

ROLE MODEL EFFECTS OF FEMALE STEM TEACHERS AND DOCTORS ON
EARLY 20TH CENTURY UNIVERSITY ENROLLMENT IN CALIFORNIA

December 2016

Zach Bleemer
UC Berkeley

Copyright 2016 Zach Bleemer, all rights reserved.

ABSTRACT

What was the role of imperfect local information in the growth, gender gap, and STEM (Science, Technology, Engineering and Math) major selection of early 20th century American universities? In order to examine pre-1950 American higher education, this study constructs four rich panel datasets covering most students, high school teachers, and doctors in the state of California between 1893 and 1946 using recently-digitized administrative and commercial directories. Students attending large California universities came from more than 600 California towns by 1910, with substantial geographic heterogeneity in female participation and STEM major selection. About 43 percent of university students in 1900 were women, and the number of women attending these universities increased by more than 500 percent between 1900 and 1940. Meanwhile, the number of California towns with female high school physics or chemistry teachers doubled between 1903 and 1923, while the proportion of towns with a female doctor increased from 20 to 26 percent (adding almost 60 towns) during the same period. Event study regression analysis shows that towns became 9-15 percentage points more likely to send at least one female student to the institutions examined in this study after the arrival of their first female high school physics or chemistry teacher or female doctor, implying a 2 percentage point increase in the likelihood of young women's college attendance, but that the arrival of female STEM teachers decreases the likelihood of a town's sending a male STEM student to university by 10 percentage points. This study establishes the role of limited information and social networks in early 20th century educational choices, and has implications for both historical growth accounting and contemporary educational practices in developing economies. It also provides a window into the tremendous socioeconomic mobility afforded by California's commitment to mass higher education. This is the first of several planned studies that are part of the new UC Cliometric History Project based at CSHE in anticipation of UC's 150th anniversary.

Keywords: Education History, California Universities, College Enrollment, Major Selection

*Children may not obey, but children will listen
Children will look to you for which way to turn
To learn what to be
Careful before you say "Listen to me"
Children will listen
~ Stephen Sondheim, Into the Woods ~*

*Jenny: 'Studying is hard and boring. Teaching is hard and boring. So you're telling me to be bored, and then bored,
and then finally bored again, this time for the rest of my life ... It's not enough to educate us any more, Mrs. Walters.
You've got to tell us why you're doing it.'
~ Nick Hornby, An Education (directed by Lone Scherfig) ~*

Between the 1900 and 1950 birth cohorts, Americans' average educational attainment increased from grade 8 to grade 12, and the proportion of both men and women who attended college more than quadrupled. How did this massive expansion of secondary and post-secondary education shape growth, gender roles, and technological advancement in the United States? Building on a large empirical literature studying the determinants of college attendance, the university gender gap, and the selection of STEM (science, technology, engineering, and math) fields in the United States¹, this study examines the function of

* Thanks to David Card, Brad DeLong, John Douglass, Barry Eichengreen, Claudia Goldin, Jasjeet Sekhon, Christopher Walters, and Basit Zafar as well as seminar participants at the All-California Labor Economics Conference, the UC Berkeley Economic History Lunch, the UC Berkeley Graduate Student Summer Seminar, and the Center for Studies in Higher Education Seminar for helpful comments. Thanks as well to Renata Ewing, Lynne Grigsby, Mary Elings, the California Digital Library, the HathiTrust Digital Library, and the UC Berkeley Bancroft Library for aiding

role models and information networks in expanding college attendance and STEM major selection in the early 20th century, especially among women.

The early 20th century is an ideal setting in which to study the upper-bound causal impact of role models on young person decision-making; educational systems were highly consolidated (few towns had more than one high school, and potential university students had few schools to choose between) and social networks were small and geographically concentrated. Until commercial radio broadcasting began after World War One, information flows outside of major metropolitan areas were largely limited to newspapers, magazines, pamphlets, and books privately sold or held in public libraries, and contemporary library records suggest that even medium-size cities held no more than a few thousand volumes in their collections.² However, while previous research has found substantial individual and macroeconomic benefits of early 20th century secondary education (Goldin and Katz, 2008) and large contemporary returns to higher education (Autor, 2014) and STEM degrees in particular (Altonji, Arcidiacono, and Maurel, 2016), little is known about post-secondary enrollment, major selection, or graduation in the early 20th century, let alone the determinants of that schooling.³

This study examines California as a case study of early 20th century American higher education, presenting four novel panel datasets covering most university students, high school teachers, and doctors in the state throughout that period.⁴ Higher education in the early 20th century California was open, polarized, and variegated. University admission was relatively non-competitive for qualified California high school graduates, implying that college enrollment trends corresponded to student demand.⁵ The University of California (UC), which enrolled more than half of California university students throughout the period, charged no tuition and placed no gender restrictions on enrollment or field of study.⁶

Nevertheless, few women studied outside the College of Letters and Sciences at UC (11 percent in 1920, mostly in commerce and pre-medicine); indeed, in 1910 UC's President Wheeler "thought that women should be trained primarily to carry out their special vocation as wives, mothers, and household managers," according to Henry May in his study of the Berkeley campus (May, 1993), and the school's first Dean of Women wrote in her autobiography that in 1906, "most of the faculty thought of women frankly as inferior beings" (Gordon, 1990).⁷ UC served students through multiple swiftly-evolving roles—as "the democratic and utilitarian people's university, ... the stronghold of polite traditional culture, ... and the center of high-powered and specialized research" (May, 1993)—making it the largest university in the country (and the world) in the 1920s and thus an ideal case study of the era's multifaceted university systems (Ferrier 1930, p. 537).

When a town hires its first female STEM-oriented professional, young female residents are provided with both the role model of a possible personal future and a valuable (perhaps unique) source of information about women's university experience and labor market outcomes. Previous studies of role model and information effects on college attendance and major selection have focused on contemporary observational and quasi-experimental evidence of small increases in college attendance (Nixon and Robinson, 1999; Bleemer and Zafar, 2015) and STEM major selection (Wiswall and Zafar, 2015) resulting from information interventions. Bettinger and Long (2005) find no effect of quasi-randomly-assigned female physics or chemistry first-year professors on STEM major selection at public Ohio universities in 1998-2000, and Carrell, Page, and West (2010) find the same for randomly-assigned female STEM professors at the US Air Force Academy. Dee (2006) proposes an alternative mechanism, presenting observational data suggesting that female high school students perform slightly better in history courses taught by women (though he finds no effect for female science teachers); although Hoffman and Oreopoulos (2009) find the achievement effect to be very small, its presence implies that examination of the effect of towns' first female doctors may provide a purer estimate of role model effects.

the digitization effort that produced the data used in this paper. This paper was awarded the UC Berkeley INET Prize in Economic History. I dedicate this study to Miriam Miller, Wellesley College '46. Any errors that remain are my own.

† UC Berkeley, Department of Economics; E-mail: bleemer@berkeley.edu.

¹ See, e.g., Attendance: Bound, Lovenheim, and Turner (2010); Lochner and Monge-Naranjo (2012); Bleemer and Zafar (2015); Gender Gap: Goldin (1997); Goldin, Katz, and Kuziemko (2006); and STEM: Wiswall and Zafar (2015); Altonji, Kahn, and Speer (2014).

² For example, the public libraries of Reedley and Willows, both above-median-population incorporated California towns in 1920, report holding between 2,000 and 3,000 volumes in that year, along with a few dozen magazines and 1-2 newspaper subscriptions (to say nothing of unincorporated communities, which made up more than 30 percent of California's population at the time; the libraries of such communities typically reported collections of fewer than 100 volumes). See the News Notes of California Libraries (1906-1972).

³ Indicative of the limited nature of pre-1930 data on college attendance, Goldin and Katz (2000) uses survey data from Iowa to estimate the 1915 return of a year of university education as 13 percent, but Feigenbaum (2015) shows that by 1940, Iowa had substantially above-average levels of college returns (Table A.3).

⁴ Three universities are examined: The (public) University of California system, Stanford University, and the University of Southern California. These schools enrolled 70-90 percent of CA university undergraduates throughout the period of interest. Further information is provided below.

⁵ Most, but not all, California public and private high schools were accredited by California universities, providing post-secondary admission for all 'recommended' students as well as all students who pass a matriculation exam, though there were exceptions; Stanford, for example, capped its number of female students. See Douglass (2007).

⁶ Through the 1920s, UC only charged a semesterly 'incidental fee' of 25 dollars. The school estimated that room and board cost 25-75 dollars per month, but students could work in town or for the university to cover this cost; see Moreman (2006) for a typical account.

⁷ In the suggestive phrasing of Harry M. Shafer, Hanford High School Principal and keynote speaker at the 1916 State Convention of California High School Principals, "Boys and girls differ so materially, both as sexes and as individuals, that no one subject, with the possible exception of the mother tongue, is essential to each person" (Wood, 1917).

Importantly, early-20th-century young Californians' smaller social networks and lower-quality information about college returns and labor market prospects suggest a far larger potential impact of female role models on college-going behavior than would be expected today. These characteristics might be shared by youth in contemporary developing economies; Muralidharan and Sheth (2015), for instance, find a larger achievement effect on female student performance in contemporary rural India than has been estimated in the United States.

Between 1900 and 1930, the number of students at California high schools and universities more than quadrupled. About 45 percent of California public and large private university students in 1900 were women, a proportion rarely matched (excluding the World Wars) until the late 20th century.⁸ While men and women were similarly-likely to graduate UC (conditional on matriculation) until the Great Depression, during which male graduation rates substantially exceeded female rates, men were far more likely to study STEM fields and commerce (or business) than women; almost half of male college students between 1900 and 1908 studied engineering, chemistry, or pre-medicine, compared to about 2.5 percent of women. STEM participation swiftly eroded in public (and large private) universities after 1908, with the proportion of men studying STEM dropping from 48 (34) percent in 1908 to 34 (16) percent in 1916 and 26 percent in 1930 (though there was no similar decline in female STEM participation), and has never recovered.⁹ California's rural population was consistently underrepresented at universities, but between 16 and 24 percent of students were from unincorporated or below-median-population towns between 1900 and 1920, and rural student representation shrunk at a far slower rate than the state's broader rural population through the 1930s.

The Progressive Era (roughly 1900 to 1920) was a period of significant social and political reform. California women obtained suffrage and political representation (four women were elected to the California Assembly in 1918), and the number of California towns with female high school physics or chemistry ("PhyChem") teachers or female doctors, the two most popular STEM-oriented occupations for women, slowly increased absolutely and proportionally throughout the period.¹⁰ Between 1903 and 1923, the number of towns with high schools more than doubled—from 123 to 291 while the number of towns with at least one doctor increased from 420 to 513. Meanwhile, the proportion of high-school towns with female PhyChem teachers increased from 22 percent to a peak of 38 percent (compared to 59 percent for all sciences and 70 percent for math), while the proportion with female doctors grew from 18 to 23 percent.

How did this expansion change the college-going behavior of young Californians? I use a difference-in-difference event study framework to examine the effect of a town's hiring its first female PhyChem teacher or female doctor on college-going behavior in that town over the subsequent ten years. In nearly all cases, the towns are statistically balanced across all measured outcomes for ten years preceding the event, suggesting exchangeability between the towns that did and did not hire female professionals in each year.¹¹

I find that a town's hiring its first female PhyChem teacher immediately increases the likelihood of that town's sending at least one woman to a public or large private university each year by 12.3 percentage points, with the gender ratio declining by more than 10 percentage points for both public and private universities. A simple calculation suggests that these findings imply an increase in the likelihood of female college attendance from 44 percent to 60 percent conditional on high school graduation, an additional one female college matriculant per year from a median-sized treated school (a 2-3 pp. increase in college attendance likelihood for the broader population of young women).

The effects are moderately persistent; the evidence suggests a medium-term (5 year) increase in the likelihood of female college participation of 5 percentage points. I do not find evidence of an increased propensity for men or women to study STEM fields or eventually practice medicine, but find that male STEM participation may decline in the short-term. Finally, I find that a town's hiring its first female doctor immediately increases both male and female college-going (about 8 percentage points). All of these findings are robust to a number of alternative specifications and controls, including town-level time trends.

Section 2 briefly details the data collected in this study, and Section 3 uses those data to describe California higher education in the early 20th century. Section 4 presents and estimates event study models and briefly considers robustness checks. Section 5 concludes.

A. Data

Previous studies of pre-1930 higher education in the United States have almost exclusively used data from post-1940 US Censuses.¹² Before 1940, the US Census asked no questions about individuals' education; starting in that year, the Census asked respondents for the "highest grade of school completed", with responses ranging from 0 ("None") to 17 ("College, 5th or

⁸ While about 45 percent of all post-secondary students were women in 1900 (Goldin, Katz, and Kuziemko, 2006), the proportion of female university students (excluding normal schools and junior colleges) was below 40 percent (Digest of Education Statistics, Table 301.20).

⁹ See Bleemer, Wiswall, and Zafar (2015) for national STEM major trends since the 1940 birth cohort.

¹⁰ I focus on physics and chemistry teachers because they were the high school laboratory courses most valued by California universities (Wood, 1917, p. 22), because they were empirically the least-likely fields for women to teach, and because those fields are the most traditionally STEM-oriented (relative to, say, botany or physiology).

¹¹ I discuss evidence for the plausible exogeneity of the arrival of female STEM professionals below.

¹² See, e.g., Smith (1984), Mare (1991), and Goldin, Katz, and Kuziemko (2006). Exceptions include Goldin and Katz (2000), which studies the 1915 Iowa Census, and Goldin and Katz (1999b), which uses aggregate (university-level) administrative data from the US Department of Education.

subsequent year”). Analysis of Census data, then, is highly restrictive. It provides no tractable way to identify individuals’ childhood towns or pre-education characteristics, prohibiting non-persistent covariate analysis (like the rural/urban divide or economic mobility). Census data does not distinguish between the partial completion of a Bachelor’s degree and partial or full completion of a two-year junior college degree, which might have importantly-different returns and implications. It cannot identify field of study, and analysis of the early 20th century is biased by differential mortality, international immigration and emigration, and misreporting. In short, there is ample reason to search for a higher-quality record of early 20th century higher education in the United States.

I collect four new comprehensive individual-level administrative and commercial datasets covering early 20th-century California:

1. **Public University Data:** An individual-year panel of all undergraduate students who attended a four-year public university in California—including University of California (UC) campuses at Berkeley, Los Angeles, San Francisco, and Davis—between 1893 and 1946.¹³ The data include name, degree program, year of study, and home town. Junior and Teacher Colleges are omitted.
2. **Private University Data:** An individual-year panel of all undergraduate students who attended a large four-year private university in California (where large is defined as having at least 1,000 students before 1940), including Stanford University (1893–1946) and the University of Southern California (USC; only available 1905–1920).¹⁴ These schools enrolled more than 60 percent of private university four-year undergraduates in California through the 1910s and 1920s.¹⁵ The data include name, major, year of study, and home town.
3. **High School Teacher Data:** An annual individual-year panel of all high school teachers in California between 1907 and 1924. The data include name, school and town, subjects taught, degrees held, and universities attended.
4. **Doctor Data:** An individual-year panel of all doctors practicing in California between 1903 and the present. Data include name, town (of practice), degrees held, and universities attended.

These data were collected from annual registers and directories published by each university, the state of California, a textbook-publishing firm, and California’s state-wide Medical Society and Medical Board.¹⁶ Each document was digitized in three stages. The first stage, largely conducted by partnerships between Google and several American universities, produced page-by-page images that were made publicly available through the HathiTrust Digital Library.¹⁷ The second stage, in which the book images were converted to text, was conducted partly by Google (again available through HathiTrust) and partly by the author using proprietary OCR software. The third stage uses algorithmic corrections to organize the variously-formatted text into uniformly-structured data.

Next, I identify each individual’s gender and self-reported home town.¹⁸ I infer gender by matching individuals’ first names with Social Security Administration records, which include all names assigned to at least five children of one gender for each year since 1880.¹⁹ Spelling errors and name changes challenge town identification; I match towns to a comprehensive list of populated areas compiled from Wikipedia (along with the names of other states and nations to identify out-of-state university students), allowing for small spelling changes and frequently-occurring errors.²⁰ Each town is matched to geographic coordinates (using MediaWiki’s GeoHack database), which are then matched to decennial counties (allowing for changing borders over time)

¹³ More than 50 percent of university students in California attended a public university throughout this period. UC Berkeley enrolled undergraduate students throughout this period. UCLA transitioned from a junior college to a four-year university in 1922, so undergraduates are included from Fall 1921; UCLA Teachers College students are omitted (since they were two-year college students). The UC School of Dentistry in San Francisco became a four-year degree-granting program in Fall 1917, adding the School of Pharmacy in 1934 and the School of Nursing in 1939. UC Davis became a four-year degree-granting university (with degrees exclusively in Agriculture) in Fall 1922. California Polytechnic State University, which started awarding four-year degrees in 1942, is omitted.

¹⁴ As late as the early 1930s, only these two universities had more than 1,000 students, and still accounted for more than 70 percent of private university students (Ferrier, 1937, p. 366-367); by 1940, they accounted for 40 percent of private students. Leland Stanford Junior University began accepting four-year undergraduate students in 1891, but the number of female students was capped at 500 (about 25 percent) until 1932. USC was founded in 1880 but had a negligible number of students (fewer than 100) until 1905. USC records are censored after 1920 due to copyright restrictions, and are thus omitted from the empirical analysis (but not the descriptive statistics) below.

¹⁵ See US Bureau of Education (1925).

¹⁶ Details about the data cleaning process are available in the Appendix.

¹⁷ A few otherwise-unavailable volumes were digitized by me or by the UC Berkeley Bancroft Library. The following volumes are presently omitted: the 1945 UC register, the 1922 teacher register and the 1909 and 1915 doctor registers.

¹⁸ According to the UC register, home town is the town in which the individual’s most recent residence was located. Every individual must report a town, including rural students.

¹⁹ These records include more than 2,000 names for each gender in each year. I begin by matching students to SSA records from 20 years earlier and both teachers and doctors to records 30 years earlier (with a floor at 1880, the first year in which the records are available), and then continue matching using subsequent and previous years. I omit names that are less than 10 times more likely for one gender. Data available at <https://www.ssa.gov/oact/babynames/limits.html>.

²⁰ In particular, a match is successful if the recorded town name is no more than one generalized Levenshtein distance away from the true town name, omitting spaces (see Levenshtein (1966)). Wikipedia is the most comprehensive source of early 20th century California towns, with about 4,800 listed incorporated cities, unincorporated communities, Census-designated places, former Census-designated places, former populated places, and current neighborhoods of Los Angeles and San Diego (many of which were formerly independent towns).

using data from the National Historical Geographic Information System.²¹

I similarly identify students' majors and degree programs and teachers' taught subjects. About 11.4 percent of students are not from California, and are omitted from my analysis. Overall, I cannot identify gender for 4.4 percent of individuals and cannot identify the home towns of another 2.1 percent; those individuals are also omitted. After omissions, I am left with 472,611 public university student-year observations, 133,088 private university observations, 67,329 high school teacher observations, and 253,275 doctor observations. Table 1 presents summary statistics for each data series.

Table 1: Summary Statistics of California Undergraduate Students, High School Teachers, and Doctors between 1895 and 1946

Population	Number of Individuals					Number of Represented Towns					Percent Male							
	1895	1905	1915	1925	1935	1946	1895	1905	1915	1925	1935	1946	1895	1905	1915	1925	1935	1946
UC Berkeley	1122	2249	4614	7761	11052	7671	131	243	425	531	565	455	60.4	57.1	57.2	54.8	61.4	39.6
Lib. Arts	852	1311	2844	5187	6842	6125	114	176	338	440	442	398	49.8	28.7	34.6	35.8	44.6	31.3
STEM	257	724	841	1271	2207	911	54	142	180	225	297	191	94.1	97.7	95.8	95.1	90.4	88.9
Prof.	13	214	912	1274	1997	604	8	61	203	246	296	147	83.3	94.1	91.3	90.6	86.6	50.4
UCLA	.	.	.	3192	4994	4140	.	.	.	212	207	205	.	.	.	52.5	58.8	30.2
Lib. Arts	.	.	.	2453	4421	3278	.	.	.	179	193	188	.	.	.	42.0	54.4	20.7
STEM	.	.	.	537	482	403	.	.	.	92	65	63	.	.	.	89.6	92.2	81.8
Prof.	.	.	.	202	91	459	.	.	.	50	41	69	.	.	.	80.8	97.8	51.9
Other Public Univ.	.	.	.	469	335	501	.	.	.	145	137	152	.	.	.	94.6	92.6	44.9
UC Davis	.	.	.	134	202	126	.	.	.	74	110	71	.	.	.	93.2	89.6	66.4
San Fran.	.	.	.	335	133	375	.	.	.	95	51	115	.	.	.	95.1	96.9	37.3
Stanford Univ.	704	1348	1376	2441	2184	4049	128	182	199	289	264	395	60.6	65.5	74.4	80.2	67.6	69.8
Lib. Arts	553	927	955	1038	1037	.	107	149	168	184	169	.	55.2	51.0	67.5	75.3	64.1	.
STEM	110	401	394	559	361	.	51	98	98	124	93	.	96.3	97.7	92.7	89.1	87.4	.
Pub. HS Teachers	.	1157	3587	7769	.	.	.	145	238	280	.	.	.	44.3	47.5	42.8	.	.
Math	.	304	608	886	.	.	.	140	237	278	.	.	.	55.0	57.5	48.2	.	.
All Sci.	.	234	520	869	.	.	.	121	232	275	.	.	.	59.6	73.7	61.7	.	.
Physchem	.	197	376	492	.	.	.	114	226	263	.	.	.	64.9	80.7	75.1	.	.
Doctors	.	3701	4577	6775	8506	8295	.	409	537	498	554	480	.	90.9	89.1	91.9	90.7	89.9

Notes: This table displays population statistics for California public university and Stanford Univ. students (conditional on being California residents), public high school teachers, and registered doctors. Liberal Arts are defined as students studying Letters, Social Sciences, Natural Sciences, and Pre-Law; STEM as studying Engineering, Chemistry, or Pre-Medicine; and Professional as studying Agriculture, Commerce or Business, Pre-Architecture, Pre-Dental, Pharmacy, or Nursing. All students at UC Davis and the UC San Francisco campus study professional fields. For high school teachers, math includes algebra, geometry, and trigonometry, while science includes physics and chemistry (PhysChem) as well as general science, zoology, botany, physiology, and general science. Doctors include physicians, surgeons, a small number of osteopaths, and (in rare cases) alternative medical practitioners. The following years of data have been replaced (see the Appendix): 1905 displays 1904 UC Berkeley data, 1915 displays 1916 Doctors data, 1925 displays 1923 HS Teachers data, and 1935 displays 1933 Doctors data. UCLA, Davis, and the San Francisco campus opened for four-year degrees in the early 1920s; Stanford first- and second-year students do not declare majors after 1942, prohibiting 1946 major assignment. By definition, all represented towns are in California. Gender is determined by matching first names to the most popular names of males and females assigned at birth in the United States around 20-30 years earlier (according to the Social Security Administration); About 7 percent of students cannot be assigned hometowns or genders, largely due to uncommon or androgynous first names, uncommon hometowns, and imperfect data cleaning.

Primary Sources: The University of California Register (1893-1946), Stanford University Annual Register (1893-1946), Heath's Directory of Secondary Schools (1907-1914), California Directory of Secondary and Normal Schools (1915-1924), and the Medical Society of California's Official Registry and Directory of Physicians and Surgeons (1903-1946).

²¹ GeoHack documentation is available from <https://www.mediawiki.org/wiki/GeoHack>; NHGIS data is available at <https://data2.nhgis.org/main>.

Figure 1: Geographic and Gender Heterogeneity of Large Four-Year California University Students, 1900-1930



Note: Circle size log proportional to the number of university students. Includes all towns with at least five students attending a large four-year university in the given year. Large universities are defined as four-year tertiary schools with more than 1,000 students in 1920 (omitting private schools in 1930 due to data unavailability), and include the University of California system (Berkeley 1900-1930, UCLA 1920-1930, Davis 1930), Stanford University (1900-1930), and the University of Southern California (1910-1920). Large universities enrolled all public university students and about half of private university students throughout the period. Student genders are determined by matching first names to the most popular names of males and females assigned at birth in the United States around 20 years before each student attends university (according to the Social Security Administration). About 7 percent of students cannot be assigned hometowns or genders, largely due to uncommon or androgynous first names, uncommon hometowns, and imperfect data cleaning.

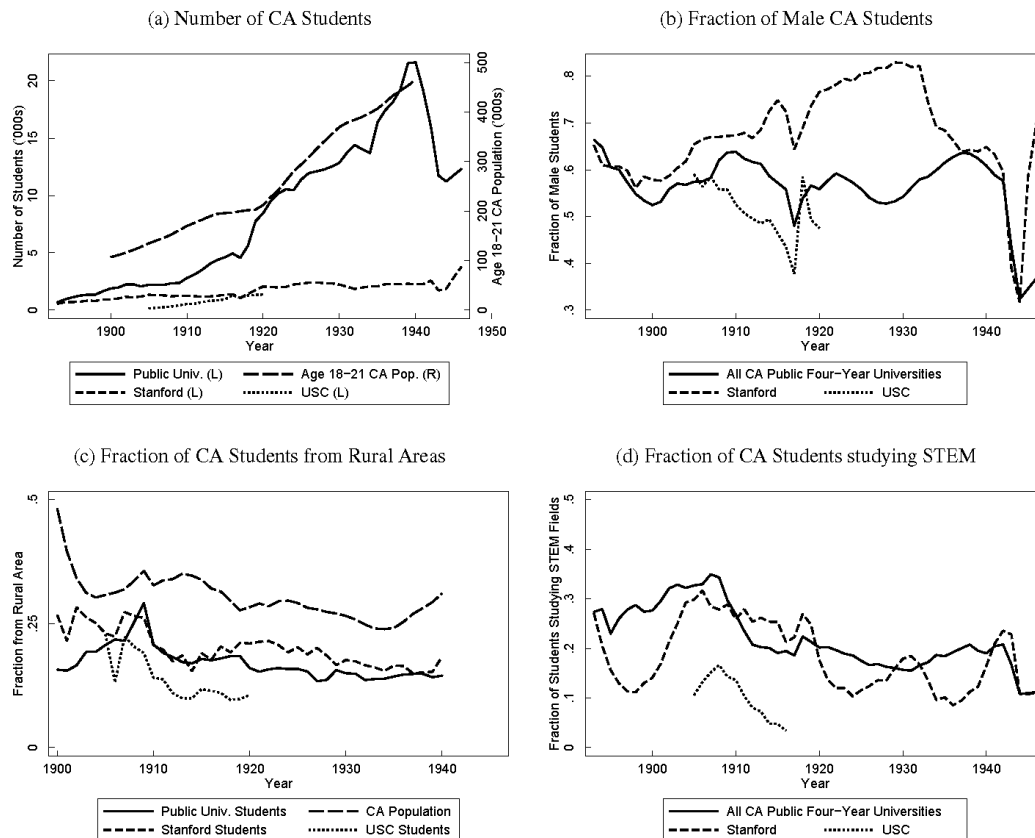
Primary Sources: The University of California Register (1893-1946), Stanford University Annual Register (1893-1946), and USC Year Book (1905-1920).

Finally, I algorithmically link student records across years into a panel using combinations of parts of their first and last names, home towns, fields of study, and year of study.²² A large plurality of student-year entries belong to students who attend university for exactly four years, though many students (17 percent of student-year entries) only appear in a single year.²³ I define new students as students who appear in the dataset for the first time, and university graduates as students who appear in their fourth (senior) year before exiting the panel.

²² See the Appendix for more details on the linking algorithms used in this study.

²³ Such single-year students were not uncommon in early 20th century California; high school teachers with four-year degrees from other schools or states were obligated to spend one year of post-graduate study at a university (where they were often categorized as fourth-year undergraduates) and many students attended universities as 'special students', taking a year of classes without earning a degree.

Figure 2: Summary Statistics on University Education in California, 1893-1946



Note: The number or fraction of students from California attending a large California university between 1893 and 1946. Large universities are defined as four-year tertiary schools with more than 1,000 students in 1920, and include the University of California (Berkeley 1893-1946, UCLA 1920-1946, Davis 1922-1946), Stanford University (1893-1920), and the University of Southern California (1905-1920). About 7 percent of students cannot be assigned hometowns or genders, largely due to uncommon or androgynous first names, uncommon hometowns, and imperfect data cleaning. Years refer to the starting year of each academic year, which runs from August to June. (a) Number of Age 18-21 CA residents calculated as the total Census-estimated population of CA multiplied by the fraction of age 18-21 individuals in the IPUMS 1 or 5 percent samples of the decennial US Census (linearly interpolated between decades). (b) Student genders are determined by matching first names to the most popular names of males and females assigned at birth in the United States around 20 years before each student attends university (according to the Social Security Administration). (c) A town is defined as rural if it is unincorporated or in the bottom half of populations of CA incorporated towns (ranging from 1,600 in 1900 to 3,100 in 1930). Town populations are interpolated from high-order polynomial fits to decennial Census counts and biannual population estimates by municipal clerks made for tax purposes, weighing the two sources equally. The CA rural fraction is the total estimated population of above-median incorporated towns divided by the total Census-estimated population of CA. (d) STEM fields are defined as engineering or pre-engineering (mechanical, electrical, civil, mining, or unknown), medicine (including dentistry), and chemistry (UC and Stanford) or fields resulting in a B.S. degree (USC).

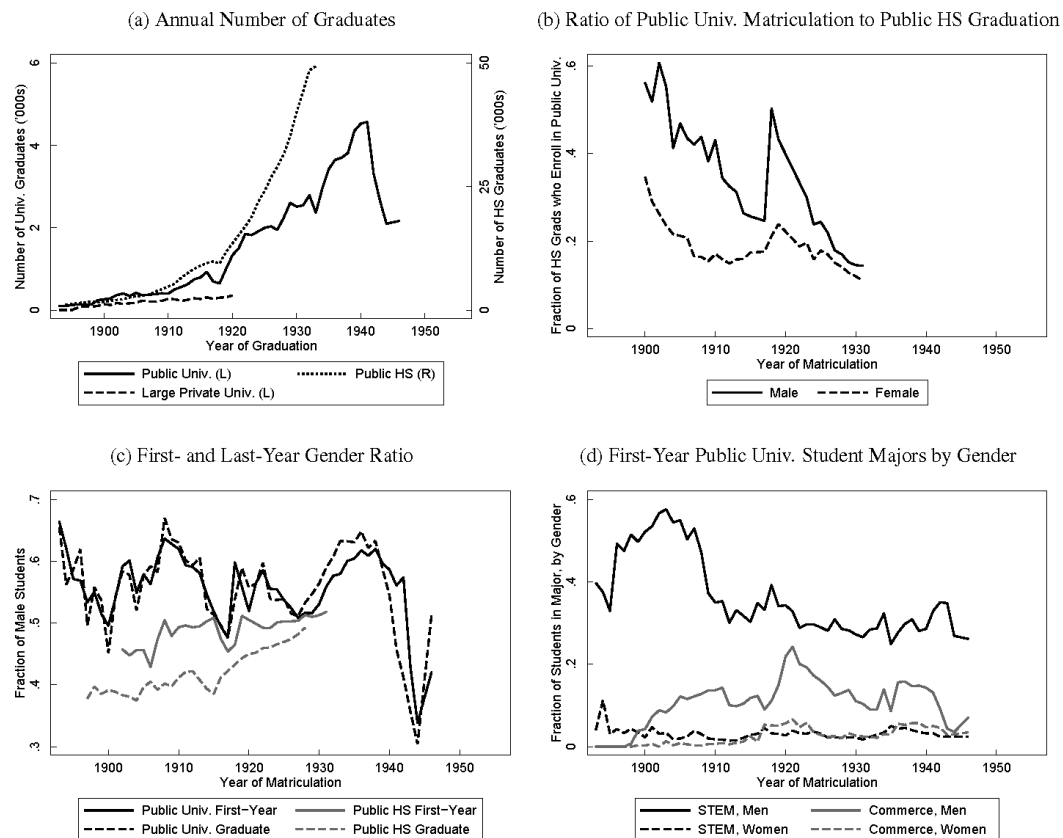
Primary Sources: The University of California Register (1893-1946), Stanford University Annual Register (1893-1946), USC Year Book (1905-1920), US Census (annual CA population), IPUMS US Census samples (CA residents age 18-21), and the Annual Report of Financial Transactions of Municipalities and Counties of California (annual CA urban population).

Given the tremendous processing required to produce this data, quality is both a concern and a high priority. I analyze more than two (and as many as four) copies of most registers used in this study, collating them and keeping only the highest-quality representation of each page (that is, the scan with the highest number of complete entries). I have also spent considerable time writing specific cleaning algorithms for each register template, producing high-quality results on random inspection.

Finally, I compare summary statistics from the UC Berkeley digitized registers (which represent about half of all collected data) to those published in their annual Statistical Summary published between 1918 and 1938, comparing the total number of students, the proportion of students who are male, the proportion of students who are in the School of Letters and Sciences (SLS, the most popular degree program), and the proportion of students in their fourth year. Appendix Figure A1 shows that the comparisons are very close, with a median 2.6 percent *absolute* gap in total enrollment, 0.9 percent gap in gender proportion, and 0.4 percent gap in the proportion enrolled in SLS across the available years.²⁴ Remaining errors are likely spurious, resulting from arbitrarily low-quality digitization efforts, but might attenuate the results presented below.

²⁴ The mean (max) gaps for each measure were 3.2 (10.4) percent for total enrollment, 1.3 (3.4) percent for the gender proportion, and 0.9 (5.4) percent for the proportion enrolled in SLS. The poorest comparisons occur around the first World War, when official figures may have been imprecise due to resource availability and population volatility.

Figure 3: University Matriculation and Graduation in California, 1893-1946



Note: The number or fraction of students from California attending a large California university between 1893 and 1946. Large universities are defined as four-year tertiary schools with more than 1,000 students in 1920, and include the University of California (Berkeley 1893-1946, UCLA 1920-1946, Davis 1922-1946), Stanford University (1893-1920), and the University of Southern California (1905-1920). About 7 percent of students cannot be assigned hometowns or genders, largely due to uncommon or androgynous first names, uncommon hometowns, and imperfect data cleaning. Years refer to the starting year of each academic year, which runs from August to June. (a) University graduates are defined as individuals who appear in their university register as being enrolled in their fourth (senior) year, though a small fraction may never actually earn a degree. High school graduates earned diploma in that year from a public CA high school. (b) Measured as total CA public university graduates over the last year's total number of high school graduates. (c) First-year public university students are those who appear for the first time in the university register (register entries are flexibly linked across years using first and last name, degree, and hometown), while public university graduates are first-year students who eventually achieve the fourth year (again using the linked files). First-year high school students are those in the ninth grade, while graduates are those who earn high school diplomas three years later. (d) The fraction of first-year public university students who choose each degree field. STEM fields are defined as engineering or pre-engineering (mechanical, electrical, civil, mining, or unknown), medicine (including dentistry), and chemistry.

Primary Sources: The University of California Register (1893-1946), Stanford University Annual Register (1893-1946), USC Year Book (1905-1920), and the Biennial Reports of the California State Department of Education (high school records).

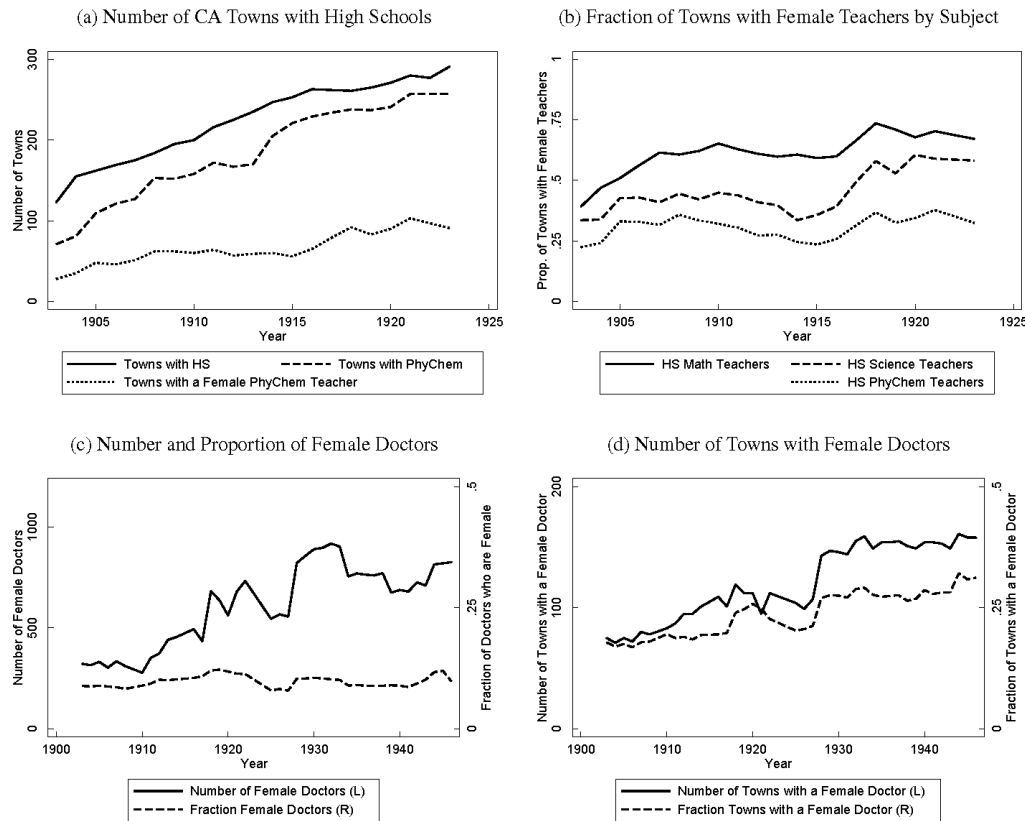
B. Descriptive Statistics: Early 20th Century Higher Education

California's youth population grew by almost 60 percent between 1900 and 1920, but the number of students attending public or large private universities increased by more than 220 percent.²⁵ The number of students increased again by 150 percent between 1920 and 1940. Such growth was typical of the United States, and has been observed previously using Census data (Goldin and Katz, 2008). Figure 1 maps the simultaneous geographic expansion of California higher education.

²⁵ Unless otherwise stated, the statistics presented in this section refer to college students from California at the University of California, Stanford University, and the University of Southern California until 1920, and only the two former universities after 1920. These schools—the only California four-year post-secondary institutions with more than 1,000 students until the late 1930s—comprised more than 80 percent of four-year university students for most of the studied period (by 1940, they enrolled 71 percent of such students (U.S. Office of Education, 1947, p. 404)).

In 1900, 72 California towns sent at least five students to UC or Stanford (by far the two largest universities in the state), and 270 towns sent any students at all; by 1940, 272 towns sent at least five students, and 655 towns at least one. In 1900, large universities had fewer than five students from 16 out of the 58 California counties; by 1940, that number was down to 4, and every county was represented by at least one student. Despite this geographic expansion, Figure 2(c) shows that rural representation slowly declined between 1910 and 1940, dropping from over 20 percent to a trough of around 14 percent of university students (despite making up 25 percent of the state's population) in the 1930s.²⁶

Figure 4: California High School Teachers and Doctors in the Early 20th Century



Note: The number of California towns with high schools and doctors or the number or fraction of working California public high school teachers and doctors, between 1903 and 1927. Through at least 1914, no town without a public high school had a private high school. About 5 percent of teachers and doctors cannot be assigned genders, largely due to uncommon or androgynous first names and imperfect data cleaning. For high school teachers, years refer to the fall semester. (b) Fraction conditional on the town's having at least one teacher in that subject. Math includes algebra, geometry, and trigonometry; science includes physics and chemistry as well as general science, zoology, botany, physiology, and general science. (c) and (d) Doctors include physicians, surgeons, a small number of osteopaths, and (in rare cases) alternative medical practitioners.

Primary Sources: Heath's Directory of Secondary Schools (1907-1914), the California Directory of Secondary and Normal Schools (1915-1924), and the Medical Society of the State of California's Official Registry and Directory of Physicians and Surgeons (1903-1946).

Figure 1 also emphasizes the substantial heterogeneity in the magnitude of the university gender gap, with towns near to (and far from) California's urban and farming centers having very high and very low proportions of women among their university students.

Figure 2(b) shows that universities were close to being gender-balanced in 1900, but that before and after that year they were strongly skewed towards men in aggregate, with the exceptions of the first and second World Wars (which induced short-term declines in the gender ratio by 10 and 30 percentage points (pp.), respectively). The naughts (1900–1910) and the Great Depression were periods of declining female representation (by nearly 10 pp. in each case), while Stanford's decision in 1932 to loosen restrictions on its number of female students (capped at 500 since 1899) leads to a nearly-20 pp. decline in the gender gap at that school in the mid-1930s.²⁷

²⁶ A town is defined as rural if it is unincorporated or in the bottom half of populations of California incorporated towns (1,600 in 1900; 3,100 in 1930). Town populations are interpolated from high-order polynomial fits to decennial Census counts and biannual population estimates by municipal clerks made for state tax purposes, weighing the two sources equally in aggregate (that is, weighing each Census observation four times more than each biannual tax estimate).

²⁷ For information about Stanford's 500-women cap, see Leland Stanford JU, 1900-1934.

Figure 2(d) displays a surprising decline in the prevalence of STEM major selection between 1910 and 1920, during which the proportion of students studying engineering, chemistry, or pre-medicine dropped by nearly a third (from which it has never recovered). The decline at UC was contemporaneous with the substantial shrinking of the Mining School, which was a popular engineering field, but the School's contraction can neither account for the magnitude of the decline nor explain why students who would have studied mining didn't switch to another engineering field. The decline appears to have been similar in magnitude at both Stanford and USC.

Figure 3, which focuses on the University of California, shows in part (d) that the decline occurred wholly among men, since few women studied STEM fields, and was met at the end of the 1910s (and after the first World War) by a large increase in the proportion of both men and women studying in the School of Commerce (today's business school). STEM field selection's association with growth suggest this 1910s decline as a potent topic of future study, but the change's abruptness makes the information frictions studied here an unlikely primary mechanism. The small but persistent rise in the proportion of women studying STEM fields throughout this period, on the other hand, could perhaps be explained by the geographic dispersion of relevant information.

The remainder of Figure 3 describes matriculation and graduation characteristics of the University of California. One concern in studying higher education in this period is the popularity of two-year 'junior' college Associate of Arts degree. While UC did not provide such a degree, students were offered a certificate of completion after two years, suggesting the possibility of students' (and, in particular, women's) use of the school to obtain higher education without studying for a Bachelor's degree, biasing measures of four-year university attendance. Panel (c), however, assuages this concern; until the Great Depression, the first-year gender ratio conditional on eventual graduation was nearly identical with the unconditional first-year gender ratio, suggesting against substantial female attrition after the second year. Panel (b), on the other hand, shows that male high school graduates had a higher rate of university matriculation than women, likely in part because women graduated high school at higher rates than men until at least the 1930s, with some male attrition occurring earlier.

Finally, Figure 4 summarizes the expansion of high schools and doctors across California throughout the Progressive Era and subsequent decades, focusing in particular on the expansion of female doctors and PhyChem (physics or chemistry) teachers. Panels (a) and (b) show that, during a period of substantial high school expansion, the number of towns with female PhyChem teachers more than tripled between 1903 and 1923, slightly increasing the share of schools with such teachers (from 23 to 31 percent). Nevertheless, the number of schools with female PhyChem teachers remained far lower than the number with any female science or math teachers; 55-70 percent of towns throughout the period had female math teachers.

Meanwhile, Panels (c) and (d) show that despite stagnation in the proportion of doctors who were women (around 10 percent), female doctors experienced similar geographic expansion: from 75 towns in 1903 (18 percent of towns with doctors) to 113 towns (23 percent) in 1922 and 219 (26 percent) in 1933. These doctors and teachers were themselves educated by California's higher education system; in 1914, the teacher directory shows that 47 percent of female PhyChem teachers in California had attended the University of California, while another 27 percent had studied at Stanford or USC. The next section will identify and examine the causal implications of the geographic expansion of these college-educated, STEM-oriented female professionals across California.

C. Event Study Models of Role Model Effects

Empirical Methods

This study does not presently include a structural model, but clearly a number of factors—including start-up costs (Bettinger et al., 2012), credit constraints (Dynarski, 2003), and expected returns (Bleemer and Zafar, 2015)—contribute to individuals' four-year college attendance and major selection decisions. The descriptive evidence above, which shows a substantial proportion of university students from rural areas and tremendous geographic expansion in the 1910s and 1920s, suggests an important role for information frictions. By focusing on the partial-equilibrium role model effects of the expanding presence of college-educated female professionals in science-oriented fields, this study provides a template for future examination of the broader relevance of information frictions in the expansion and evolving major distribution of early 20th century higher education.

I focus on seven outcome variables of interest: whether any men or women from the town enroll in a university, the proportion of university students from the town who are men, whether any men or women from that town select a STEM field of study, and whether any men or women from that town become doctors licensed in California. Consider the following event: a female high school PhyChem teacher or a female doctor is observed in a town—and stays for at least one year—for the first time.²⁸ Let Y_{it} be an outcome measure among new university matriculants from town i in year t . Define E_{it0} as an indicator for the event's occurring in that town-year, and E_{itj} as an indicator for the event's occurring between 1 and j years before t in town i .²⁹ I estimate linear least-squares regressions of the form:

²⁸ Towns which are observed with female teachers or doctors in the first or second observed year are defined as never experiencing that event, and doctors or teachers who only appear in a single year are excluded (since their appearance may be a clerical or data error). Teacher and doctor records are available annually between 1903 and 1923.

²⁹ For $j < 0$, define E_{itj} as an indicator for the event's occurring between 1 and $-j$ years after t in town i .

(1)

Table 2: Event Study Estimates for the Arrival of a Female High School PhyChem Teacher

Timing (Inclusive)	University Matriculation			STEM Selection		Grad. Employment	
	At Least One Male	Female	Proportion Students Male	At Least One STEM Male	Female	At Least One Doctor Male	Female
One Year After Event	-0.047 (0.048)	0.118 (0.041)	-0.130 (0.041)	-0.094 (0.058)	-0.013 (0.048)	0.010 (0.006)	0.027 (0.045)
Five Years After Event	0.001 (0.023)	0.049 (0.027)	-0.036 (0.020)	-0.013 (0.036)	-0.027 (0.023)	0.002 (0.002)	0.010 (0.024)
Ten Years After Event	-0.002 (0.023)	0.035 (0.025)	-0.030 (0.020)	-0.030 (0.034)	-0.006 (0.027)	0.001 (0.001)	0.000 (0.019)
One Year Before Event	-0.027 (0.046)	-0.038 (0.051)	0.013 (0.036)	-0.010 (0.058)	0.011 (0.042)	-0.002 (0.001)	-0.006 (0.036)
Five Years Before Event	0.023 (0.034)	-0.007 (0.029)	0.017 (0.026)	0.031 (0.040)	-0.001 (0.021)	-0.001 (0.002)	-0.008 (0.018)
Ten Year Pre-Trend	0.000 (0.004)	-0.002 (0.003)	0.002 (0.003)	0.004 (0.005)	0.000 (0.003)	-0.000 (0.000)	-0.000 (0.002)

Note: Note: β coefficients from separate OLS estimates of equations (1) and (2) by outcome and number of years (inclusive), with clustered standard errors (by town) in parentheses. The regressions control for indicators of the number of high school teachers and the number of PhyChem (physics or chemistry) teachers in the town as well as a quartic in interpolated log town population (see the Appendix). The event is defined as the first occasion since 1903 in which a town has a female PhyChem teacher who stays at least one year, having had any such teachers for at least the prior two years. Teacher genders are determined by use of 'Miss' or 'Mrs.' (1907-1914) or by matching first names to the most popular names of males and females assigned at birth in the United States around 30 years earlier (according to the Social Security Administration) (1915-1924). Matriculation is defined as students who appear for the first time in University of California or Stanford University directory in that year (see the Appendix for linking algorithms).

Primary Sources: University of California Register (1893-1946), Stanford University Annual Register (1893-1946), Heath's Directory of Secondary Schools (1907-1914), and the California Directory of Secondary and Normal Schools (1915-1924).

$$Y_{it} = \beta_j E_{itj} + \alpha_i + \gamma_j + \delta X_{it} + \epsilon_{it}$$

where the coefficient of interest is β_j when $j > 0$, which estimate the change in the level of Y_{it} in impacted towns in the j years after the event. Following a difference-in-difference event study framework, the model includes town (α_i) and year (γ_j) fixed effects, with time-varying town-level characteristics X_{it} —indicators for either the number of teachers and PhyChem teachers or the number of doctors in the town, as well as a quartic in log town population—included to improve balance and efficiency.³⁰ Standard errors are clustered at the town level.

Estimates of β_j when $j < 0$ are also presented below; under the hypothesis of pre-treatment exchangeability, necessary for the causal interpretation of the β_j coefficients, they will be approximately equal to 0. To further examine the causal interpretability of the estimates below, I also estimate the following regression for each outcome:

$$Y_{it} = \beta' t E_{it,-10} + \alpha_i + \gamma_j + \delta X_{it} + \epsilon_{it} \quad (1)$$

where $\beta' t$ estimates the ten-year trend preceding the event. The presence of a pre-trend in any outcome would provide evidence that the arrival of female STEM professionals was endogenous.

All students enrolled at a four-year California university must have attended high school (in order to satisfy admission requirements), but not all California towns had high schools. However, student records include only individuals' home towns, which could downwardly bias the estimated impact of town-level role model effects (since some 'treated' individuals may be included in the control group). To avoid this bias, I measure the distance between every California town and every high school open in a given year (using the Haversine great-circle formula) and assign all students to the town-with-high-school closest to their actual hometown, showing results without reassignment as a robustness check.³¹ Robustness checks also include restricting the sample to public or private universities, excluding 1918 and 1919 (in which World War One might confound treatment estimates), and including town-level time trends (or some combination of those).³²

³⁰ Log town population is interpolated from decennial Census and annual municipal tax records; see Note 23.

³¹ Without the reassignment step, an additional assumption must be made about town entrance and exit; I include towns that have appeared at least once in any register in the dataset.

³² I omit USC from the event study analysis, since its registers are not available throughout the analyzed period. Combined, the University of California system and Stanford University enroll more than 65 percent of California four-year university students in most covered years, with the public university comprising more than half of California university enrollment throughout the period, but that leaves a large portion of the university-going population unobserved. Nevertheless, findings reflected in both student populations suggest robustness across the broader student population.

Table 3: Robustness Analysis of the Arrival of a Female High School PhyChem Teacher

Panel A: Female College Attendance							
Timing (Inclusive)	Baseline	Subsamples Public	Private	Original Location	Exclude WWI Years	Town Time Trends Baseline	No WWI
One Year After Event	0.118 (0.041)	0.131 (0.046)	0.141 (0.054)	0.139 (0.048)	0.127 (0.042)	0.109 (0.045)	0.110 (0.045)
Five Years After Event	0.049 (0.027)	0.047 (0.029)	0.019 (0.027)	0.049 (0.028)	0.050 (0.029)	0.036 (0.034)	0.018 (0.036)
Ten Years After Event	0.035 (0.025)	0.050 (0.028)	0.041 (0.029)	0.034 (0.029)	0.032 (0.027)	0.012 (0.033)	-0.013 (0.034)
One Year Before Event	-0.038 (0.051)	0.004 (0.054)	-0.144 (0.035)	-0.050 (0.053)	-0.082 (0.057)	-0.043 (0.050)	-0.087 (0.055)
Five Years Before Event	-0.007 (0.029)	-0.021 (0.035)	-0.049 (0.027)	-0.031 (0.032)	-0.019 (0.032)	-0.003 (0.034)	-0.024 (0.038)
Ten Year Pre-Trend	-0.002 (0.003)	-0.002 (0.004)	-0.008 (0.003)	-0.004 (0.004)	-0.003 (0.004)	-0.002 (0.004)	-0.004 (0.004)

Panel B: Proportion of College Attendance Male							
Timing (Inclusive)	Baseline	Subsamples Public	Private	Original Location	Exclude WWI Years	Town Time Trends Baseline	No WWI
One Year After Event	-0.130 (0.041)	-0.093 (0.044)	-0.211 (0.070)	-0.132 (0.045)	-0.140 (0.044)	-0.115 (0.044)	-0.124 (0.047)
Five Years After Event	-0.036 (0.020)	-0.016 (0.024)	-0.052 (0.034)	-0.021 (0.022)	-0.034 (0.022)	-0.025 (0.026)	-0.020 (0.027)
Ten Years After Event	-0.030 (0.020)	-0.021 (0.024)	-0.091 (0.033)	-0.029 (0.026)	-0.026 (0.021)	-0.007 (0.026)	0.002 (0.027)
One Year Before Event	0.013 (0.036)	-0.038 (0.043)	0.207 (0.042)	0.022 (0.040)	0.038 (0.041)	0.012 (0.036)	0.036 (0.040)
Five Years Before Event	0.017 (0.026)	0.010 (0.028)	0.081 (0.031)	0.035 (0.025)	0.033 (0.025)	0.015 (0.029)	0.033 (0.028)
Ten Year Pre-Trend	0.002 (0.003)	-0.000 (0.003)	0.012 (0.004)	0.004 (0.003)	0.004 (0.003)	0.001 (0.003)	0.004 (0.004)

Note: β coefficients from separate OLS estimates of equations (1) and (2) by outcome and number of years (inclusive), with clustered standard errors (by town) in parentheses. The dependent variable of each regression in Panel A is an indicator for at least one woman from the town matriculating at a university in that year, and in Panel B is the proportion of college matriculants in the town-year that are male. See the notes to table 2 for an explanation of the baseline specification. The robustness specifications are as follows: columns (2) and (3) restrict the matriculation sample to Public (University of California) and Private (Stanford University) students; column (4) does not re-assign students' hometowns to the nearest town with a high school, thus estimating effects on a larger and higher-variance sample; column (5) excludes events which occur in 1918 and 1919, during American participation in World War One; column (6) includes town-level time trends as an additional control; and column (7) both includes town-level time trends and excludes 1918 and 1919 events.

Primary Sources: University of California Register (1893-1946), Stanford University Annual Register (1893-1946), Heath's Directory of Secondary Schools (1907-1914), and the California Directory of Secondary and Normal Schools (1915-1924).

Estimated Results

Estimating the event study specifications discussed above, I find strong quasi-experimental evidence of Nixon and Robinson's (1999) teacher gender role model effects leading to increased female college attendance. Table 2 displays marginal university attendance by gender before and after a town hires its first female high school PhyChem teacher. The first column shows no short- or long-term effect on whether towns send at least one male student to university.

However, column (2) shows that there is an immediate (significant at 1 percent) and somewhat-persistent increase in the likelihood of the town's sending a female student to university. The immediate effect is an 11.8 percentage point increase the year after the teacher arrives, implying near-universal female university participation, with a persistent effect closer to 4 pp. over 10 years (though not statistically significant). The proportion of new university students who are male, meanwhile, declines by 12.5 pp., with a 3.6 pp. 5-year decline significant at the 10 percent level. Back-of-the-envelope calculations, given that the median treated town has a 122-student high school and a 60-40 university student gender gap, suggest that the likelihood of a female high school graduate's attending college increases by 16 percentage points, from 44 to 60.³³ This effect is large, but

³³ I assume that 6.4 percent of high school students at each school female fourth-years, using the California state average during the period, since I do not observe age distributions at the high school level.

recall that it is conditional on high school graduation in a period when only 10-20 percent of the population graduated, implying a 1.9 pp. increase in college enrollment across the broader young female population.

There are several explanations for the role model effect's attenuated persistence. New teachers of either gender might influence students' college-going through differential enthusiasm or experience; however, analysis of the arrival of new *male* PhyChem teachers (available from the author) suggests that such novelty effects are negligible. Female teachers' short tenures provide a second explanation. The median tenure of female PhyChem teachers between 1903 and 1923 was four years in California's ten largest cities but only two years outside those cities; while information about female college attendance outlasts female PhyChem teachers' presence, in part through their teaching younger students, the teachers' departure might mitigate their effect over time. Third, independent information about college-going might have become more broadly-available towards the end of the period, which would mitigate the long-term information impact of a new female teacher's arrival relative to the untreated towns.

Table 4: Event Study Estimates for the Arrival of a Female Doctor

Timing (Inclusive)	University Matriculation			STEM Selection		Grad. Employment	
	At Least One Male	Female	Proportion Students Male	At Least One STEM Male	Female	At Least One Doctor Male	Female
One Year After Event	0.076 (0.036)	0.078 (0.046)	-0.000 (0.032)	0.024 (0.062)	0.063 (0.053)	-0.007 (0.004)	-0.047 (0.022)
Five Years After Event	0.010 (0.023)	0.004 (0.033)	-0.002 (0.024)	-0.014 (0.035)	-0.014 (0.023)	0.003 (0.004)	0.004 (0.017)
Ten Years After Event	0.015 (0.027)	0.027 (0.033)	-0.007 (0.025)	0.030 (0.035)	-0.014 (0.024)	0.002 (0.003)	0.020 (0.022)
One Year Before Event	-0.016 (0.058)	0.046 (0.062)	-0.029 (0.051)	-0.052 (0.077)	0.106 (0.061)	-0.002 (0.002)	-0.027 (0.024)
Five Years Before Event	0.002 (0.032)	-0.024 (0.039)	0.034 (0.026)	0.032 (0.042)	0.017 (0.022)	-0.003 (0.003)	-0.005 (0.016)
Ten Year Pre-Trend	-0.000 (0.004)	0.000 (0.005)	0.000 (0.003)	0.001 (0.005)	0.003 (0.003)	-0.000 (0.000)	-0.001 (0.002)

Note: β coefficients from separate OLS estimates of equations (1) and (2) by outcome and number of years (inclusive), with clustered standard errors (by town) in parentheses. The regressions control for indicators of the number doctors in the town as well as a quartic in interpolated log town population (see the Appendix). The event is defined as the first occasion since 1903 in which a town has a female doctor who stays at least one year, having had any such doctors for at least the prior two years. Doctor genders are determined by matching first names to the most popular names of males and females assigned at birth in the United States around 30 years earlier (according to the Social Security Administration). Matriculation is defined as students who appear for the first time in University of California or Stanford University directory in that year (see the Appendix for linking algorithms). Primary Sources: University of California Register (1893-1946), Stanford University Annual Register (1893-1946), Medical Society of the State of California's Official Registry and Directory of Physicians and Surgeons (1903-1946).

Table 3 displays a number of robustness checks for each of these results:

1. Public and private university matriculation are separately modeled.
2. Rather than assigning university students whose hometown has no high school to the geographically-nearest school, I estimate the model using students' reported hometowns.
3. American participation in World War One occurs in the middle of my estimation period, and 1918 and 1919 were the most common years in which towns obtained their first female PhyChem teachers (likely because their previously-male teachers went to war). Since the teachers hired under these circumstances might have been different in quality or formal position (perhaps being treated as temporary workers), I omit events from those two years.
4. Town-level time trends are included as an additional control variable.

None of these changes substantively alter the estimated effects presented above.

Columns (4) and (5) of Table 2 show the event study effects of a new female PhyChem teacher on STEM field selection. Male STEM participation declines by 9.4 percentage points in the year after a town's first female PhyChem teacher arrives (significant at the 10 percent level). The decline is driven wholly by public university students, where most STEM students attended at the time, and might reflect either an aversion arising from knowledge of female scientific practitioners or poorer academic performance in a female-taught course (à la Dee (2006)). The absence of an effect on STEM selection among men after the initial arrival of female doctors (see Table 4) provides suggestive evidence of the latter mechanism. There is no aggregate impact on STEM selection among women, though additional robustness analysis suggests a shift in composition among medical-bound women from public schools to private schools, perhaps providing evidence of an increase in income or class among female STEM students (though the evidence is imperfectly balanced). There is also no measurable impact on students' becoming

doctors.

While the role model effects literature has focused on student-teacher matches due to data availability and potential endogeneity, student-doctor matches (when identifiable) provide a cleaner measure of role model effects on young person decisions, since doctors do not formally educate students (which might provide an alternative mechanism through which teachers influence students' behavior).

Table 5: Robustness Analysis of the Arrival of Female Doctors: Female University Attendance

Timing (Inclusive)	Baseline	Subsamples		Original Location	Exclude WWI Years	Town Time Trends Baseline	Trends No WWI
		Public	Private				
One Year After Event	0.078 (0.046)	0.060 (0.051)	0.094 (0.060)	0.033 (0.048)	0.073 (0.069)	0.081 (0.047)	0.085 (0.070)
Five Years After Event	0.004 (0.033)	0.011 (0.033)	-0.007 (0.030)	0.019 (0.035)	0.009 (0.048)	0.039 (0.038)	0.027 (0.045)
Ten Years After Event	0.027 (0.033)	0.047 (0.034)	-0.002 (0.028)	0.046 (0.035)	0.044 (0.047)	0.060 (0.037)	0.055 (0.045)
One Year Before Event	0.046 (0.062)	0.098 (0.062)	0.064 (0.073)	-0.010 (0.056)	-0.004 (0.114)	0.038 (0.058)	-0.026 (0.105)
Five Years Before Event	-0.024 (0.039)	-0.014 (0.039)	0.006 (0.027)	-0.017 (0.032)	-0.034 (0.050)	-0.045 (0.040)	-0.059 (0.052)
Ten Year Pre-Trend	0.000 (0.005)	0.002 (0.005)	0.000 (0.003)	-0.001 (0.004)	-0.003 (0.006)	-0.002 (0.005)	-0.006 (0.007)

Note: See notes to Tables 3 and 4. The dependent variable in each regression is an indicator for at least one woman from the town matriculating at a university in that year. Primary Sources: University of California Register (1893-1946), Stanford University Annual Register (1893-1946), Medical Society of the State of California's Official Registry and Directory of Physicians and Surgeons (1903-1946).

Table 4 shows that towns that hire their first female doctor experience immediate increases in female college-going of nearly the same magnitude as those which hire their first female PhyChem teacher, with a 7.8 pp. increase in the proportion of towns that send at least one female student to college (significant at 10 percent). However, the proportion of *male* students attending college also increases by 7.6 percent, mitigating the impact on the university gender ratio to economic and statistical insignificance. Table 5 provides robustness analysis of the increase in female university participation, showing that the estimates are robust to time trends but become noisier when towns which initially obtained female doctors during WWI are excluded.

Unlike in the case of female PhyChem teachers, Table 4 presents some evidence that female doctors' initial appearance leads to a long-term increase in STEM field selection by young women. While there is no evidence that STEM participation declines among men, the likelihood with which a new female student from the town studies a STEM field (usually pre-medicine) increases by 6.3 pp. in the year after the initial female doctor's arrival, though it is not statistically significant. Further analysis, available from the author, shows that the increase is statistically significant 6 and 9 years after the initial female doctor's arrival, but the ten-year increase in female STEM participation is negligible. However, the female doctor's arrival appears to decrease the likelihood with which young women from that town become California-educated doctors themselves. Female doctors seem to not only provide young women (and, perhaps, their parents) with the knowledge that education and labor participation are possible and desirable, but may also (knowingly or unknowingly) propel young women into STEM fields, though not into their own profession.

These results provide substantial evidence of large role model effects among women in early 20th century California. More broadly, they suggest a number of new mechanisms which played substantial roles in the important decisions (like education attainment and participation in technology-related fields) that drove growth and the wealth distribution through the 20th century, including geographic information frictions and social networks. Finally, they provide evidence not only for the importance of these mechanisms, but also for their tractability and identification—despite previously-insurmountable data limitations—through new widely-available digitization techniques.

D. Conclusion

In this study, I summarize four new data series describing higher education in early 20th century California, showing that previously-observed volatility in the university gender gap occurred simultaneously with persistent rural college representation and a sharp decline in STEM field selection in the early 1910s. The Progressive Era also brought California a large persistent increase in the proportion and geographic expansion of female high school physics/chemistry teachers and doctors. I present evidence that this expansion produced a substantial positive feedback loop through a role model (information) mechanism, pushing more women to college and (in some cases) to study in STEM fields.

This study is the first of several studies that are part of the UC Cliometric History Project based at the Center for Studies in

Higher Education on the UC Berkeley campus, and pursued to help celebrate the 150th anniversary of the University of California in 2018. Future papers will extend its analysis in several ways. First, I am currently working with the Office of the President of the University of California and the Registrar's offices of several UC campuses to digitize and collect student-level administrative data since 1946, enabling more contemporary analysis of the roles of California universities in promoting the state's growth, economic mobility, and gender equality. Second, I have collected 1940 Census data and CABI birth records (by county, year, and mothers' maiden name) to estimate the general equilibrium effect of teacher and doctor gender assignment on long-term economic and demographic outcomes.³⁴ Third, I have also collected the full-count 1890-1930 Censuses, enabling limited panel analysis of economic mobility. Finally, I am constructing a machine-learning algorithm using Census data to identify individuals' ethnicity using their first and last names, and will expand my event study analysis to examinations of ethnicity. All of these projects rely on unique comprehensive data sources providing novel insight into the historical role of higher education in 20th century California.

³⁴ See Goldin and Katz (1999a). As late as the 1920s female teachers were often prevented from marrying; see Goldin (1991).

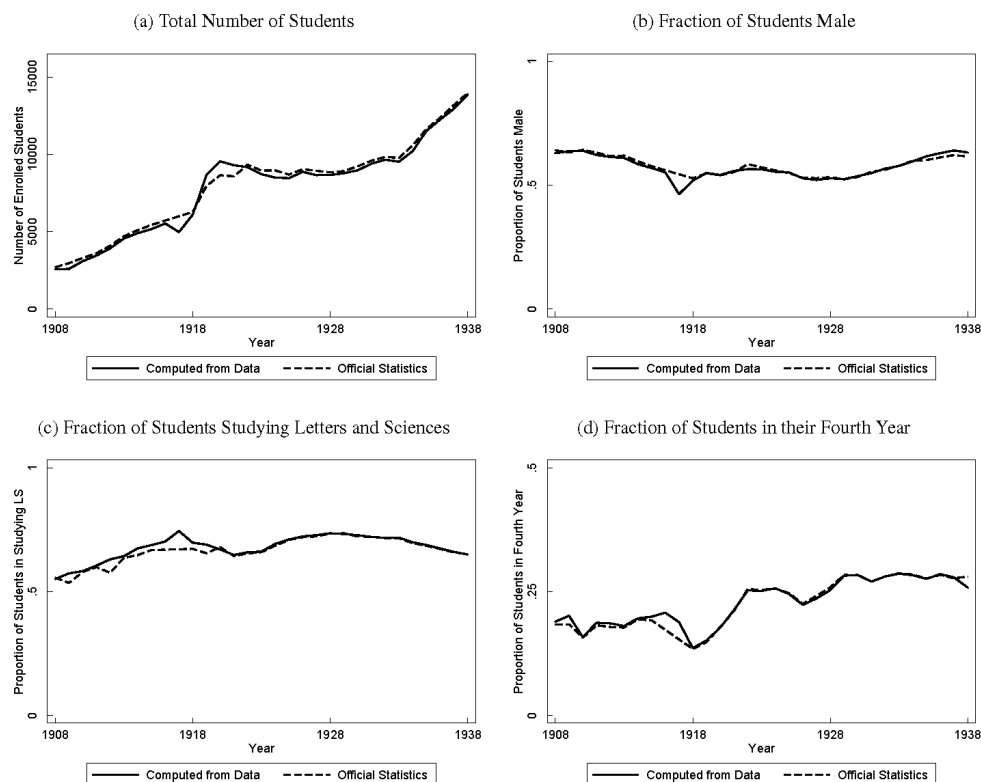
REFERENCES

- Altonji, Joe, Peter Arcidiacono, and Arnaud Maurel. 2016. "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects." Handbook of the Economics of Education 5 (7):305–396. URL [Link](#).
- Altonji, Joseph G., Lisa B. Kahn, and Jamin D. Speer. 2014. "Trends in Earnings Differentials across College Majors and the Changing Task Composition of Jobs." American Economic Review 104 (5):387–393. URL [Link](#).
- Autor, David. 2014. "Skills, education, and the rise of earnings inequality among the 'other 99 percent'." Science 344 (6186):843–851. URL [Link](#).
- Bettinger, Eric P. and Bridget Terry Long. 2005. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." The American Economic Review 95 (2):152–157. URL [Link](#).
- Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment." Quarterly Journal of Economics 127 (3):1205–1242. URL [Link](#).
- Bleemer, Zachary, Matthew Wiswall, and Basit Zafar. 2015. "Long-Term Trends in Fields of Study." Manuscript.
- Bleemer, Zachary and Basit Zafar. 2015. "Intended College Attendance: Evidence from an Experiment on College Returns and Cost." Federal Reserve Bank of New York Staff Reports (739). URL [Link](#).
- Bound, John, Michael Lovenheim, and Sarah Turner. 2010. "Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources." American Economic Journal: Applied Economics 2 (3):129–157. URL [Link](#).
- California. 1890-1934. Biennial Report of the California State Department of Education. Sacramento: California State Printing Office. URL [Link](#).
- . 1910-1938. Annual Report of Financial Transactions of Municipalities and Counties of California. Sacramento: California State Printing Office. URL [Link](#).
- . 1922a. Report of the Adjutant General of the State of California. Sacramento: California State Printing Office. URL [Link](#).
- . 1922b. Report of the Selective Service Administration of California. Sacramento: California State Printing Office. URL [Link](#).
- California State Board of Education. 1915-1925. Directory of Secondary and Normal Schools. Sacramento: California State Printing Office. URL [Link](#).
- California State Library. 1906-1922. News Notes of California Libraries. Sacramento: California State Printing Office. URL [Link](#).
- Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." Quarterly Journal of Economics 125 (3):1101–1144. URL [Link](#).
- Chilcote, Gaylord H. and Charles Fletcher Scott. 1907-1915. Heath's Directory of California Secondary Schools. San Francisco: D.C. Heath and Co. URL [Link](#).
- Dee, Thomas. 2006. "Teachers and the Gender Gaps in Student Achievement." Journal of Human Resources 42 (3):528–554. URL [Link](#).
- Douglass, John. 2007. The Conditions for Admission: Access, Equity, and the Social Contract of Public Universities. Stanford: Stanford University Press. URL [Link](#).
- Dynarski, Susan. 2003. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion." American Economic Review 93 (1).
- Feigenbaum, James. 2015. "Intergenerational Mobility during the Great Depression." Manuscript URL [Link](#).
- Ferrier, William Warren. 1930. Origin and Development of the University of California. Berkeley, CA: Sather Gate Book Shop.
- . 1937. Ninety Years of Education in California. Berkeley, CA: Sather Gate Book Shop.
- Goldin, Claudia. 1991. "Marriage Bars: Discrimination Against Married Women Workers, 1920's to 1950's." In Favorites of Fortune: Technology, Growth and Economic Development Since the Industrial Revolution, edited by Patrice L. R. Higonnet, David S. Landes, and Henry Rosovsky. Cambridge: Harvard University Press, 511–538. URL [Link](#).
- . 1997. "Career and Family: College Women Look to the Past." In Gender and Family Issues in the Workplace, edited by Francine D. Blau and Ronald G. Ehrenberg. New York: Russell Sage Foundation Press, 20–58. URL [Link](#).
- Goldin, Claudia and Lawrence Katz. 1999a. "Human Capital and Social Capital: The Rise of Secondary Schooling in America, 1910–1940." Journal of Interdisciplinary History 29 (4):683–723. URL [Link](#).
- . 1999b. "The Shaping of Higher Education: The Formative Years in the United States, 1890 to 1940." Journal of Economic Perspectives 13 (1):37–62. URL [Link](#).
- . 2000. "Education and Income in the Early Twentieth Century: Evidence from the Prairies." Journal of Economic History 60 (3):782–818. URL [Link](#).
- . 2008. The Race between Education and Technology. Cambridge, MA: Harvard University Press.
- Goldin, Claudia, Lawrence Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." Journal of Economic Perspectives 20 (4):133–156. URL [Link](#).
- Gordon, Lynn D. 1990. Gender and Higher Education in the Progressive Era. New Haven: Yale University Press.
- Hoffman, Florian and Philip Oreopoulos. 2009. "A Professor Like Me: The Influence of Instructor Gender on College Achievement." Journal of Human Resources 44 (2). URL [Link](#).
- Leland Stanford Junior University. 1893-1946. Annual Register. Stanford: Stanford University Press. URL [Link](#).
- Levenshtein, Vladimir I. 1966. "Binary Codes Capable of Correcting Deletions, Insertions and Reversals." Soviet Physics-Doklady 10 (8):707–710. URL [Link](#).
- Lochner, Lance and Alexander Monge-Naranjo. 2012.

- "Credit Constrints in Education." Annual Review of Economics 4 (1):225–256. URL [Link](#).
- Mare, Robert D. 1991. "Five Decades of Educational Assortative Mating." American Sociological Review 56 (1):15–32. URL [Link](#).
- May, Henry F. 1993. Three Faces of Berkeley: Competing Ideologies in the Wheeler Era, 1899-1919. University of California, Berkeley: Center for Studies in Higher Education and Institute of Governmental Studies. URL [Link](#).
- Medical Society of the State of California. 1903–1933. Official Registry and Directory of Physicians and Surgeons in the State of California. San Francisco. URL [Link](#).
- Moreman, Grace E. 2006. Student Life at University of California, Berkeley During and After World War I. Lewiston, CA: The Edwin Mellen Press.
- Muralidharan, Karthik and Katki Sheth. 2015. "Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers." Journal of Human Resources URL [Link](#).
- Nixon, Lucia A. and Michael D. Robinson. 1999. "The Educational Attainment of Young Women: Role Model Effects of Female High School Faculty." Demography 36 (2):185–194. URL [Link](#).
- Ruggles, Stephen, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2015. Integrated Public Use Microdata Series: Version 6.0. Minneapolis: University of Minnesota.
- Smith, James P. 1984. "Race and Human Capital." American Economic Review 74 (4):685–698. URL [Link](#).
- University of California. 1893-1946. University of California Register. Berkeley: The University Press. URL [Link](#).
- University of Southern California. 1905-1920. Year Book. Los Angeles: USC Press. URL [Link](#).
- US Bureau of Education. 1925. Bulletin 1924-14: Biennial Survey of Education 1920-1922, Volume II. Washington: Government Printing Office. URL [Link](#).
- U.S. Office of Education. 1947. Biennial Surveys of Education in the United States, 1938-1942. Washington: Government Printing Office. URL [Link](#).
- Wiswall, Matthew and Basit Zafar. 2015. "Determinants of College Major Choice: Identification Using an Information Experiment." Review of Economic Studies 82 (2):791–824. URL [Link](#).
- Wood, Will C. 1917. Proceedings of the State Convention of California High School Principals of 1916. Sacramento: California State Printing Office.
- Zafar, Basit. 2013. "College Major Choice and the Gender Gap." Journal of Human Resources 48 (3):545–595. URL [Link](#).

APPENDIX – Data Sources

Figure A1: Comparison Between Register Estimates and Published Summary Statistics for UC Berkeley



Note: The number or fraction of students attending UC Berkeley between 1908 and 1938. Solid lines represent estimates computed from the digitized records used to produce the figures presented in this paper; dashed lines represent official statistics published by UC Berkeley in its annual Statistical Summary. In the computed data, student genders are determined by matching first names to the most popular names of males and females assigned at birth in the United States around 20 years before each student attends university (according to the Social Security Administration). About 5 percent of students cannot be assigned genders, largely due to uncommon or androgynous first names and imperfect data cleaning. Years refer to the starting year of each academic year, which runs from August to June. Letters and Sciences (LS), the largest program at UC Berkeley throughout the period, includes all students studying in the schools of Letters (today's humanities), Social Sciences, and Natural Sciences, which were combined into LS in 1914, as well as students studying Architecture or Jurisprudence.

Primary Source: The University of California Register (1893-1946).

The following is a list of the sources and providers of the data used in this study, as well as the available content in each of those sources:

1. University of California Register, 1893-1946: Annual administrative records of students attending four-year undergraduate degree programs at the University of California campuses at Berkeley (1893–1946), San Francisco (1917–1946), Los Angeles (1921–1946), and Davis (1922–1946). Available in HathiTrust records [007130126](#), [011249103](#), [007910193](#), [100024883](#), and [003915007](#), which were digitized by partnerships between Google and the University of California, the University of Illinois at Urbana-Champaign, Cornell University, and the University of Michigan. Records include first and last name (1893–1946), middle name (1893–1946), hometown (1893–1904, 1907–1946), year of school (1893–1946), school of enrollment (Letters (1893–1914), Social Sciences (1893–1914), Natural Sciences (1893–1914), Letters and Science (1915–1946), Mechanical Engineering (1893–1946), Civil Engineering (1893–1946), Mining Engineering (1893–1946), Agriculture (1893–1946), Commerce or Business Administration (1893–1946), Chemistry (1893–1946), Engineering (1930–1946), Dentistry (1915–1946), Pharmacy (1935–1946), Optometry (1941–1946), Nursing (1935–1946), Applied Arts (1939–1946), Teacher's College (1921–1938)), professional sub-field (Pre-Medicine (1905–1946), Pre-Law or Jurisprudence (1915–1926), Pre-Architecture (1895–1941)), and local address (1893–1946). Students attending the UCLA Teachers College are omitted, since they are pursuing two-year post-secondary degrees; students studying pharmacy prior to 1935 are omitted because they are pursuing a three-year degree. The 1903 Register is available in print from the UC Berkeley Bancroft Library, which digitized it (not publicly available); the 1945 Register is unavailable.
2. Stanford University Register, 1893-1946: Annual administrative records of students attending graduate or undergraduate degree programs at Leland Stanford Junior University. Records include first and last name (1893–1946), middle name

(1893–1946), hometown (1893–1946), field of study (1893–1946; disaggregated), number of credit-hours earned (1908–1918, 1920–1946), graduate student status (1893–1946), and local address (1893–1946). Graduate students are omitted. Students are assumed to have earned one additional year of standing (in years of school) for every 30 (1907–1916) or 45 (1917–1946) credit-hours earned. [Available](#) from the University Publications Division of the Digital Collections of Stanford University Libraries and Academic Information Resources, which digitized the records. Starting in 1920, only third and fourth-year students are assigned fields of study.

3. University of Southern California Year-Book and Circular of Information, 1905-1920: Annual administrative records of undergraduate students attending the University of Southern California. Available in HathiTrust records [100630461](#) and [000056358](#), which were digitized by partnerships between Google and both the University of Illinois at Urbana-Champaign and the University of Michigan. Records of first and last name (1905–1920), middle name (1905–1920), hometown (1905–1920), number of credit hours earned (1905–1908), year of school (1909–1920), field of study (1905–1908), and degree pursued (Bachelor of Arts (1909–1916), Bachelor of Sciences (1909–1916), Pre-Medical (1916)). Students are assumed to have earned one additional year of standing (in years of school) for every 30 (1905–1908) credit-hours earned.
4. Throop College of Technology Annual Catalogue, 1912-1919, and Bulletin of the California Institute of Technology, 1920-1946: Annual administrative records of undergraduate students attending either Throop College (until 1919) or CalTech (thereafter). [Available](#) from the Caltech CampusPubs repository of the Caltech Library, which digitized the records, and in HathiTrust record [100607120](#), which was digitized by a partnership between Google and the University of Illinois at Urbana-Champaign. No registers were published in 1942 and 1943. Records of first and last name (1912–1946), middle name (1912–1946), hometown (1912–1946), year of school (1912–1946), field of study (1912–1946; disaggregated), and local address (1912–1921).
5. Mills College Catalogue, 1903-1919: Annual administrative records of undergraduate students attending Mills College. Available in HathiTrust record [005808070](#), which was digitized by partnerships between Google and both of the University of Illinois at Urbana-Champaign and the University of Michigan. Records include first and last name (1903–1919), middle name (1903–1919), hometown (1903–1919), and year of school (1903–1919).
6. Heath's Directory of California Secondary and Normal Schools, 1903-1914: Annual privately-collected records of all high school and junior high school teachers employed by participating California public and private high schools (that is, whose clerk provides identifying information to Heath's), as well as all teachers whose identifying information was collected by Heath's clerks (purportedly universal). Available after 1906 in HathiTrust record [006110712](#), which was digitized by a partnership between Google and the University of California, and before 1907 in print from Stanford University Library (digitized by the author; not publicly-available). Records include first and last name (1903–1914), title (1903–1914; Ms., Mrs., Dr.), middle name (1903–1914), employing high school and town of high school (1903–1914), town of residence (1903–1914), subjects and classes taught (1903–1914; disaggregated), degrees earned and time attending post-secondary schools (1903–1914; B.A., Ph.B., M.A., Ph.D., "Summer School", "Studied one year", "Graduate Work", etc.), post-secondary schools attended (1903–1914), years of post-secondary degrees or attendance (1903–1914). Private schools and junior high schools are omitted.
7. California State Board of Education Directory of Secondary and Normal Schools, 1915-1924: Annual administrative records of high school teachers employed by California public high schools. Available in HathiTrust record [010236895](#), which was digitized by a partnership between Google and the University of California. Records include first and last name (1915–1923), middle name (1915–1923), employing high school and town of high school (1915–1923), town of residence (1915–1923), subjects and classes taught (1915–1923), certification level (either five years of post-secondary education ("full certification") or "special certification"). The 1922 Register is unavailable.
8. The Medical Society of the State of California's Official Register and Directory of Physicians and Surgeons in the State of California, 1903-1956: Available in HathiTrust records [011933633](#), [000045888](#), [010753440](#), [100194831](#), and [100553463](#), which were digitized by partnerships between Google and the University of California, the University of Illinois at Urbana-Champaign, and Harvard University. Registers from 1909, 1915, 1922, and 1923 are unavailable.
9. Biennial Report of the Superintendent of Public Instruction and Biennial Report of the California State Board of Education, 1888-1931: Available in HathiTrust records [000061846](#) and [000060049](#), which were digitized by partnerships between Google and the University of Michigan, the University of California, and the New York Public Library.
10. Biennial Survey of Education of the Department of the Interior's Bureau of Education, 1916-1924: [Available](#) in some years from Archive.org, digitized by the American Printing House for the Blind, and others from the [ERIC](#) platform of the US Department of Education and the Institute of Education Sciences.
11. Annual Report of Financial Transactions of Municipalities and Counties of California, 1910-1940: Available until 1922 in HathiTrust record [008959980](#), which was digitized by a partnership between Google and both the University of California and Princeton University, and in print from the UC Berkeley Doe Library.