The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence*

Stefanie Fischer
Cal Poly State University, San Luis Obispo
sjfische@calpoly.edu

March 13, 2017

Abstract

This paper investigates whether class composition can help explain why women are disproportionately more likely to fall out of the "STEM" pipeline. Identification comes from a standardized enrollment process at a large public university that essentially randomly assigns freshmen to different mandatory introductory chemistry lectures. Using administrative data, I find that women who are enrolled in a class with higher ability peers are less likely to graduate with a STEM degree, while men's STEM persistence is unaffected. The effect is largest for women in the bottom third of the ability distribution. I rule out that this is driven solely by grades.

JEL Codes: I20, I230, I240
Keywords: Higher Education, Gender, STEM, Classroom Composition Effects

*I thank the office of Institutional Research, Planning and Assessment at UC Santa Barbara for access to administrative data. This paper has benefited from helpful comments from Kelly Bedard, Peter Kuhn, Shelly Lundberg, Dick Startz, Maya Rossin-Slater, Jon Sonstelie, UCSB Labor Lunch Seminar participants, members of the UCSB Human Capital Working Group, Brian Duncan, Elira Kuka, and other participants at the annual Student Dissertation Workshop at the WEAI, participants at the All-California Labor Economics Conference, participants at the Society of Labor Economists Annual Meeting, participants at the Banff Empirical Microeconomics Conference, as well as seminar participants at Colgate University, Utah State University, UC Irvine Ed Department, and Cal Poly - SLO. All errors are my own.

1 Introduction

Women exit STEM (science, technology, engineering, and math) majors at much higher rates than men, which in part contributes to the gender-wage gap because they miss out on the sizable premiums associated with employment in these fields.[1] To date, little is known regarding women's decision to drop out of the STEM pipeline in college. Recent research finds that the instructor-student gender match plays a role in this decision (Carrell et al., 2010; Hoffmann and Oreopoulos, 2009), and that women's responsiveness to grades explains some of the phenomenon as well (Rask and Tiefenthaler, 2008; Ost, 2010), but much remains unexplained. Gaining a better understanding of the factors that cause women to leave STEM fields during college is important for developing policy aimed at bolstering this group's STEM persistence. This paper examines a novel pathway: *To what extent does the composition of a woman's introductory university STEM course impact STEM persistence?*

There are many ways to define classroom composition. In this study it is defined as the share of high ability students in a mandatory introductory STEM lecture, where ability includes effort and motivation, innate ability, and any resources that might aid in academic success.[2] There are several ways in which being assigned to a college lecture with relatively higher ability peers could affect one's outcomes. On one hand, this type of environment could be performance enhancing. Students may benefit directly from higher ability classmates through knowledge spillovers during class, office hours, or out-of-class group study sessions. Additionally, the average class ability can affect the overall standard, and students may be motivated to work harder to keep up with their high achieving peers.

On the other hand, a high achieving classroom environment may be harmful in more subtle ways by negatively impacting self-perception. The higher the ability of the peers in a classroom, the harder it is to be ranked high. While in many situations high ability peers can improve perfor-

[1] Paglin and Rufolo (1990); Murnane et al. (1995); Grogger and Eide (1995); Brown and Corcoran (1997); Weinberger (1999); Weinberger (2001); Murnane et al. (2000); Rose and Betts (2004)

[2] Section 2.1 and Section 2.2 outline why this measure of average class ability is favorable, and from where exogenous variation in the share of high ability students per lecture arises.

mance, contest theory suggests that large gaps in skills between individuals can have the perverse effect of reducing effort incentives. Brown (2011) provides empirical evidence for this theoretical prediction by showing that the presence of a superstar in a PGA golf tournament is associated with lower performance by the other competitors. In a classroom setting, marginal students may see themselves as relatively weaker and reduce their effort or even choose to opt out if possible (i.e., opt out of the STEM pipeline).

Moreover, women's STEM persistence may be particularly affected by the class ability composition. STEM courses have a distinct environment, they tend to be large, difficult, and competitive relative to non-STEM courses. A National Science Foundation study from the early 2000s, which surveyed roughly 25,000 undergraduate women in STEM, revealed that women are most likely to leave STEM during their first year. The two most cited reasons for leaving were dissatisfaction with grades and the heavy workload, and a distaste for the overall competitive climate (Goodman, 2002). In a related vein, Niederle and Vesterlund (2007) and Garratt et al. (2013) show that women shy away from competition more than men and that the gender performance gap is exacerbated under competition (Gneezy et al., 2003). To the extent that increasing the share of high ability students in the class increases the competitive environment, it may induce the marginal women to leave.

In addition to the sparse literature on women's STEM persistence, there exists relatively little research on post-secondary classroom composition effects because isolating the causal impact is difficult. Often students, or more indirectly administrators, influence the student make-up of a classroom. Several studies at the elementary school level, which for the most part rely on data from the large scale randomized experiment Project STAR, find a positive relationship between average classmate ability and achievement (Whitmore, 2005; Hanushek et al., 2003; Boozer and Cacciola, 2001; Hoxby, 2003). Whether these results extend to a higher education setting, however, is an open question.

To date, the most convincing peer effects study in higher education – which estimates small positive effects on freshman grade point average – relies on the random assignment of students from the United States Air Force Academy to squadrons, which are essentially cohorts (Carrell et al., 2009).[3] This peer group measure is an improvement over previous studies, which define dorm-mates as the peer group, because squadrons capture a more comprehensive set of students' peer interactions.[4] None of these studies, however, capture the effects that students within a classroom may have on individual outcomes because dorm-mates and squadron members do not necessarily attend the same classes.

In this study, I employ administrative data from a large public research institution, University of California Santa Barbara, to estimate the relationship between class ability composition in an introductory STEM course and STEM persistence. This setting is ideal because it circumvents several empirical challenges associated with estimating composition effects. Firstly, the standardized way in which students load into their first STEM course (General Chemistry) – a mandatory prerequisite for nearly all STEM majors at most universities – leaves little room for non-random class enrollment (see Section 2.3). This aspect of the empirical setting allows for the estimation of class ability composition effects free of the usual problem of self-selection.

Secondly, the Transfer Admission Guarantee (TAG) agreement – a program that offers students from California community colleges guaranteed admissions to this university and several other University of California campuses, conditional on meeting certain requirements – generates variation in the ability composition across sections of this introductory STEM course. In particular, the TAG agreement creates stark differences in ability between those students who obtain admissions into the university directly out of high school and those who enter through TAG (see Section 2.1). Thirdly, the introductory STEM course analyzed, General Chemistry, is a required prerequisite for most STEM majors and students cannot circumvent the course by applying Advanced Placement or community college credits, which eliminates another avenue for selection.

[3] Similarly, Carrell et al. (2013) run an experiment at the USAFA and show low ability students are harmed by their higher ability squadron members – results which are in-line with my findings. Lyle (2007) uses a similar military dataset (USMA) and cohort approach and finds no evidence of peer effects. A drawback of both of these studies is that students from military institutions likely are not representative of the general university population, especially women.

[4] The following studies use the random assignment of students to dorms to estimate peer effects and find mixed results. Stinebrickner and Stinebrickner (2006) find small positive peer effects on grades for women. Zimmerman (2003) and Sacerdote (2001) find small positive peer effects on students' grades, grade point average, and the take-up of social networks such as fraternities/sororities. Foster (2006) finds no evidence of peer effects.

Exploiting this quasi-experimental setting, I show that being in a class with higher ability peers reduces the probability that women graduate with a STEM degree and has no significant effect on men. More specifically, a 15 percent increase in the number of high ability students in a General Chemistry lecture (one standard deviation) reduces the probability that the average woman graduates with a STEM degree by 3.1 percentage-points (6.8 percent). As one might expect, I further show that the effect is strongest in the bottom third of the math ability distribution. I rule out grades in the same course as the underlying mechanism by showing that the gender differential effect persists even after controlling for one's grade in this introductory course.

The results are informative for at least two reasons. One, they are among the first to show that classmate influences are an important factor in determining students' academic success in higher education, at least for women.[5] Two, in contrast to most previous work, I focus on STEM major completion rather than grades because STEM completion closely relates to occupation choice, which is an important piece to the gender wage gap story (Murnane et al., 2000; Rose and Betts, 2004).

The remainder of the paper is organized as follows: Section 2 describes the empirical setting. Section 3 presents the econometric specification and results. Section 4 concludes.

2 Empirical Setting

All data in this study are drawn from the University of California Santa Barbara (UCSB) administrative data system. UCSB is a selective, public research institution with a large undergraduate population; enrollment is roughly 23,000 students and the acceptance rate is around 36%. In this analysis, I study the group of UCSB students who are enrolled in the course CHEM1A – which is the first quarter of the yearlong introductory STEM sequence, General Chemistry – in a fall quar-

[5] Feld and Zölitz (2017) in a recent working paper concurrently show that low ability students are harmed by relatively higher ability peers. Our studies differs in a variety of ways but one key way is that students in their sample are placed in very small classes (15) compared with the average class size of 330 in my sample. As such, the way in which peers affect academic outcomes are quite different across these two settings.

ter between 1997 and 2007.[6] This includes 11 entering cohorts of about 1,200 students each, and tracks them through graduation. Each fall quarter there are four or five sections of CHEM1A with on average 329 students per section. At this university, and at the majority of other universities, General Chemistry is the first prerequisite for most STEM majors.[7] Students are required to take this course at UCSB; they cannot apply advanced placement credits or test out of the class, which is a strategy often used by students to circumvent introductory math and statistics.

General Chemistry is an ideal setting to study the relationship between STEM persistence and class ability composition because it is a mandatory prerequisite and it has all of the characteristics common to STEM; the course is difficult and competitive. Twenty-five percent of the university's on campus tutoring resources (Campus Learning Assistant Services) go to General Chemistry students. Often the course is considered a STEM major weed-out. Earning high grades can be difficult as the typical curve is 20 percent "As", 30 percent "Bs", and 30 percent "Cs" per CHEM1A section. The final course grade is weighted heavily toward exams with two midterms, two quizzes, and a final. Homework only counts for 10 percent of one's final grade. After each exam, students are notified of their place in the overall grade distribution. Given these characteristics of the course, studying the STEM persistence behavior of the group enrolled in CHEM1A captures the students who are most attached to earning a STEM degree.

2.1 Measure of Average Class Ability

To construct the class ability composition variable, the data are divided into two distinct groups, on-track students (freshmen) and late-track students (sophomores or higher). Importantly, because of a key University of California policy, these two groups differ in observable and unobservable ways. On-track students are students who are enrolled in CHEM1A in the fall quarter of their first year at the university. They were admitted into the university directly out of high school and are on-track to graduate with a STEM degree in four years. Late-track students are upperclassmen taking

[6]This gateway sequence consists of CHEM1A, CHEM1B and CHEM1C; CHEM1A is offered in the fall, CHEM1B in the winter, and CHEM1C in the spring.

[7]Appendix Table A1 lists the STEM majors that require CHEM1A.

CHEM1A in a fall quarter other than their freshman year; they are behind schedule to graduate with a STEM degree in four years. Roughly 85 percent of the observations in the sample are on-track, totalling 12,230 students. The other 15 percent of students (1,912) are late-track. I leverage this variation to construct the main measure of average class ability, the share of on-track students in a section. From here forward, I use the term section to refer to what many would consider a lecture or class. Sections are defined as a unique time, instructor, and year.

In order for the share of on-track students in a section to be a good measure of average class ability, the following must hold: (1) late-track students, on average, must differ in ability from on-track students, and (2) there must exist exogenous variation across CHEM1A sections in the share of on-track students (or alternatively, the share of late-track students). Below I outline the institutional structure which suggests that late-track students are relatively lower ability than on-track students. The institutional structure also provides insight as to how variation in the share of on-track students across sections arises. Finally, I provide evidence from the data supporting conditions (1) and (2).

There are several reasons to believe that late-track CHEM1A students differ on observable (and unobservable) characteristics from on-track students – in particular that they are, on average, lower ability. The late-track group includes three types of students: upperclassmen who switched to a STEM major at some point after entering the university, upperclassmen that needed a year of preparatory courses – remedial math and science – before starting pre-major requirements, and transfer students. In this sample, conditional on being late-track, 621 are transfers and 1,291 are non-transfer upperclassmen. The first type of upperclassmen, although they are behind in their STEM courses, are not necessarily lower ability. The second type, those who come to the university with an inadequate background and take a year of preparatory courses, are, on average, lower ability.

Transfer students, the third type, typically come from the local city college (or community college). In fact, in the 2015 entering cohort, 94 percent of the transfer students came from a California community college.[8] Transferring from a two-year state institution to one of the University of California (UC) campuses is particularly common in California because of the TAG agreement (Transfer Admission Guarantee). This program guarantees admission to one of six UCs if a student completes 30 units with the given minimum grade point average at a public California two-year college.[9] The TAG agreement was put into place in the early 1980s and the participating UCs include: Santa Barbara, Davis, San Diego, Irvine, Riverside, and Merced.

Central to this paper's empirical approach, TAG and other transfer students are likely financially constrained or did not earn admissions into UCSB directly out of high school. Either way, on average, they have lower high school grades and socioeconomic characteristics relative to the on-track group. To provide additional context on TAG students, in 2014 UC San Diego petitioned to opt out of the TAG agreement stating that it was squeezing out too many traditional students who had more competitive applications.[10]

Using the share of on-track students in a section as the main measure of average section-mate ability in this setting is preferred to other measures typically used in the literature such as the section average of a predetermined characteristic that proxies for ability, i.e. SAT scores. The variable on-track potentially captures a more comprehensive set of observable and unobservable characteristics related to an individual's ability compared with, for example, SAT score. Furthermore, twenty-five percent of the sample's SAT scores are missing since TAG students and other transfers are not required to take the SAT, which, in fact, is the group that generates variation in ability across sections.[11]

Finally, the data support the notion that on-track students are on average higher ability than late-track students on observable characteristics. Table 1 reports average predetermined student characteristics and outcomes for each group. Column 1 reports averages for transfer students –

[8] This statistic comes from the following website: https://admissions.sa.ucsb.edu/docs/default-source/PDFs/ucsb-admission-guide-2016.pdf?sfvrsn=2. Unfortunately, these numbers were not available for the time period of the dataset.

[9] Grade point average requirements vary by institution, but range from 2.8 to 3.3

[10] Information on UCSD's decision to opt out of the TAG agreement can be found in the San Diego Union-Tribune: http://www.sandiegouniontribune.com/news/2012/may/01/ucsd-ends-community-college-transfer-program/

[11] In a robustness check that is available upon request, I use a more conventional linear-in-means, leave-me-out approach and find similar results.

a subset of late-track students, Column 2 reports averages for all late-track students, Column 3 includes averages for on-track students, and Column 4 reports the difference in means between on-track and late-track students and indicates if the difference is statistically different from zero. Note that transfer students conditional on reporting an SAT score, on average, have 44 percent of a standard deviation lower SAT math score than the on-track group, a characteristic that tends to be a strong predictor of STEM success. Put differently, the average transfer SAT math score is around the 69th percentile on the national scale while the average on-track score is around the 81st percentile. Moreover, on-track students on average have a higher high school grade point average than non-transfer, late-track students.[12] Finally, on-track students on average are more likely to graduate in STEM, and their actual grade in CHEM1A is a full letter grade higher.

2.2 Variation in the Share of On-Track Students Across Sections

Variation in CHEM1A class composition arises from late-track student enrollment patterns. While on-track students face a no-priority registration policy, late-track students are able to selectively enroll and they do.[13] Late-track students register for fall classes the previous spring before on-track students, who register in the summer. Based on estimated freshmen fall enrollment, the university holds a share of the CHEM1A seats in each section for incoming freshmen – the majority of seats are held for freshmen as this is typically a first-year course – and offers the other seats to late-track students. In some cases late-track students fill all of the seats allotted to them in a given section and in other cases they do not. Once late-track students have enrolled, all of the remaining seats are filled with on-track students who load into sections through a standardized process which achieves as good as random assignment.

For illustrative purposes, suppose that there are four sections of CHEM1A offered in a given year and each has a maximum enrollment of 100. Further suppose that the university decides to

[12]Unfortunately, the reported high school grade point average for transfer students is from their previous college (typically Santa Barbara City College) and all other reported high grade point averages are from high school. The former is on a 4 point scale while the latter is on a 4.75 scale to account for Advance Placement courses. As such, the two GPAs are not comparable which is why in Table 1 " High School GPA" is not reported for transfers.

[13]See Section 2.3.4 for details on late-track student's enrollment patterns.

make available 30 percent of the seats in each section for late-track students. If in one of the sections all 30 seats are filled, in the second 25 are filled, and in the third and fourth only 20 and 15 respectively are filled, then variation will arise in the number of late-track students across sections.

In the second stage of the process, on-track students fill the remaining seats in each section in a standardized way. Because CHEM1A is a highly demanded class, it is rare for a section not to fill to capacity. The assignment process for on-track students and the enrollment patterns of late-track students are discussed in detail in Section 2.3.

2.3 Student Assignment to Lectures

2.3.1 Identifying Assumption

The aim of this study is to understand how the ability composition of a section differentially affects STEM persistence for men and women. In order to interpret the estimates as causal a sufficient but not necessary assumption is that characteristics explaining a student's academic achievement are uncorrelated with the quality of classmates in a student's section. Random assignment of students to sections will ensure the validity of this assumption. However, because the goal is to understand how the average class ability *differentially* effects men and women, in order to interpret the estimated effect as causal, it is only necessary that there is no gender based selection. For example, the causal interpretation of the estimated effect is questionable if high ability women systematically enroll in sections with a low share of on-track students and high ability men systematically enroll in sections with a high share of on-track students. In this scenario, it is unclear if the composition estimate is capturing the true effect of ability composition on STEM persistence or if it is reflecting student selection into sections.

To alleviate concerns of student selection and gender based student selection, in the analysis I only study the outcomes of freshmen (on-track students) and rely on UCSB's standardized no-priority enrollment process for this group, which achieves assignment to sections that is as good as random (see Section 2.3.2). While the institutional structure is such that there should be essentially no sorting of on-track students into sections, in Section 2.3.3 I provide multiple tests of randomization that support both the stronger assumption of no student selection in general, and the weaker but more relevant assumption of no differential selection by gender. Even though late-track students do not directly enter the analysis, they do through the composition variable. As such, Section 2.3.4 documents late-track student's enrollment patterns and provides evidence that their sorting is not a threat to identification. Finally, as a way to further reduce the possibility of student selection introducing bias, I include several fixed effects: year, time of day, and instructor (see Section 3.1). The results are also robust to the inclusion of additional fixed effects: instructor by year, instructor by time of day, instructor by student gender, and time of day by student gender.

2.3.2 Institutional Background – Student Assignment to Sections

During my sample period, 95% of all first year students attend a two-day summer orientation either in June, July or August where they register for their fall quarter courses.[14] Importantly, there is no priority based registration during or before summer orientation for this group. In each orientation session, a percentage of the total seats available in a given "first year" course are made available to that particular orientation session. This equalizes the probability of enrolling in a particular section across all orientation sessions and eliminates the issue of students who attend earlier orientation dates getting all the "good" classes.

At orientation each student is assigned to a group of 15 students. They are placed into orientation groups by declared major, but groups within major are formed randomly. With this group, students attend seminars about university life, map out a class schedule for the first quarter (and first year) under the guidance of a trained orientation leader for their declared major, and register for their first quarter classes. For instance, students in a pre-Biology major orientation group are advised by their leader to take General Chemistry during their first quarter so that they are on-track to get into the major and ultimately graduate in four years.

The structure of course registration is such that only one student in an orientation group is able to register at a time; each group has one laptop and one leader who facilitates registration. Within

[14]Each summer 12 freshman orientation dates are offered and students can attend the date of their choice.

each group a registration queue is formed by random draw (i.e. creating registration position 1-position 15). The student who draws position 1 registers first. Registration opens at the same time for all groups within an orientation session. According to the Director of Orientation Programs and Parent Services at UCSB, the high demand for CHEM1A, the limited seats, and the random registration order makes strategic CHEM1A registration very unlikely. Making selection even harder, the Chemistry Department strictly enforces a no switching policy. A student can only switch lectures during the first week of the quarter and he must have a student in his desired lecture replace him in his original one: a one for one switch.[15]

One concern related to student selection that I cannot directly rule out is the possibility that students who attend summer orientation differ from those who do not. Students who do not attend summer orientation register for their fall classes in mid-September prior to the start of the quarter but after all orientation attendees have registered. Composition estimates will be biased if students who do not attend summer orientation are non-representative and systematically register for sections based on the ratio of on-track to late-track students. If I could observe CHEM1A registration dates I could test for balance using observed predetermined traits between students who attend summer orientation and those who do not. Since these data are unavailable for my sample period, I have instead obtained registration data for all freshmen enrolled in CHEM1A in fall 2013. Although these students are not in my main sample, the registration behavior should be similar, as the general structure of freshman registration is the same between the two periods.

For this group of students – all freshmen enrolled in CHEM1A in fall of 2013 – I observe their CHEM1A registration date and time as well as demographics and CHEM1A instructor characteristics. Ninety percent of this sample attended a summer orientation/registration, slightly lower than the 95% in the main sample. Comparing the observable characteristics of students who attend orientation and those who do not, underrepresented minorities is the only group who is underrepresented in orientation attendance; 38% of the orientation attending group are URMs compared

[15] All information regarding freshman orientation and registration comes from an interview with Kim Equinoa (kim.equinoa@sa.ucsb.edu) who was the director of Orientation Programs and Parent Services at UCSB during the years in which the data for this analysis are from. Information on the Chemistry Department's no switching rule comes from the administrative office within the Chemistry Department.

to 49% of the non-orientation attending group. There appears to be no selection into orientation attendance by gender, parent's education level, whether one has a high school GPA in the top half of the distribution for the sample, whether one scores in the top half of the SAT math or SAT verbal distribution for the sample, type of high school one attended, and English language learner status. Most importantly, the data indicate that there is no statistically significant difference in the share of on-track students in a section for those who attend summer orientation and those who do not. Appendix Table A2 reports these results.

Finally, even though the way in which on-track students load into this introductory STEM course leaves little room for selective enrollment, suppose a student does have enough flexibility to manipulate her schedule. It is possible a student has preferences for a particular section start time, the day of the week that the section meets, or for a particular instructor. There is no evidence, however, of students sorting in a differential fashion related to start time – that is, high ability women are no more likely to select into the morning sections than men – and, all specifications include a control for section start time. In this setting, it is also not possible for students to sort by the day of the week that the section meets because all CHEM1A sections meet on the same days: Mondays, Wednesdays and Fridays for fifty minutes each meeting. Selective enrollment based on instructor preferences is also of little concern because on-track students lack information about instructors; they enroll in classes before they move to campus. Moreover, the data for this study come from a time period that, for the most part, predates the online site Rate My Professor and other similar sites that provide a public forum to disseminate information on instructor quality. Instructor fixed effects are also included in all regressions, further alleviating the concern that sorting based on time-invariant instructor characteristics could bias the point estimates.

2.3.3 Is On-track Student Assignment to Sections Really Random?

While the identifying assumption – no student sorting across sections and/or no gender based sorting – is fundamentally untestable, I provide several indirect tests of its plausibility. First consider a series of balance regressions presented in Table 2. Each column in this table corresponds to a separate regression where a different predetermined student characteristic is regressed on lnO_{itnd} and $F_i * lnO_{itnd}$. lnO_{itnd} is the log of the number of on-track students in a section for student i who takes CHEM1A in year t with instructor n at time of day d. $F_i * lnO_{itnd}$ is that variable interacted with a female indicator.[16] Year fixed effects are also included as students within a year face the same no-priority registration, but not necessarily across years. Excluded from the test are two sections that have enrollment of less than 100 as they were likely added last minute to meet a larger than expected demand for CHEM1A. Results for the main analysis are robust to the exclusion of these two sections.

If selection is present, the coefficient on lnO_{itnd} will attain statistical significance. Furthermore, if gender based selection exists, the interaction term $F_i * lnO_{itnd}$ will be significant. Although the coefficient on lnO_{itnd} is statistically different from zero at the 10% level in three cases, the estimates are essentially zero in magnitude. More importantly for my purposes, there is no evidence of a gender differential in selection across section composition; $F_i * lnO_{itnd}$ is not statistically different from zero in any of the five cases.

As a second way to empirically test the assumption of random assignment, I employ the resampling technique used by Carrell and West (2008) and Lehmann et al. (1986). I test for student selection into CHEM1A sections by five pre-treatment characteristics: high school grade point average, math SAT score, verbal SAT score, parent is a college graduate, and unrepresented minority status. For each section I calculate the number of on-track students in a section and then randomly draw 10,000 samples of equal size without replacement from the group of all on-track students enrolled in any CHEM1A section in a given year. For each randomly sampled section, I compute the average pre-treatment characteristic (i.e. high school grade point average) and an empirical p-value representing the share of simulated sections with an average pre-treatment characteristic less than the observed average for the given section. If assignment is random, the distribution of empirical p-values will be uniform since any p-value is equally likely to be observed. I obtain 55 p-values, one per year per pre-treatment trait. I test for uniformity of the empirical p-values by

[16] An explanation for the log transformation is presented in Section 3.

year by trait using a Kolmogorov-Smirnov one-sample equality of distribution test.

Table 3 Panel A reports the results and shows that for all eleven years and the five pre-treatment characteristics, the null hypothesis of random assignment is not rejected at the five percent level. In summary, I find very little empirical evidence that on-track students are selectively enrolling into sections by academic ability.

2.3.4 Enrollment Patterns of Late-Track Students

While on-track students face a no-priority registration policy, late-track students are able to selectively enroll, and they do. In fact, as mentioned in Section 2.2, the enrollment patterns of late-track students is exactly how variation in the share of on-track students arises. At the time of registration, late-track students observe the time of day of the CHEM1A sections and the instructors.[17] As such, it is possible that they enroll based on start times and/or based on the reputation of the instructor's quality or difficulty. Although the dataset used in this analysis pre-dates websites that publicly provide information on instructor quality (including listing grade distributions, syllabi etc.), some of these students may have information on instructors from friends.

Through a series of section-level regressions which examine the correlation between observable section characteristics and the share of late-track students, I show late-track students sort by time of day, and not by instructor. Results for this exercise are presented in Table 4. The outcome is the share of late track students in a given section. The first two columns examine observable instructor characteristics. Column 1 shows very little correlation between the perceived difficulty of the instructor and share of late-track students, where the measure of instructor difficulty is the average grade assigned by a given instructor in all previous sections. There is also no evidence of students sorting by the instructor's gender (Column 2). Perhaps unsurprising, the margin by which late-track students select into sections is time of day. There is a larger share of late-track students in afternoon sections (Column 3).

Recall the identifying assumption, student characteristics that explain academic achievement

[17] They also observe the day of the week of the section, but there is no variation. All sections meet on the same days for the same amount of time.

must be uncorrelated with the ability composition of the class. That is, even though late-track students selectively enroll based on section start time, there is little concern that this threatens the causal interpretation of the results. First, all regressions include time of day fixed effects. Second, similar to the indirect tests for selective enrollment conducted for on-track students, a balance test for late-track students reported in Table 5 shows no sign that this group is sorting into sections based on observable characteristics that are correlated with the ability composition of the section. In addition, I employ the resampling exercise for the group of late-track students, which is reported in Table 3 Panel B, and reject the null hypothesis of random assignment in only two of the 55 cases at the 10 percent level of significance.[18]

Finally, to further alleviate concerns surrounding the enrollment patterns of late-track students, because the are able to selectively enroll, I only include on-track students in the analysis.[19] The only way in which late-track students enter the analysis is through the ability composition variable (the share of on-track students per section).

2.4 Data

Table 6 presents summary statistics for the sample, which includes only on-track students. From 1997-2007 there are 46 CHEM1A sections taught by 13 different instructors. The average section size is 329 students, but there is little variation. Since CHEM1A is a highly demanded course, sections typically fill to capacity. That is, section size is determined by the number of seats in the lecture hall. A majority of the lectures are 300-370 students.[20] On average, on-track students make up 85 percent of each section; the minimum is 71 percent and the maximum is 96 percent.

The main outcome of interest is STEM completion, defined as graduating with a STEM major from UCSB within five years. Among entering freshmen who take CHEM1A, the average STEM

[18] Recall from Section 2.2 that on-track students and late-track students enroll in CHEM1A at different times and under different rules; which in fact is the very reason that there are differences in average ability across sections. As such it is necessary to perform this simulation for the two groups separately. A simulation where the two groups are considered together will, by construction, result in the rejection of random assignment.

[19] In a robustness check presented in Table A5, I show the results hold when late-track students are also included.

[20] The results are not sensitive to the exclusion of the 25 percent of lectures that are smaller than 300 students.

completion rate is 53 percent for men and 45 percent for women. The average CHEM1A grade for males is a 2.65 GPA (on a 4 point scale) and a 2.49 for women.

UCSB administrative data also include several socioeconomic measures: race, sex, high school grade point average, SAT math and verbal scores, type of high school (public or private), parents' highest education level, English proficiency and age. Limited information is also available regarding instructors and course times. These include an instructor's sex and a unique instructor identification number, as well as the year, day, and time of the section. These data are linked to students.

3 Econometric Specification and Results

The primary specification is the following linear probability model:

$$G_{itnd} = \alpha_1 + \beta_1 F_i + \beta_2 lnO_{tnd} + \beta_3 F_i * lnO_{tnd} + \alpha_2 C_{tnd} + \alpha_3 X_i + \alpha_3 M_{tnd} + \phi_t + \rho_n + \varepsilon_{itnd} \quad (1)$$

The variable G_{itnd} denotes STEM major completion for on-track student i who takes CHEM1A in year t with instructor n at time of day d; tnd uniquely identifies an individual section in a specific year. F is a female indicator variable and O_{tnd} is the number of on-track students in a specific section. The log transformation allows one to interpret the on-track estimate as a percent change and takes into account that a one student change is proportionately larger from a small base.[21] The coefficient β_2 captures the effect of the number of on-track students per section on the outcome for men. The coefficient on the interaction term $F_i * lnO_{tnd}$ is the differential effect of the number of on-track students per section for women. Thus, for women the percentage-point change in STEM graduation associated with a percent increase in the number of on-track students is $\beta_2 + \beta_3$.

C_{tnd} controls for several class level characteristics: the log of the total number of students enrolled in a given section and the percent female. The section size variable includes the log of the total number of on-track and late-track students enrolled in a given section. X_i is a vector of student

[21] In an alternative specification, I use percent of on-track students in a lecture as the measure of class composition and obtain similar results, see Section 3.3.

background characteristics including: race, if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores, age, and the student's intended major at entry (hard sciences, Biology/Environmental Science, Social Sciences, Humanities/Arts/interdiciplinary, and undeclared). M indicates that the section took place in the morning (starting at 8 a.m. or 9 a.m.). Year fixed effects (ϕ_t) are included to control flexibly for time trends in STEM completion. Since many instructors appear repeatedly, I include instructor fixed effects (ρ_n) to control for time-invariant instructor differences. All standard errors are clustered at the section level (instructor/year/time of day).

Fewer than 1 percent of on-track observations are missing SAT scores, and 0.5 percent of on-track observations are missing high school grade point average. To deal with these missing values, I impute values using the average of those with a reported value by sex. For example, for a female missing the SAT math score, I fill in with the average SAT value of all female on-track students in the sample. I also include a vector of indicator variables, one for each pretreatment variable, in all regressions which takes on a value of one if the pretreatment variable is missing.

3.1 Main Results

Results from the main specification (Equation 1) – which estimates the differential impact of the number of on-track students in a section for men and women – are reported in Table 7. Column 1 reports results for the full sample and shows that increasing the number of on-track students in a class by 15 percent reduces the probability that a woman graduates with a STEM major by 3.1 percentage-points (see Column 1, panel B). Increasing the number of on-track students by 15 percent in the average section is equivalent to adding 44 more on-track students to a section with 281 on-track students, which is about one standard deviation. To give context to the magnitude of the results, the average STEM graduation rate for women is 45 percent and 53 percent for men. Thus, a 15 percent increase in the number of on-tracks student in a section decreases the STEM graduation rate for an average woman from 45 percent to about 42 percent, which is a decrease of 6.6 percent. For men, Column 1 suggests that there is no statistically significant relationship between the ability of the students in his CHEM1A section and the rate at which he persists in STEM. Importantly, the interaction term – which estimates the differential effect of the number of on-track students per section for women – is statistically different from zero (see Panel A, Row 2) indicating that the effects for men and women are different from one another.

Table A3, a companion to Table 7, reports the main results with varying levels of controls. If student sorting across sections is present and introducing bias, then including various controls – e.g., fixed effects for instructor or time of day, or student level characteristics – should influence the point estimates. However, the main effect – both gender differential effect and total effect – changes very little across eight different specifications, further providing reassurance that student selection into sections is not driving the results.

Column 1 is the most parsimonious specification and only includes year fixed effects, recall student assignment is random within a year. Columns 2, 3 and 4 add to this parsimonious specification by including instructor fixed effects, time of day fixed effects, and class and student level predetermined characteristics, respectively. Note that Column 4 reproduces the main specification for reference. Columns 5-8 extend the main specification by adding various two-way fixed effects.

Column 5 adds instructor by year fixed effects to control for characteristics of an instructor that differ across years but are common within a year. Column 6 adds instructor by time of day fixed effects to the main specification, which controls for aspects of an instructor that are common to a certain time of the day. Two additional specifications are included to control for the potential that female and male students may respond differentially to the time of the day of the section or to a particular instructor. Column 7 adds instructor by student gender fixed effects to the main specification, which controls for instructor characteristics that affect male and female students differentially . Finally, Column 8 adds time of day by student gender fixed effects to the main specification which rules out, for example, the possibility that women perform better in the afternoon and thus are more likely to graduate with a degree in STEM.

3.2 What Does the Estimated Composition Effect Include?

While the aim of this analysis is to understand the total effect of a student's section-mates on that student's STEM outcomes, it is worth noting that in principle the total estimated composition effect can embody three distinct effects: correlated effects, endogenous peer effects, and exogenous peer effects (also known as contextual effects) (Manski, 1993). Correlated effects are present when groups of individuals form based on common characteristics – i.e. ability – and, as a result, behave similarly. This is often caused by students self-selecting into a group. Because, in my setting, students load into sections in an as good as random way, the composition estimates are free of correlated effects.

Endogenous peer influence is often described as the "reflection" problem and refers to the empirical challenge of disentangle the effect of a group on an individual's outcome from the effect that an individual has on the group (Moffitt et al., 2001; Sacerdote, 2001). The two are often determined simultaneously as peer interactions are reciprocal in nature. In an attempt to mitigate reflection, I follow a strategy common to this literature; I control for previous peer achievement (Carrell et al., 2009, 2013; Hanushek et al., 2003).

Lastly, contextual effects capture the effect of a student's classmate's predetermined characterises – high school grade point average, SAT scores etc. – on her own outcomes. It is often the goal of the empiricist to isolate the exogenous effect net of the other two effects, but in this study I am interested in estimating the total composition effect. The presence of an endogenous effect does not undermine the empirical findings.[22]

A final concern is that the results are an artifact of a mechanical relationship between the measure of own and peer ability as described in Angrist (2014). A mechanical relationship stems from measurement error and is distinct from selection bias. In this study, however, because I show assignment of students to sections is as good as random, any bias stemming from measurement error will only attenuate the composition estimates. That is, the estimated negative composition

[22]If one believes that endogenous effects are present, and assuming that both the exogenous and endogenous effects are negative, my estimate of the "total effect" will overstate the exogenous effect. That is, the estimate will be inflated by a social multiplier, and be more negative than the true exogenous effect.

effect is an upper bound; the true effect may be even more negative (Feld and Zölitz, 2017).

3.3 Sensitively and Heterogeneity Analysis

There are two sections in the sample that are substantially smaller than the rest having fewer than 100 students. Although the simulation presented in Table 3 reveals that students are not sorting into sections based on observable characteristics, investigating these small sections is warranted. More than likely these sections were added last minute to meet a larger than expected CHEM1A demand, but the data do not allow one to observe added sections.[23]

If these small sections are added, the on-track students assigned to them have the greatest potential to be non-representative. For instance, the small percent of on-track students who do not attend a summer orientation session and also enroll in CHEM1A (which is on average five percent of an incoming freshman class) are most likely assigned to an add-on lecture during the first week of school. One would expect this non-summer orientation attending group of students to be less advantaged, thereby dampening the estimated on-track student effect found in the main specification. [24]

Column 2 of Table 7 reports the estimates for the subsample which excludes sections with fewer than 100 students; variation in the share of on-track students per section ranges from 75 to 96 percent [25] Results for this subsample indicate that small sections are not driving the main findings. In fact, the magnitude of the estimated composition effects for women and men are not statistically different from the estimated effects using the whole sample.

The results are robust to an alternative measure of average class ability. Table A4 reports the results using the percent of on-track students in a section as the measure of the ability composition instead of the log of the number of on-track students. The main difference between these two mea-

[23] Although many years the Chemistry Department accurately estimates the demand for CHEM1A, there are cases where they add an additional lecture the week before the fall term begins.

[24]Non-orientation attending students are likely less advantaged because summer orientation is an additional cost. According to UCSB office of Orientation Programs and Parent Services, the most common reason students do not attend orientation is due to summer employment.

[25]A balancing test for the subsample is statistically the same as the balancing test for the main sample. This table is available upon request.

sures is that percent on-track assumes a constant marginal effect regardless of the base. Although somewhat noisier, the estimated effects for men and women are quite similar in magnitude; the interaction term is just shy of statistical significance at the 10 percent level. I also estimate Equation 1 as a probit model rather than a linear probability model and obtain similar results. In a final robustness check, I expand the sample to include late-track and on-track students. These results are presented in Table A5 and are quite similar to those found in the original sample.

Next, understanding which group of students is driving the main result is important for developing and implementing interventions. Columns 3-5 of Table 7 report results disaggregated by SAT math score. The effect is strongest for women in the bottom third of the SAT math distribution. A 15 percent increase in the number of on-track students in a class reduces the probability by 5.7 percentage-points that a women in this SAT math group completes college with a degree in STEM, and this effect is statistically different from the estimated effects in the other two SAT categories (Columns 4 and 5). Consistent with the main finding, the men in all subsamples, including those in the bottom third of the SAT math distribution, appear to be unaffected by the classroom composition.

It is intuitive that women in the lower part of the math ability distribution are the group most affected by classroom composition since they are the group most at risk of dropping out of STEM. In a working paper, Astorne-Figari and Speer (2017) document a similar pattern. They find that women switch out of relatively more competitive majors, while men do not, and this result is particularly pronounced at the lower end of the ability distribution. They also show that this gender switching gap is not reduced by controlling for grades, something I document as well (see Section 3.4). Alternatively, our results oppose the findings in Carrell et al. (2010). They find that the group influenced by STEM interventions are women at the top of the SAT math distribution. In particular, they document that women in the top 25 percent of the SAT math distribution with female STEM instructors are more likely to graduate with a STEM major.

One might wonder if the results truly are a gender effect. It is possible that I am capturing an underrepresented minority effect or merely picking up the fact that all students at the bottom end of the SAT math distribution are less likely to graduate with a degree in STEM. Columns 1 and 2 of Table 8 report results from the specifications outlined in Equations 3 and 4 respectively. These models, which are extensions of Equation 1, include a triple interaction term allowing one to disentangle differences in the ability composition effect across gender and race (Column 1), as well as gender and position in the SAT math distribution (Column 2).

$$G_{itnd} = \alpha_1 + \beta_1 F_i + \beta_2 lnO_{tnd} + \beta_3 URM_i + \beta_4 F_i * URM_i + \beta_5 lnO_{tnd} * URM_i + \beta_6 F_i * lnO_{tnd} + \quad (2)$$

$$\beta_7 F_i * URM_i * lnO_i + \alpha_2 C_{tnd} + \alpha_3 X_i + \alpha_3 M_{tnd} + \phi_t + \rho_n + \varepsilon_{itnd}$$

$$G_{itnd} = \alpha_1 + \beta_1 F_i + \beta_2 lnO_{tnd} + \beta_3 Low_i + \beta_4 F_i * Low_i + \beta_5 lnO_{tnd} * Low_i + \beta_6 F_i * lnO_{tnd} + \quad (3)$$

$$\beta_7 F_i * Low_i * lnO_i + \alpha_2 C_{tnd} + \alpha_3 X_i + \alpha_3 M_{tnd} + \phi_t + \rho_n + \varepsilon_{itnd}$$

URM_i denotes whether a student is an underrepresented minority (black, Hispanic, American Indian or Filipino) and Low_i indicates whether a student falls in the bottom third of the SAT math distribution for the sample in a given year. All other variables are as defined in Equation 1. Column 1 shows that all women, regardless of race, have STEM persistence rates that are negatively affected by the number of on-track students in her CHEM1A class. As reported in Column 1 Panel B, URM and non-URM women experience a 3.0 percentage-point decline in STEM persistence as a result of an increased number of on-track students. Again, there is no detectable class composition effect for men, URM or non-URM.

Results presented in Table 8 Column 2 further support a gender story. These results show that only women (and not men) in the bottom third of the SAT math distribution for the sample have STEM persistence rates that are affected. In fact, women in this group are 4.0 percentage-points less likely to graduate in STEM as a result of a 15 percent increase in the number of on-track students in a class. The results for the women are statistically different from zero and statistically different from men in this same SAT math group.

$$G_{itnd} = \alpha_1 + \beta_1 F_i + \beta_2 lnO_{tnd} + \beta_3 B_i + \beta_4 F_i * B_i + \beta_5 lnO_{tnd} * B_i + \beta_6 F_i * lnO_{tnd} + \quad (4)$$

$$\beta_7 F_i * Low_i * lnO_i + \alpha_2 C_{tnd} + \alpha_3 X_i + \alpha_3 M_i + \phi_t + \rho_i + \varepsilon_{itnd}$$

Finally, socioeconomic status may also play a role in one's willingness to leave STEM when placed in a lecture with a higher share of on-track students. I use a similar triple difference specification – as outlined in Equation 4 – and examine the differential effect of the ability composition on STEM persistence by gender and by parent's level of education. B_i indicates if a student has at least one parent with a bachelor's degree. Results presented in Column 3 of Table 8 show that all women, regardless of whether her parent is a college graduate, have an increased probability of exiting STEM. Consistent with all other specifications, the persistence rate for men in all subgroups seems to be statistically unrelated to the composition of the class. Together, these findings provide strong evidence that *women* in the bottom third of the math ability distribution are the group most affected by the ability of their classmates. There is no evidence to support the conjecture that it is merely reflecting minority status, being in the bottom of the ability distribution, or socioeconomic status.[26]

3.4 Grades as a Possible Mechanism

These reported findings raise the question: *Why are women less likely to graduate with a STEM degree if their first experience with STEM is in a setting with higher ability classmates, and why are men unaffected by this factor?* While little is known regarding post-secondary classroom composition effects and student outcomes in general, less is known about the mechanisms at work, and in particular why composition matters more for women.

One possible candidate is grades. Several studies in economics find that women are more responsive to grades than men, and as a result exit STEM majors (Rask and Tiefenthaler, 2008; Ost, 2010).[27] It is also likely the class ability composition influences the grade a student earns in this class. Students who randomly end up in sections with more high ability classmates will receive lower grades relative to their counterpart (those in lectures with fewer on-track students) if there is

[26] In an additional unreported analysis, I further investigate if the gender composition of the late-track group plays a role in explaining female's stem persistence, and find no conclusive results.

[27] The average grade in STEM courses is much lower than humanities, social sciences, arts and interdisciplinary courses.

a similar grading curve in each class. For instance, suppose that there are two CHEM1A sections with equal enrollment in a given fall quarter and one has more on-track students than the other. Relying on the fact that sections with more on-track students are overall higher ability (Table 1), a student receiving a score of 77 percent in the section with more on-track students will be assigned a lower final letter grade than if she were in a section with fewer on-track students.

This scenario is likely since all CHEM1A instructors assign final course grades based on a similar curve. The curve is applied within a section and the typical distribution is 20 percent A grades, 30 percent B grades, and 30 percent C grades.[28] Once enrolled in the course, from the syllabus students can observe that there is a curve imposed in each section and that the final course grade is largely determined by performance on exams. The final grade breakdown is usually 40 percent final exam, 20 percent midterm 1, 20 percent midterm 2, 10 percent quizzes, and 10 percent homework. Quizzes are typically multiple choice and midterms and the final are a combination of multiple choice and short answer. Instructors write and grade their own exams; that is, exams are not common across sections within the same fall quarter.

If grades in the introductory course are driving this longer-run negative effect on STEM completion, then controlling for CHEM1A grade in the main specification (Equation 1) should diminish the composition estimates. This does not seem to be the case. Table 9 reports such estimates and shows that the effect remains despite controlling for grade. It appears that men also experience a negative composition effect once controlling for CHEM1A grade, but importantly the gender differential effect persists suggesting that women are still more negatively affected. In summary, the results presented in Table 9 suggest that grades are likely not the main driver underling the negative composition effect.

There are, however, various other ways in which the ability composition may discourage women that are consistent with my findings. One possibility is that the composition of the class could affect students' self-perception about their immediate and future success in the major. Pre-

[28]It is possible for instructors to adjust the curve up or down by a few percentage points depending on the section (i.e. 22 percent earn an A grade, 28 percent earn a B grade etc.). Nevertheless, "good" grades remain scarce and students are aware of this from the onset.

sumably, all students enter the initial course with an expectation about how they will do. Throughout the course they learn about their relative standing and update beliefs about themselves accordingly. Individuals in sections with relatively higher ability classmates may adjust these beliefs differently compared to those who are not. For example, Pop-Eleches and Urquiola (2013) show that students who just make it into better high schools receive better exam scores but also report feeling marginalized and relatively weaker compared to students who are placed in classes with lower ability classmates. To the extent that women's self-perception about their future success is more negatively affected by the ability of those around them, it could explain their much lower retention rate. In a related vein, if women are more risk averse, then the marginal ones may switch to majors where they perceive having a higher chance of "making it" while marginal men gamble by staying in STEM. Consistent with this idea, Kuziemko et al. (2014) show that men are more likely to gamble to avoid low rank whereas women accept it.[29]

Along these lines, I show that women still graduate but that they respond to the composition of their introductory STEM course by switching into majors that are on average lower paying, less quantitative, and arguably less competitive.[30] Table 10, Column 1 reports that increasing the number of on-track students in a class by 15 percent leads to a 3.2 percentage-point increase in the probability that a woman graduates with a humanities, social sciences, art, or interdisciplinary major.[31]

4 Conclusion

It has been well documented that women are less likely than men to persist in STEM majors and careers. This study targets a unique group of students, those taking General Chemistry in their first quarter of college, to better understand how one's first collegiate experience in STEM explains

[29] Although data for this study comes from the laboratory and manipulates an individual's rank in the wealth distribution, it is reasonable that the detected behavioral response extends to a classroom ability distribution.

[30] Appendix Table A6 outlines by sex the percent of students in each major category at entry and at graduation.

[31] I get similar results when I exclude Economics (Econ, Econ-Math, and Econ-Accounting) and Psychology from this group.

STEM major graduation rates. Relying on data containing roughly 12,000 first year university students from 11 entering cohorts between 1997-2007, I estimate the causal relationship between the ability of one's classmates in a required STEM major course and a student's STEM major completion.

In summary, I find women who are assigned to a STEM lecture with higher ability peers at the start of their university career are less likely to persist in STEM while men's persistence behavior is unaffected. I rule out the possibility that women earn lower grades in classes with higher ability classmates and as such are less likely to persist.

This study is the first to provide an analysis of the relationship between classroom composition and STEM degree completion in higher education, and to document the differential response by gender. It is also among the first to examine the effects of classroom composition in higher education. Broadly, the results from this study suggests that women's longer run STEM persistence is affected by her experience in the gateway course, and, in particular, that the classroom ability composition plays a crucial role. By identifying this new channel through which women opt out of STEM, the estimates presented in this paper provide potentially important information to policymakers attempting to bolster the participation of women in STEM fields.

References

Angrist, Joshua D, "The Perils of Peer Effects," *Labour Economics*, 2014, *30*, 98–108.

Astorne-Figari, Carmen and Jamin Speer, "Are Changes of Major, Major Changes? The Roles of Grades, Gender, and Preferences in College Major Switching," *Working Paper*, 2017.

Boozer, Michael and Stephen E Cacciola, "Inside the'Black Box'of Project STAR: Estimation of Peer Effects using Experimental Data," *Yale Economic Growth Center Discussion Paper*, 2001, (832).

Brown, Charles and Mary Corcoran, "Sex-Based Differences in School Content and the Male-Female Wage Gap," *Journal of Labor Economics*, 1997, *15* (3), 431–65.

Brown, Jennifer, "Quitters Never Win: The (adverse) Incentive Effects of Competing with Superstars," *Journal of Political Economy*, 2011, *119* (5), 982–1013.

Carrell, Scott E and James E West, "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," Technical Report 2008.

_ , Bruce I Sacerdote, and James E West, "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 2013, *81* (3), 855–882.

_ , Marianne E Page, and James E West, "Sex and Science: How Professor Gender Perpetuates the Gender Gap," *The Quarterly Journal of Economics*, 2010, *125* (3), 1101–1144.

_ , Richard L Fullerton, and James E West, "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 2009, *27* (3), 439–464.

Feld, Jan and Ulf Zölitz, "Understanding Peer Effects: On the Nature, Estimation and Channels of Peer Effects," *Journal of Labor Economics*, 2017.

Foster, Gigi, "It's Not your Peers, and it's Not your Friends: Some Progress Toward Understanding the Educational Peer Effect Mechanism," *Journal of Public Economics*, 2006, *90* (8), 1455–1475.

Garratt, Rodney J, Catherine Weinberger, and Nick Johnson, "The State Street Mile: Age and Gender Differences in Competition Aversion in the Field," *Economic Inquiry*, 2013, *51* (1), 806–815.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini, "Performance in Competitive Environments: Gender Differences," *The Quarterly Journal of Economics*, 2003, *118* (3), 1049–1074.

Goodman, Irene F., "Final Report of the Women's Experience in College Engineering (WECE) Project," *Online Submission*, 2002.

Grogger, Jeff and Eric Eide, "Changes in College Skills and the Rise in the College Wage Premium," *Journal of Human Resources*, 1995, pp. 280–310.

Hanushek, Eric A, John F Kain, Jacob M Markman, and Steven G Rivkin, "Does Peer Ability Affect Student Achievement?," *Journal of Applied Econometrics*, 2003, *18* (5), 527–544.

Hoffmann, Florian and Philip Oreopoulos, "A Professor Like Me the Influence of Instructor Gender on College Achievement," *Journal of Human Resources*, 2009, *44* (2), 479–494.

Hoxby, Caroline, "The Power of Peers: How Does the Makeup of a Classroom Influence Achievement," Technical Report, Education Next 2003.

Kuziemko, Ilyana, Ryan W Buell, Taly Reich, and Michael I Norton, "Last-place Aversion: Evidence and Redistributive Implications," *The Quarterly Journal of Economics*, 2014, *129* (1), 105–149.

Lehmann, Erich Leo, Joseph P Romano, and George Casella, *Testing Statistical Hypotheses*, Vol. 150, Wiley New York et al., 1986.

Lyle, David S, "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point," *The Review of Economics and Statistics*, 2007, *89* (2), 289–299.

Manski, Charles F, "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542.

Moffitt, Robert A et al., "Policy Interventions, Low-level Equilibria, and Social Interactions," *Social Dynamics*, 2001, *4*, 45–82.

Murnane, R.J., J.B. Willett, and F. Levy, "The Growing Importance of Cognitive Skills in Wage Determination," Technical Report, National Bureau of Economic Research 1995.

_ , _ , Y. Duhaldeborde, and J.H. Tyler, "How Important are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings," *Journal of Policy Analysis and Management*, 2000, *19* (4), 547–568.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away from Competition? Do Men Compete Too Much?," *The Quarterly Journal of Economics*, 2007, *122*, 1067–1101.

Ost, Ben, "The Role of Peers and Grades in Determining Major Persistence in the Sciences," *Economics of Education Review*, 2010, *29* (6), 923–934.

Paglin, M. and A.M. Rufolo, "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences," *Journal of Labor Economics*, 1990, pp. 123–144.

Pop-Eleches, Cristian and Miguel Urquiola, "Going to a Better School: Effects and Behavioral Responses," Technical Report, American Economic Review 2013.

Rask, Kevin and Jill Tiefenthaler, "The Role of Grade Sensitivity in Explaining the Gender Imbalance in Undergraduate Economics," *Economics of Education Review*, 2008, *27* (6), 676–687.

Rose, H. and J.R. Betts, "The Effect of High School Courses on Earnings," *Review of Economics and Statistics*, 2004, *86* (2), 497–513.

Sacerdote, Bruce I, "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, 2001, *116* (2), 681–704.

Stinebrickner, Ralph and Todd R Stinebrickner, "What Can Be Learned About Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds," *Journal of Public Economics*, 2006, *90* (8), 1435–1454.

Weinberger, Catherine J, "Mathematical College Majors and the Gender Gap in Wages," *Industrial Relations: A Journal of Economy and Society*, 1999, *38* (3), 407–413.

Weinberger, C.J., "Is Teaching More Girls More Math the Key to Higher Wages?," *Squaring Up: Policy Strategies to Raise Womens Income in the US*, 2001.

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

Zimmerman, David J, "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics*, 2003, *85* (1), 9–23.

Table 1: On-Tracks are Relatively Higher Ability than Late-Tracks

	Transfers	Late-Track	On-Track	Diff. (3) -(2)
	(1)	(2)	(3)	(4)
Predetermined Characteristics				
Women	0.46	0.52	0.49	-.03**
	(0.50)	(0.50)	(0.50)	(0.01)
URM (underrepresented minority)	0.33	0.32	0.32	0.00
	(0.47)	(0.47)	(0.47)	(0.01)
High school G.P.A	–	3.68	3.75	0.06***
		(0.35)	(0.32)	(0.01)
SAT math score	577.67	604.34	612.74	8.4**
	(81.22)	(80.95)	(80.84)	(2.25)
SAT verbal score	544.40	575.91	569.58	-6.33**
	(84.04)	(83.66)	(84.59)	(2.33)
Public high school	0.94	0.86	0.85	-0.01
	(0.24)	(0.35)	(0.36)	(0.01)
English spoken in home	0.67	0.69	0.67	-0.02
	(0.47)	(0.46)	(0.47)	(0.01)
No parent college grad.	0.38	0.29	0.33	0.04 **
	(0.48)	(0.46)	(0.47)	(0.01)
Outcomes				
Graduate with STEM major	0.45	0.44	0.49	0.05**
	(0.50)	(0.50)	(0.50)	(0.01)
Graduate	0.76	0.82	0.81	-0.01
	(0.43)	(0.38)	(0.39)	(0.01)
CHEM1A grade	2.05	2.26	2.57	0.31**
	(1.16)	(1.11)	(0.93)	(0 .03)
Took follow-on (CHEM1B)	0.65	0.63	0.82	0.19**
	(0.48)	(0.48)	(0.38)	(0.01)
Grade in follow-on (CHEM1B)	2.31	2.53	2.58	0.05
	(0.96)	(0.97)	(0.87)	(0.03)
Observations	621	1,935	12,230	

Notes: On-track students are enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB from 1997 to 2007. Late-track students are enrolled in CHEM1A during this time frame but are taking the course as an upperclassman or transfer student. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. Note that only 159 of the 621 transfers report SAT scores. High school GPA is not reported for transfers because the data only contain their previous college GPA and not their high school GPA. Level of significance is indicated as follows: ** p<0.01, * p<0.05, + p<0.1. Standard deviations are in parentheses.

Table 2: Are On-Track Students Selectively Enrolling Based on Ability Composition?

	URM	SAT Math	SAT Verbal	H.S. GPA	Parent is College Grad
	(1)	(2)	(3)	(4)	(5)
Panel A: Point Estimates					
Ln(no. of on-track)	-0.119+	7.372	19.740	0.063	0.084
	(0.059)	(8.062)	(12.210)	(0.030)	(0.073)
Ln(no. of on-track) X fem.	0.057	5.852	-1.788	-0.016	0.042
	(0.078)	(4.634)	(5.250)	(0.036)	(0.075)
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track					
The differential effect	0.80	0.82	-0.25	0.00	0.01
Women	-0.87	1.85	2.51	0.01+	0.02
Men	-1.70+	1.03	2.76	0.01+	0.01
Observations	12,122	12,036	12,036	12,054	12,122

Notes: Each column is a separate regression and also includes year fixed effects as well as a female indicator. On-track students are those enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB between the years 1997 and 2007. This sample excludes two small classes with enrollment less than 100. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. Level of significance is indicated as follows: ** p<0.01, * p<0.05, + p<0.1. Robust standard errors are in parentheses.

Table 3: Randomization Check for Section Assignment

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Panel A: On-Track Students											
H.S. GPA											
K-S Test P-values	0.274	0.17	0.284	0.277	0.281	0.273	0.274	0.283	0.284	0.28	0.178
SAT math											
K-S Test P-values	0.273	0.166	0.287	0.272	0.277	0.266	0.28	0.286	0.28	0.266	0.176
Sat verbal											
K-S Test P-values	0.275	0.167	0.281	0.283	0.268	0.275	0.274	0.282	0.263	0.266	0.173
Col. grad parent											
K-S Test P-values	0.281	0.174	0.27	0.284	0.274	0.277	0.283	0.279	0.272	0.273	0.171
URM											
K-S Test P-values	0.266	0.177	0.278	0.275	0.273	0.271	0.278	0.284	0.274	0.277	0.165
Panel B: Late-Track Students											
H.S. GPA											
K-S Test P-values	0.26	0.164	0.272	0.281	0.274	0.267	0.267	0.281	0.26	0.186	0.169
SAT math											
K-S Test P-values	0.273	0.175	0.209	0.23	0.244	0.244	0.25	0.131	0.259	0.273	0.10
Sat verbal											
K-S Test P-values	0.218	0.171	0.269	0.233	0.216	0.192	0.256	0.193	0.208	0.053	0.146
Col. grad parent											
K-S Test P-values	0.287	0.174	0.29	0.271	0.296	0.285	0.266	0.273	0.27	0.288	0.16
URM											
K-S Test P-values	0.272	0.172	0.283	0.301	0.275	0.271	0.274	0.278	0.273	0.285	0.168

Note: The reported P-values correspond to a Kolmogorov-Smirnov test of uniformity. The values indicate that the null hypothesis of random assignment is not rejected at any conventional level of statistical significance for on-track students. It is only rejected at the 10% level in 2 of the 55 cases for late-track students. Both set of results suggest there is little concern of student sorting across sections.

Table 4: How are Late-Track Students Selectively Enrolling?

Outcome: Share Late-Track	(1)	(2)	(3)
Ave. grade [n-1]	0.012		
	(0.048)		
Prof. female		0.013	
		(0.017)	
Late morning			-0.0025
			(0.015)
Afternoon			0.043*
			(0.022)
Year FE	X	X	X
Observations	46	46	46

Note: Each column is a separate regression and also includes year fixed effects. Data are collapsed to the section level. Each regression includes 46 observations. Level of significance is indicated as follows: ** p<0.01, * p<0.05. + p<0.1. Robust standard errors are in parentheses.

Table 5: Are Late-Track Students Selectively Enrolling Based on Ability Composition?

	URM	SAT Math	SAT Verbal	H.S. GPA	Parent is College Grad
	(1)	(2)	(3)	(4)	(5)
Panel A: Point Estimates					
Ln(no. of on-track)	-0.097	-11.890	7.272	0.285	0.0613
	(0.092)	(20.700)	(21.680)	(0.223)	(0.094)
Ln(no. of on-track) X fem.	-0.047	18.470	12.710	0.013	0.062
	(0.125)	(26.220)	(28.850)	(0.104)	(0.278)
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track					
The differential effect	-0.65	2.59	1.79	0.18	0.87
Women	-2.00	0.92	2.80	4.2	1.72
Men	-1.35	-1.66	1.02	4.0	0.86
Observations	1,902	1,416	1,416	1,863	1,902

Notes: Each column is a separate regression and also includes year fixed effects. On-track students are those enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB between the years 1997 and 2007. This sample excludes two small classes with enrollment less than 100. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. Level of significance is indicated as follows: ** p<0.01, * p<0.05. + p<0.1. Robust standard errors are in parentheses.

Table 6: Summary Statistics (sample includes on-tracks only)

	Women (1)	Men (2)
Classroom Characteristics		
% on-track in a lecture	0.85	0.85
	(0.05)	(0.05)
CHEM1A lecture size	329.37	328.53
	(46.73)	(47.37)
Student Background Characteristics		
URM (underrepresented minority)	0.34	0.31
	(0.47)	(0.46)
High school grade point average	3.80	3.70
	(0.31)	(0.33)
SAT math score	587.59	636.41
	(78.16)	(75.49)
SAT verbal score	564.88	573.97
	(82.63)	(85.60)
Attended public high school	0.86	0.84
	(0.35)	(0.37)
English is only language spoken in home	0.69	0.65
	(0.46)	(0.48)
No parent graduated from college	0.36	0.29
	(0.48)	(0.46)
Outcomes		
Graduate with STEM major	0.45	0.53
	(0.50)	(0.50)
Graduate	0.82	0.80
	(0.39)	(0.40)
CHEM1A grade	2.49	2.65
	(0.95)	(0.91)
Took follow-on course (CHEM1B)	0.80	0.85
	(0.40)	(0.36)
Grade in follow-on course (CHEM1B)	2.58	2.59
	(0.87)	(0.87)
Observations	5,942	6,288

Notes: The sample includes only on-track students, those enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB between the years 1997 and 2007. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. Standard deviations are reported in parentheses.

Table 7: The Effect of the Number of On-Track Students on STEM Major Completion

	Full Sample	Lectures > 100	Bottom 1/3 SAT Math	Middle 1/3 SAT Math	Top 1/3 SAT Math
	(1)	(2)	(3)	(4)	(5)
Panel A: Point Estimates					
Ln(no. of on-track)	-0.112	-0.037	-0.232	-0.71	-0.057
	(0.106)	(0.103)	(0.171)	(0.299)	(0.138)
Ln(no. of on-track) X fem.	-0.112**	-0.137**	-0.166*	0.010	-0.055
	(0.030)	(0.045)	(0.067)	(0.061)	(0.077)
Instructor, year, time of day FE	X	X	X	X	X
Student & class controls	X	X	X	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track					
Women	-3.14*	-2.40+	-5.70*	-0.90	-1.60
Men	-1.57	-0.50	-3.20	-1.00	-0.80
Observations	12,230	12,122	4,206	3,438	4,586

Note: Each column is a separate specification. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. The student's intended major at entry is coded into five categories: hard sciences, Biology/Environmental Science, Social Sciences, Humanities/Arts/interdiciplinary, and undeclared. Clustered standard errors are in parentheses , ** p<0.01, * p<0.05, + p<0.1. Clusters are by CHEM1A section. A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table 8: Heterogeneity Analysis – STEM Major Completion for Various Subgroups

	URM Effect?	Low Ability Effect?	Low SES Effect?
	(1)	(2)	(3)
Panel A: Point Estimates			
Ln(no. of on-track)	-0.096	-0.116	-0.143
	(0.114)	(0.106)	(0.119)
Ln(no. of on-track) X fem.	-0.125*	0.802+	-0.087
	(0.060)	(0.472)	(0.070)
Ln(no. of on-track) X URM	-0.050		
	(0.070)		
Ln(no. of on-track) X fem. X URM	0.033		
	(0.101)		
Ln(no. of on-track)X bottom 1/3		0.013	
		(0.064)	
Ln(no. of on-track) X fem. X bottom 1/3		-0.144+	
		(0.084)	
Ln(no. of on-track) X parent col. grad			0.054
			(0.063)
Ln(no. of on-track) X fem. X parent col. grad			-0.042
			(0.087)
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track			
Non-URM – women	-3.10*		
Non-URM – men	-1.30		
URM – women	-3.30*		
URM – men	-2.00		
Bottom 1/3 – women		-4.00*	
Bottom 1/3 – men		-1.40	
Top 2/3 – women		-2.00	
Top 2/3 – men		-1.60	
Col. grad parent – women			-3.00*
Col. grad parent – men			-1.20
No col. grad parent – women			-3.20*
No col. grad parent – men			-2.10
Observations	12,230	12,230	12,230

Note: Each column is a separate specification. The Column 1 regression also includes a dummy for URM and an interaction term between URM and woman. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. The Column 2 regression also includes a dummy for being in the bottom 1/3 of the SAT math distribution and the interaction between being in the bottom and a woman. The Column 3 regression also includes a dummy for having at least one parent with a college degree and the interaction between being that dummy and woman. Additionally, all three regressions include controls for percent female in a class, class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , ** p<0.01, * p<0.05, + p<0.1. Clusters are by CHEM1A section. A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students).

Table 9: STEM Major Completion – Controls for CHEM1A Grade

	Full Sample	Bottom 1/3 SAT Math	Middle 1/3 SAT Math	Top 1/3 SAT Math
	(1)	(2)	(3)	(4)
Panel A: Point Estimates				
Ln(no. of on-track)	-0.256+	-0.467**	-0.231	-0.184
	(0.140)	(0.171)	(0.302)	(0.134)
Ln(no. of on-track) X fem.	-0.117**	-0.152*	0.0267	-0.0754
	(0.034)	(0.061)	(0.055)	(0.079)
CHEM1A Grade	0.169**	0.171**	0.178**	0.167**
	(0.005)	(0.009)	(0.009)	(0.009)
Instructor, year, time of day FE	X	X	X	X
Student & class controls	X	X	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track				
Women	-5.20**	-8.67**	-2.85	-3.64+
Men	-3.60+	-6.53**	-3.23	-2.58
Observations	12,230	4,206	3,438	4,586

Note: Each column is a separate specification. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , ** p<0.01, * p<0.05, + p<0.1. Clusters are by CHEM1A section. A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table 10: Where are the Women Going?

	Graduate in Non-STEM	Graduate
	(1)	(2)
Panel A: Point Estimates		
Ln(no. of on-track)	1.22	-0.035
	(0.099)	(0.113)
Ln(no. of on-track) X fem.	0.107**	-0.074+
	(0.030)	(0.038)
Instructor, year, time of day FE	X	X
Student & class controls	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track		
Women	3.20*	-0.50
Men	1.70	-1.50
Observations	12,230	12,230

Note: Each column is a separate specification. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , ** p<0.01, * p<0.05, + p<0.1. Clusters are by CHEM1A section. A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Appendix

Table A1: STEM Majors

Major Requires CHEM1A	STEM Majors
X	Biology
X	Biochemistry
X	Biopsychology
X	Chemistry
X	Engineering
X	Computer Engineering
	Computer Science
X	Earth Science
X	Ecology
X	Environmental Science
	Mathematics
	Statistics
X	Geophysics
X	Hydrology
X	Zoology
X	Pharmacology
X	Physics
X	Physiology
X	Physical Geography

Table A2: Balance Test – Orientation Attendees vs. Non-Attendees

	Summer Orientation Group	Did Not Attend Orientation	Diff. (1) - (2)
	(1)	(2)	(3)
Student Background and Class Characteristics			
No. On-track is above median	0.48	0.46	0.02
	(0.01)	(0.04)	(0.04)
Lecture is at 8 or 9 a.m.	0.35	0.39	-0.04
	(0.01)	(0.04)	(0.04)
Female	0.49	0.52	-0.03
	(0.01)	(0.04)	(0.04)
URM (underrepresented minority)	0.38	0.48	-0.10*
	(0.01)	(0.04)	(0.04)
High school GPA is above median	0.51	0.46	-0.05
	(0.01)	(0.04)	(0.04)
SAT math score is above median	0.51	0.51	0.00
	(0.01)	(0.04)	(0.04)
SAT verbal score is above median	0.54	0.48	0.06
	(0.01)	(0.04)	(0.04)
Attended public high school	0.92	0.93	-0.01
	(0.01)	(0.02)	(0.02)
English is only language spoken in home	0.54	0.48	0.06
	(0.01)	(0.04)	(0.04)
No parent graduated from college	0.37	0.43	-0.05
	(0.01)	(0.04)	(0.04)
Observations	1,624	178	1,802

Note: URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. Column 3 uses an asterisk system to denote whether the differences in means are significant. Level of significance is indicated as follows: ** p<0.01, * p<0.05, + p<0.1. Standard deviations are in parentheses. The sample includes only those students enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB for the year 2013.

Table A3: The Effect of the Number of On-Track Students on STEM Major Completion Including Various Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Point Estimates								
Ln(no. of on-track)	-0.262*	-0.347*	-0.211*	-0.112	-0.283*	-0.183+	-0.126	-0.031
	(0.128)	(0.138)	(0.102)	(0.106)	(0.120)	(0.101)	(0.105)	(0.105)
Ln(no. of on-track) X fem.	-0.085*	-0.098**	-0.103**	-0.112**	-0.114**	-0.113**	-0.075*	-0.152**
	(0.034))	(0.029)	(0.029)	(0.030)	(0.030)	(0.030)	(0.034)	(0.035)
Year FE	X	X	X	X		X	X	X
Instructor FE		X	X	X				X
Time of day FE			X	X	X		X	
Student & class controls				X	X	X	X	X
Instructor X year FE					X			
Instructor X time of day FE						X		
Instructor X student gender							X	
Time of Day X student gender								X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track								
Women	-4.85**	-6.23**	-4.40**	-3.14*	-5.56**	-4.14**	-2.81+	-2.58+
Men	-3.66*	-4.86**	-2.95+	-1.57	-3.96**	-2.56+	-1.76	-0.44
Observations	12,230	12,230	12,230	12,230	12,230	12,230	12,230	12,230

Note: Each column is a separate specification. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , ** $p<0.01$, * $p<0.05$, + $p<0.1$. Clusters are by CHEM1A section. A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table A4: The Effect of Percent On-Track in a Class on STEM Major Completion

	Full Sample	Lectures > 100	Bottom 1/3 SAT Math	Middle 1/3 SAT Math	Top 1/3 SAT Math
	(1)	(2)	(3)	(4)	(5)
Panel A: Point Estimates					
Percent On-Track	-0.050	-0.030	-0.204	0.188	-0.181
	(0.163)	(0.171)	(0.287)	(0.263)	(0.279)
Percent On-Track X fem.	-0.272	-0.232	-0.291	-0.572+	0.353
	(0.189)	(0.199)	(0.333)	(0.299)	(0.322)
Instructor, year, time of day FE	X	X	X	X	X
Student & class controls	X	X	X	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in share of on-track					
Women	-4.83*	-3.93+	-7.43*	-5.75	2.58
Men	-0.75	-0.44	-3.07	2.83	-2.80
Observations	12,230	12,122	4,206	3,438	4,586

Note: Each column is a separate specification. Controls include percent female in a class, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , ** p<0.01, * p<0.05, + p<0.1. Clusters are by CHEM1A section. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table A5: The Effect of the Number of On-Track Students on STEM Major Completion for Various Groups

	Late-Track	Late-Track & On-Track	Late-Track & On-Track		
			Bottom 1/3	Middle 1/3	Top 1/3
	(1)	(2)	(3)	(4)	(5)
Panel A: Point Estimates					
Ln(no. of on-track)	-0.279	-0.143	-0.206	-0.135	-0.081
	(0.190)	(0.088)	(0.152)	(0.291)	(0.133)
Ln(no. of on-track) X fem.	-0.150+	-0.117**	-0.162**	-0.013	-0.115+
	(0.083)	(0.031)	(0.048)	(0.076)	(0.065)
Instructor, year, time of day FE	X	X	X	X	X
Student & class controls	X	X	X	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track					
Women	-6.00*	-3.64**	-5.15*	-2.07	-2.74
Men	-3.91	-2.00	-2.89	-1.90	-1.13
Observations	1,935	14,165	4,770	3,389	5,556

Note: Each column is a separate specification. “Bottom 1/3” refers to the bottom 1/3 of the SAT math distribution. “Middle 1/3” and “Top 1/3” refer to the middle 1/3 and top 1/3 of the SAT math distribution for the class respectively. Columns 3-5 include both on-track and late-track students. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student’s declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , ** p<0.01, * p<0.05, + p<0.1. Clusters are by CHEM1A section. A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table A6: Major Composition by Gender (%)

	Women		Men	
	Intended Major at Entry	Major at Grad.	Intended Major at Entry	Major at Grad.
Hard science	15.07	14.29	49.37	33.77
Bio/environ. sci.	51.81	29.88	24.84	17.45
Social science	3.48	19.27	2.79	15.79
Human./arts/interd.	3.08	36.57	2.55	32.97
Undeclared	26.55	0.00	20.45	0.01

Note: See Appendix Table A7 for the majors that fall into each major category: hard science, biology/environmental studies, social sciences, and humanities/arts/interdisciplinary. The sample includes only those students enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB between the years 1997 and 2007.

Table A7: Majors by Category

Hard Sci.	**Bio & Env. Studies**	**Social Sci.**	**Human., Arts, & Interdis.**
Biochemistry	Biochemistry-Molecular Biology	Anthropology	Art History
Chemistry	Biological Sciences	Geography	Art Studio
Chemical Engineering	Biopsychology	Geophysics	Asian & American Studies
Computer Engineering	Physiology	Physical Geography	Asian Studies
Electrical Engineering	Biology	Economics-Accounting	Black Studies
Earth Science	Cell & Develp. Biology	Economics-Mathematics	Chicana and Chicano Studies
Hydrological Sciences	Microbiology	Economics	Chinese
Mechanical Engineering	Environmental Studies	Political Science	Classics
Pharmacology		Psychology	Communication Studies
Physics		Sociology	Comparative Literature
Computer Science		Business Economics	Creative Studies
Mathematics			Dance
Financial Math & Stats			Dramatic Art
Statistics			English
Zoology			Feminist Studies
Aquatic Biology			Film & Media Studies
Ecology and Evolution			Financial Mathematics & Statistics
Computer Science			French
			Germanic Languages
			Global Studies
			History or History of Public Policy
			Interdisciplinary Studies
			Italian Cultural Studies
			Japanese
			Language, Culture & Society
			Latin Am/Iberian Studies
			Law & Society
			Linguistics
			Medieval Studies
			Middle Eastern Studies
			Music & Music Composition
			Philosophy
			Portuguese
			Religious Studies
			Slavic Languages & Literatures
			Spanish