peer Groups on STEM Persistence: Average Marginal Effects from a Probit Estimation

Variables	Specification (1)	(2)
Female	-0.117* (0.060)	-0.107* (0.057)
Female Friends	0.028 (0.025)	0.022 (0.023)
For Men		-0.037 (0.032)
For Women		0.075** (0.032)
Varsity Sports	-0.185** (0.088)	-0.154* (0.079)
For Men		-0.352*** (0.091)
For Women		0.041 (0.125)
Intramural	-0.049 (0.048)	-0.065 (0.048)
For Men		-0.033 (0.057)
For Women		-0.093 (0.074)
Foreign Language Group	0.038 (0.082)	0.057 (0.077)
For Men		0.101 (0.088)
For Women		0.018 (0.123)
Greek life	0.026 (0.057)	-0.001 (0.058)
For Men		-0.028 (0.070)
For Women		0.023 (0.091)
Political Organization	-0.245*** (0.070)	-0.233*** (0.069)
For Men		-0.278*** (0.101)
For Women		-0.179 ** (0.087)
Environmental Group	0.169** (0.074)	0.173** (0.074)
For Men		0.092 (0.106)
For Women		0.258** (0.113))
Career Development	-0.052 (0.057)	-0.074 (0.058)
For Men		-0.060 (0.082)
For Women		-0.085 (0.079)
Religious Organization	0.031 (0.048)	0.031 (0.045)
For Men		0.026 (0.065)
For Women		0.035 (0.064)
Arts	0.035 (0.061)	0.016 (0.058)
For Men		0.022 (0.083)
For Women		0.011 (0.083)
Volunteer Organization	-0.027 (0.044)	-0.037 (0.042)
For Men		-0.101 (0.062)
For Women		0.021 (0.057)
N	3,018	3,018
Pseudo-R^2	0.068	0.088

Note: Standard errors in parenthesis.
Standard errors are clustered at the student level.
* indicates statistical significance at the 10% level.
** indicates statistical significance at the 5% level.
*** indicates statistical significance at the 1% level.

not statistically different from zero. This difference is consistent with the differences reported for STEM major choice. It also appears that the marginal effect for women of environmental groups may be driving the overall marginal effect as it is only statistically significantly different from zero for women.

Number of Math Courses Taken

In terms of math course taking, the various peer groups have largely the same relationship with math course taking as we would expect and is consistent with results for STEM major choice. Being involved in varsity sports, politics, the arts, or a foreign language group has a statistically significant negative relationship with number of math courses taken. For many of these activities, it may be the case that being involved in the activity is highly correlated with a particular major that has a very low math requirement. When interaction terms are included, I find that students involved in Greek organizations take 0.124 fewer math courses than students not involved in Greek life; however, the interaction term is positive, providing evidence that the relationship is different for male students and female students. The coefficient on this interaction term provides further limited evidence that peers interact with academic choices in different ways for male students and female students.

Also notable from Table 1.13 is the fact that even among STEM majors, female students still take fewer math courses than male students. As expected, the coefficient on being a STEM major is positive (STEM majors take almost an additional half of a math course relative to non-STEM majors). This result relates to the descriptive statistics presented in Table 1.5; female students make up a higher proportion of the biological sciences and a lower proportion of mathematically-demanding STEM fields. However, I cannot disentangle whether female students are not taking math courses and thus not

Table 1.13: Effects of Micro-Level Peer Groups on Math Course Taking: Linear Estimation

Variables	Specification (1)	(2)
Female	-0.099***	-0.244***
	(-0.037)	(-0.084)
Female Friends	-0.004	-0.023
	(-0.017)	(-0.028)
Female Friends*Female		0.033
		(-0.032)
STEM	0.418***	0.540***
	(-0.042)	(-0.068)
STEM*Female		-0.259***
		(-0.081)
Varsity Sports	-0.129***	-0.240***
	(-0.038)	(-0.055)
Varsity Sports*Female		0.228***
		(-0.079)
Intramural	0.002	-0.016
	(-0.039)	(-0.059)
Intramural*Female		0.040
		(-0.075)
Foreign Language Group	-0.067*	-0.143**
	(-0.036)	(-0.062)
Foreign Language*Female		0.119
		(-0.075)
Greek Life	-0.056*	-0.124**
	(-0.033)	(-0.052)
Greek Life*Female		0.126*
		(-0.067)
Political Organization	-0.109***	-0.187***
	(-0.030)	(-0.053)
Political Organization*Female		0.120*
		(-0.064)
Environmental Group	-0.008	0.345
	(-0.095)	(-0.278)
Environmental Group*Female		-0.495*
		-(0.282)
Career Development	-0.028	-0.114**
	(-0.032)	(-0.058)
Career Development*Female		0.133*
		(-0.070)
Religious Organization	-0.049	-0.098*
	(-0.031)	(-0.056)
Religious Organization*Female		0.082
		(-0.065)
Arts	-0.098***	-0.191***
	(-0.033)	(-0.059)
Arts*Female		0.152**
		(-0.068)
Volunteer Organization	-0.076***	-0.132**
	(-0.029)	(-0.056)
Volunteer*Female		0.077
		(0.063)
N	7,821	7,821
Pseudo-R^2	0.129	0.145

Note: Standard errors in parenthesis.
Standard errors are clustered at the student level.
* indicates statistical significance at the 10% level.
** indicates statistical significance at the 5% level.
*** indicates statistical significance at the 1% level.

choosing these majors or if they are choosing other majors and thus have a lower mathematics requirement to fulfill.

Tests of Linear Combinations of Variables

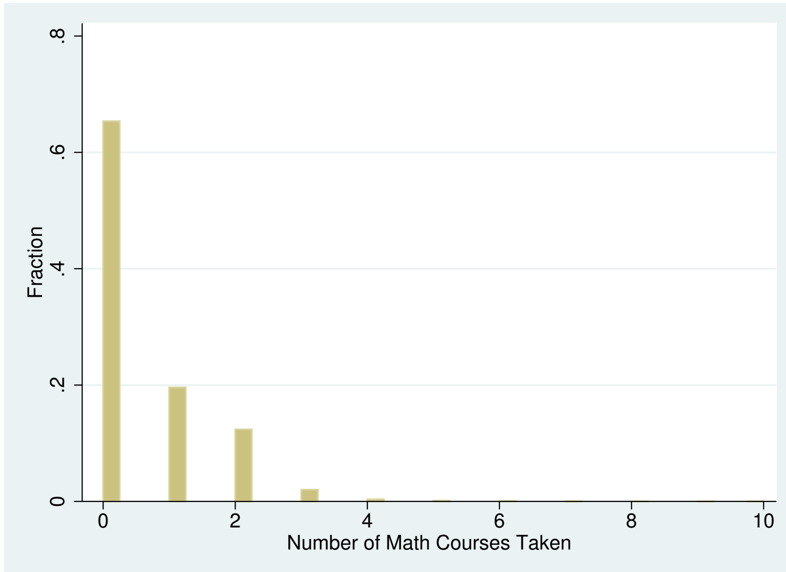
I test the linear combination of variables where the coefficient on the interaction term was statistically significant. I find that the linear combination of being female, involved in the arts, and their interaction is -0.182 and is statistically significant at the 10% significance level. Additionally, the linear combination of being female and the interaction term between being female and a STEM major is -0.388 and is statistically significant at the 1% significance level. This effect can be interpreted as saying that female STEM majors take 0.388 fewer math courses than male STEM majors. This result is as expected given the breakdown by gender within STEM majors described in Table 1.5.

1.7.3 Potential Problems

One possible problem with this model is that there are a large number of zero values for number of math courses taken by students (see Figures 1.4 and 1.5). STEM majors have fewer zero-values than non-STEM majors, but a large number of students still report taking zero math courses in college, possibly because of credit gained from joint enrollment or AP classes while still in high school. Hurdle models or zero-inflated models are possible models that could be used in the future to address this and may be better than the linear model that I use here.

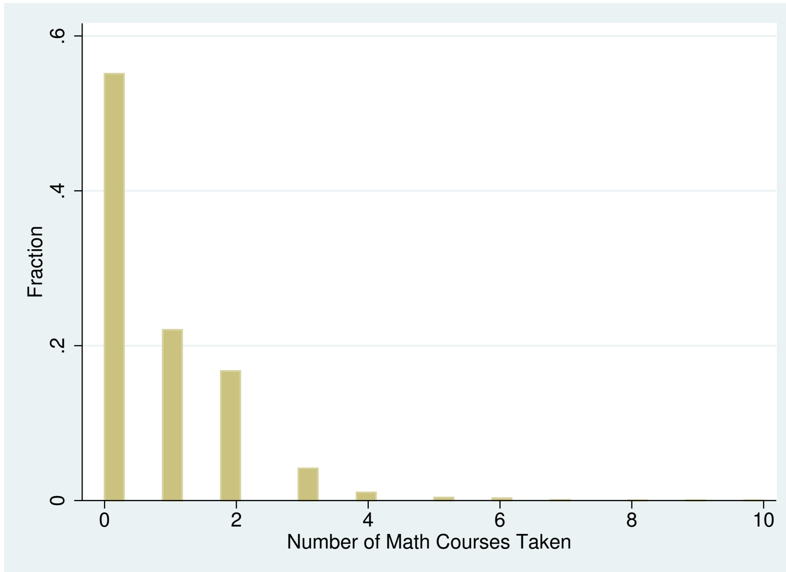
Possible weaknesses of this paper are difficulty fully controlling for selection into peer groups and weaknesses with the data set. Many of the peer groups, particularly extracurricular activities, are selected into based on unobservable characteristics that are likely also correlated with major selection. Additionally, although this data set contains a rich set of survey information, many questions are missing a response from a substantial

Figure 1.4: Number of Math Courses Taken



number of students, posing problems when the students who do not answer are likely systematically different from those who do. Because this is survey data, there are limitations to some of the responses. Many of the responses don't make sense; for instance, an alarmingly high percentage of students cannot remember whether they took calculus or physics in high school when asked about it their freshman year. Furthermore, the dataset includes a great deal of information about peer groups at certain waves in the study, but not at others. Because of this, I will be able to draw conclusions about the effects at those points and about decisions in the future, but I am unable, without making very bold assumptions, to draw conclusions about earlier decisions. Finally, my measure of the gender composition of STEM majors is potentially flawed since it is based off of the gender composition of the sample rather than the institution as a whole.

Figure 1.5: Number of Math Courses Taken by STEM Majors



1.8 Conclusion

Addressing the lack of women in the STEM pipeline is crucial to creating a diverse workforce that has the skills necessary for success in the 21st century. A considerable number of students who take advanced STEM classes in high school or express initial interest in a STEM field ultimately do not end up finishing their STEM majors. A crucial component of many of these fields is proper understanding of mathematical concepts, which is linked with taking math classes, another area where we see a gender gap between men and women. Part of understanding why we see these gaps in math course taking and STEM majors is the way that peer groups influence students at the college level with regards to their course taking and major choices.

This paper finds largely the expected relationship between gender and taking math courses and selecting a STEM major. In this sample, women are less likely to be STEM majors and take fewer math courses relative to men. I also find the expected relationships between various peer groups and the outcomes of interest (i.e. students involved in artistic and political activities are less likely to be STEM majors). I find limited evidence that peer groups affect men and women in different ways.

This research could be extended to younger age groups where peer groups are possibly more formative and less well established. At this age, there are additionally larger policy implications in terms of how cohorts are divided into classes that would not be implementable at the college level due to the fact that students in college have much greater choice in selection into certain institutions, majors, and classes.

1.9 References

Bettinger, E. (2010). "To Be or Not to Be: Major Choices in Budding Scientists American Universities in a Global Market" (pp. 69-98): University of Chicago Press.

Carrell, Scott E. and Richard L. Fullerton and James E. West (2009). "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," Journal of Labor Economics, University of Chicago Press, vol. 27(3), pages 439-464, 07.

Eccles, J. S. (2006). "Where are all the women? Gender differences in participation in physical science and engineering." In Why aren?t more women in science? Top researchers debate the evidence, ed. S. J. Ceci and W. M. Williams, 199?210. Washington, DC: American Psychological Association.

Fryer, Roland G., and Steven D. Levitt (2010). "An Empirical Analysis of the Gender Gap in Mathematics." American Economic Journal: Applied Economics2.2: 210-40. Web.

Griffith, Amanda L. (2008). "Determinants of Grades, Persistence and Major Choice for Low-Income and Minority Students?. Cornell Higher Education Research Institute Working Papers. Web.

Griffith, Amanda L. (2010). "Persistence of Women and Minorities in STEM Field Majors: Is It the School That Matters?" Economics of Education Review29.6: 911-22. Web.

Hoxby, Caroline (2000). "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper Series: n. pag. Web

Lavy, Victor and Analia Schlosser (2011). “Mechanisms and Impacts of Gender Peer Effects at School,” American Economic Journal: Applied Economics, American Economic Association, vol. 3(2), pages 1-33, April.

Manski, Charles F. (1993). “Identification of Endogenous Social Effects: The Reflection Problem.” The Review of Economic Studies 60.3: 531-42. JSTOR. Web. 10 Sept. 2014.

Rendall, Andrew and Michelle Rendall (2014). “Math Matters: Educational Choices and Wage Inequality?. University of Zurich Working Papers Series. . n. 160. Web.

Sacerdote, Bruce and David Marmaros (2006). “How Do Friendships Form?” The Quarterly Journal of Economics 121, 1: 79-119.

Sacerdote, Bruce and James West (2013). “From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation ,? Econometrica , 81(3): 855 - 882 , 2013

1.10 Appendix A: List of STEM Majors

Biological Science Majors

Biochemistry
Bio-engineering
Biological Basis of Behavior
Biology
Environmental Science
Neuroscience
Nursing
Pharmacy
Physical Therapy
Pre-Med
Sports science
Zoology

Engineering Majors[4]

Bio-engineering[5]
Chemical Engineering
Civil Engineering
Electrical Engineering
Aerospace Engineering
Engineering, unspecified

Other STEM Majors

Actuarial Science
Chemistry
Computer Science
Material Science
Mechanical Science
Physics
Mathematics[6]
Science, unspecified

[4]Computer engineering is indistinguishable in the data from computer science so I classify it as that

[5]Bio-engineering counts for both a biological science major and an engineering major

[6]Strangely, Statistics is not listed as a major choice. In the list of departments that students can take classes in, it is listed as mathematics/statistics.

1.11 Appendix B: Coefficient Estimates from Probit Estimation

Table 1.14: Coefficient Estimates from a Probit Estimation on STEM Major: Macro-Level Peer Groups

	Specification					
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.138**	-0.139**	-0.153**	-0.139**	-1.507	
	(-0.068)	(-0.068)	(-0.067)	(-0.068)	(-0.986)	
Percent Female			0.0928	2.595	1.412	3.85
			(-0.484)	(-2.947)	(-3.135)	(-7.196)
Precent Female*Female					0.778	-1.8
					(-2.083)	(-9.975)
Percent Female STEM			0.169	-0.0642	-0.997***	-1.812***
			(-0.261)	(-0.277)	(-0.384)	(-0.653)
Percent Female*STEM					2.036***	4.142***
					(-0.502)	(-0.922)
SAT Score	0.0002	0.0005*	0.0005*	0.0005*	0.0005*	
	(-0.0002)	(-0.0003)	(-0.0003)	(-0.0003)	(-0.0003)	
STEM Freshman	1.120***	1.099***	1.089***	1.099***	1.091***	
	(-0.0757)	(-0.079)	(-0.0775)	(-0.079)	(-0.079)	
Liberal Arts			-0.179			
			(-0.129)			
Private Research			-0.180**			
			(-0.0772)			
Includes race dummies	X	X	X	X	X	
Year fixed effects	X	X	X	X	X	
Institution fixed effects		X		X	X	
Includes high school courses	X	X	X	X	X	
Individual fixed Effects						X
N	3,228	3,228	3,228	3,228	3,228	3,228
$PsuedoR^2$	0.037	0.069	0.045	0.069	0.074	

Note: Standard errors in parenthesis.
Standard errors are clustered at the student level.
* indicates statistical significance at the 10% level.
** indicates statistical significance at the 5% level.
*** indicates statistical significance at the 1% level.

Table 1.15: Coefficient Estimates from a Probit Estimation on STEM Persistence: Macro-Level Peer Groups

	Specification					
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.051	-0.113	-0.118	-0.113	-1.479	
	(-0.114)	(-0.119)	(-0.115)	(-0.119)	(-1.594)	
Percent Female			0.411	1.529	0.323	4.497**
			(-0.890)	(-4.905)	(-5.251)	(1.951)
Percent Female * Female					0.958	-6.531***
					(-3.468)	(1.718)
Percent Female STEM			0.581	-0.346	-1.229	-2.480*
			(-0.583)	(-0.579)	(-0.796)	(1.289)
Percent Female STEM * Female					1.836	5.994***
					(-1.143)	(1.779)
SAT	0.0006	0.0008*	0.0009**	0.0008*	0.0008*	
	(-0.0004)	(-0.0005)	(-0.0005)	(-0.0005)	(-0.0005)	
Liberal Arts			-0.0287			
			(-0.2320)			
Private Research			-0.162			
			(-0.1220)			
Includes race dummies	X	X	X	X	X	
Year fixed effects	X	X	X	X	X	
Institution fixed effects		X		X	X	
Includes high school courses	X	X	X	X	X	
Individual fixed Effects						X
N	3,303	3,303	3,303	3,303	3,303	3,303
$PsuedoR^2$	0.037	0.069	0.045	0.069	0.074	

Note: Standard errors in parenthesis.
Standard errors are clustered at the student level.
* indicates statistical significance at the 10% level.
** indicates statistical significance at the 5% level.
*** indicates statistical significance at the 1% level.

Table 1.16: Coefficient Estimates from a Probit Estimation on STEM Major: Micro-Level Peer Groups

Variables	Specification (1)	(2)
Female	-0.148	-0.624**
	(-0.105)	(-0.263)
Female Friends	-0.021	-0.098
	(-0.046)	(-0.062)
Female Friends*Female		0.140
		(-0.093)
Varsity Sports	-0.376**	-0.704***
	(-0.155)	(-0.194)
Varsity Sports*Female		0.691**
		(-0.289)
Intramural	-0.004	-0.081
	(-0.093)	(-0.124)
Intramural*Female		0.184
		(-0.188)
Foreign Language Group	-0.097	-0.245
	(-0.169)	(-0.206)
Foreign Language Group * Female		0.248
		(-0.306)
Greek Life	-0.091	-0.157
	(-0.098)	(-0.133)
Greek Life*Female		0.134
		(-0.191)
Political Organization	-0.590***	-0.614***
	(-0.114)	(-0.161)
Political Organization*Female		0.053
		(-0.229)
Environmental Group	0.521***	0.700***
	(-0.144)	(-0.251)
Environmental Group*Female		-0.246
		(-0.307)
Career Development	-0.092	-0.104
	(-0.118)	(-0.203)
Career Development*Female		-0.016
		(-0.249)
Religious Organization	0.214**	0.131
	(-0.095)	(-0.147)
Religious Organization*Female		0.129
		(-0.192)
Arts	-0.235**	-0.17
	(-0.107)	(-0.166)
Arts*Female		-0.137
		(-0.210)
Volunteer Organization	-0.110	-0.188
	(-0.082)	(-0.125)
Volunteer Organization*Female		0.116
		(-0.163)
N	7,821	7,821
Pseudo-R^2	0.142	0.149

Note: Standard errors in parenthesis.
Standard errors are clustered at the student level.
* indicates statistical significance at the 10% level.
** indicates statistical significance at the 5% level.
*** indicates statistical significance at the 1% level.

Table 1.17: Coefficient Estimates from a Probit Estimation on STEM Persistence: Micro-Level Peer Groups

	Specification	
Variables	(1)	(2)
Female	-0.319*	-1.166***
	(-0.167)	(-0.356)
Female Friends	0.075	-0.103
	(-0.069)	(-0.091)
Female Friends * Female		0.310**
		(-0.132)
Varsity Sports	-0.504**	-1.013***
	(-0.247)	(-0.312)
Varsity Sports*Female		1.126**
		(-0.466)
Intramural	-0.135	-0.0928
	(-0.131)	(-0.161)
Intramural*Female		-0.167
		(-0.261)
Foreign Language Group	0.103	0.293
	(-0.228)	(-0.265)
Foreign Language Group*Female		-0.243
		(-0.427)
Greek Life	0.0703	-0.0794
	(-0.156)	(-0.197)
Greek Life*Female		0.144
		(-0.318)
Political Organization	-0.671***	-0.783**
	(-0.208)	(-0.309)
Political Organization*Female		0.264
		(-0.416)
Environmental Group	0.483**	0.266
	(-0.230)	(-0.317)
Environmental Group*Female		0.459
		(-0.471)
Career Development	-0.141	-0.167
	(-0.155)	(-0.228)
Career Development*Female		-0.0714
		(-0.317)
Religious Organization	0.0862	0.0735
	(-0.131)	(-0.183)
Religious Organization*Female		0.0224
		(-0.257)
Arts	0.067	0.063
	(-0.166)	(-0.234)
Arts*Female		-0.033
		(-0.328)
Volunteer Organization	-0.073	-0.281
	(-0.120)	(-0.173)
Volunteer Organization*Female		0.339
		(-0.234)
N	3,018	3,018
Pseudo-R^2	0.068	0.088

Note: Standard errors in parenthesis.
Standard errors are clustered at the student level.
* indicates statistical significance at the 10% level.
** indicates statistical significance at the 5% level.
*** indicates statistical significance at the 1% level.