Creating Career Women: Access to Teaching Jobs and Women's Long-Run Outcomes*

Rebecca Brough
University of California, Davis

October 20, 2026

Click here for the most recent version

Abstract

How do career opportunities change the trajectory of women's work and family lives? I study this question in the context of early twentieth-century teaching, which accounted for nearly a quarter of women's white-collar work in 1920. I digitize records from Illinois teacher-certification exams (1916–1923) and link them to the full-count censuses (1910–1940) to implement a regression discontinuity design around the exam's passing threshold. I find that marginal passers are 13.3 percentage points (34 percent) more likely to ever become teachers and 10.9 percentage points (18 percent) more likely to ever be in the labor force; for marginal failers, the alternative to teaching was typically not a different occupation, but to not enter the labor force at all. Marginal passers remained more likely to be in the labor force over the next 20 years. These long-run effects were concentrated in teaching and persisted after marriage where local policies permitted married women to teach. Women who continued teaching over the long term acquired additional education. Access to teaching also increased the likelihood of remaining unmarried and reduced fertility, but it was not associated with marrying higher-earning partners. These results provide new quasi-experimental evidence that career opportunities in early adulthood can induce women into sustained labor force participation, with corresponding implications for their family outcomes.

*I am grateful for the support of Marianne Bitler, Katherine Eriksson, Marianne Page, Giovanni Peri, and Santiago Pérez. I also thank Matt Freedman, Grace Ortuzar, Joaquín Paleo, David Phillips, Melissa Spencer, Kathya Tapia, Patrick Turner, and Wenni Yang for helpful feedback on earlier drafts. I thank the staff at the Illinois State Archives for their assistance in accessing records. This paper benefited from comments by seminar participants at UC Davis, the Bay Area Economic History Workshop, the All-UC Labor Economics Workshop, and the IPUMS data conference, as well as poster session attendees at the NBER Summer Institute, the Institute for Research on Poverty, and the Economic History Association Annual Meeting. I gratefully acknowledge funding from the Russell Sage Foundation, the Economic History Association, and the UC Davis Provost's Dissertation Year Fellowship. All errors are my own.

Email: rjbrough@ucdavis.edu

1 Introduction

Women's work and family lives, and how they interacted, changed dramatically over the course of the twentieth century in the US. Female labor force participation (LFP) tripled between 1900 and 1990; the majority of this growth was among married women in white-collar and professional work (Figure 1). This growth reflected a fundamental shift in how women engaged in the labor market, both in their persistence in the labor force after marriage and in the types of work they did, which increasingly rewarded education and experience (Goldin, 1990). Occurring alongside these changes there was, with the exception of the baby-boom era, a significant decline in women's fertility.

Work histories suggest that new entrants to the labor force, and the kinds of occupations they first entered, were key determinants of this growth in labor force participation (Goldin, 1989). However, because women who pursue professional work differ in ability, motivations, and preferences, it remains unclear whether access to such work causally generated long-run persistence or instead drew in women who would have worked anyway—and how it affected family outcomes. Existing papers address this by exploiting changes across places and over time in labor demand (Acemoglu, Autor and Lyle, 2004; Fernández, Fogli and Olivetti, 2004; Goldin and Olivetti, 2013), technology (Rashid, 2025; Feigenbaum and Gross, 2024; Greenwood, Seshadri and Yorukoglu, 2005; Vidart, 2024), reproductive technology (Goldin and Katz, 2002), and educational opportunities (Goldin, 2006; Doxey, Karger and Nencka, 2026; Bald, 2025) to study women's entry into work or specific occupations. These aggregate approaches capture changes in opportunities, as well as broader labor-market and institutional changes that potentially accompany them.

In this paper, I use newly digitized historical records on teacher-certification exams to estimate how access to teaching affected women's labor-market trajectories and subsequent education, marriage, and fertility outcomes. In contrast to existing work, I exploit quasi-random, individual-level variation in women's access to professional work, which allows me to hold fixed the labor demand, technologies, and institutions that shaped women's opportunities in aggregate. Access to a teaching career provides a valuable case study of professional work because it was one of the few socially acceptable and highly sought-after professional occupations available to women. In 1920, one-fourth of women's white-collar work was concentrated in teaching, making it the single largest white-collar occupation for women (Figure 2). For many women, a barrier to teaching was a barrier to professional work itself. Teaching was also distinct among occupations available to women in that it had several career-like attributes, including credentialing requirements, relatively high pay that scaled with experience, pensions, and ongoing professional-development requirements.

Teaching is also one of the few occupations in which mandated entry barriers generated quasi-random variation in who gained access. In 1920, all states except Vermont offered an examination pathway to attain teaching credentials (Cook, 1921).[1] These systems imposed strict passing criteria that generated individual-level variation in access to teaching. I focus on the certification system in Illinois, in which examination grading was centralized and done anonymously by the State Examining Board. Applicants completed 13 examinations and were eligible for teaching positions if their cumulative average score met the passing threshold, with no subject score below 60 percent. Using a regression discontinuity design (RD), I compare the life-cycle outcomes of women whose first-attempt average scores fell just above and below the cumulative average score threshold.

To implement this design, I scan and digitize elementary school teacher-certification examination ledgers spanning 1916 to 1923. The records cover all Illinois counties except Cook County[2] and contain approximately 48,000 applicants with detailed information on each applicant and her subject scores. During these years, about half of new elementary certificates issued to women outside Cook County came through examinations (Figure 3). I focus on the 79 percent of applicants who were women. Because applicants were young—18.2 years old on average, with fewer than 1 percent married—I link them to the census immediately preceding the examination, when they are generally observed in their childhood households, and then follow them forward through the 1920, 1930, and 1940 censuses to observe their long-run labor market, education, marriage, and fertility outcomes.

The main threats to my regression discontinuity design are endogenous retaking and manipulation around the passing threshold, but the structure of the examination and grading process allows me to address both. First, applicants could retake the examination, and the decision to retake is endogenous and can bias estimates (Jepsen, Mueser and Troske, 2016). Because applicants could also retain passing subject scores and retake only the subjects they had failed, subsequent attempts differed in nature from the first. I therefore focus on first-attempt scores. Second, I find evidence that graders increased candidates' minimum subject scores to the required 60 percent when this change would allow them to pass the exam. Because graders disproportionately increased minimum scores for candidates whom the increase would put in a passing position, this behavior creates bunching in both the minimum-score and cumulative-score distributions. Unlike in most settings with manipulation, however, this pattern allows me to detect the relevant manipulation and identify the affected test takers. I exclude applicants whose minimum subject score equals 60, after

[1] States varied in whether examinations were required in addition to or in place of other educational credentials.

[2] Cook County was exempt from the statewide certification system.

which the distribution of cumulative scores is smooth around the threshold. Other concerns about strategic manipulation are mitigated by the institutional setting, in which examinations were graded anonymously by the State Board of Examiners, which limited graders' ability to influence scores for specific named applicants.

Passing the teacher examination substantially increased women's entry into both teaching and the labor force. Marginal first-attempt passers are 13.3 percentage points (34 percent) more likely ever to be observed as teachers in a subsequent census and 10.9 percentage points (18 percent) more likely ever to be observed in the labor force. I cannot reject that these coefficients are equal, consistent with women on the margin of entering teaching being on the margin of entering the labor force at all. Because failers were able to retake the examination, passing on the first attempt raised the probability of ever clearing the cumulative threshold by 47 percentage points, which implies that the effects of certification itself are approximately twice the size of the reduced-form effects. I report intent-to-treat estimates throughout.

I find that the labor effects persist for more than two decades. By 1930, marginal passers were 6.6 percentage points more likely to be teachers and 7.2 percentage points more likely to be in the labor force. Even by 1940, 17 to 24 years after the examination, they remained 8.9 percentage points more likely to be teachers and 8.6 percentage points more likely to be in the labor force. Pooling the 1930 and 1940 censuses, women are 33 percent more likely to be teachers, and 15 percent more likely to be in the labor force in the long run. There is also suggestive evidence that women had higher wage income ($91.5, or 35 percent) by 1940, which is explained entirely by their greater extensive-margin labor force participation. This is strong evidence that access to teaching generated persistent careers for women.

I next examine for whom and under what conditions the labor force effects persisted. First, long-run employment was driven almost entirely by persistence in teaching as opposed to entry into other occupations, consistent with the similar magnitudes of the teaching and labor force effects across census years. Second, roughly half of the long-run labor force participation effect (measured in 1940) was concentrated among women who continued to work after marriage. Teaching did not crowd out married women's work in other occupations. If anything, there is suggestive evidence of additional entry into other white-collar work. Finally, education compounded the effects of initial career access. Marginal passers had 0.3 more years of schooling by 1940, with the effect concentrated at the two-year college margin and among women who remained teachers through 1930 and 1940.

Although my sample is not large enough to support extensive heterogeneity analysis, two suggestive patterns emerge. First, whether women could sustain these careers after marriage depended on local restrictions on married women's work; I build on data collected in Yang and Brough (2026) to identify Illinois counties that likely operated teacher marriage

bars, which were school-board policies that restricted married women from working as teachers. Marginal passers persisted in the labor force after marriage only in counties without a marriage bar. In counties with a marriage bar, the married-teacher effect disappeared, and there is little evidence that women substituted into other work. These results suggest that marriage bars kept women out of the labor force, and that women's desire to continue their careers after marriage may have been stronger than aggregate labor force participation rates suggest.

Second, I find that the labor force participation and income effects are concentrated among farmers' daughters, who were disproportionately rural and had the fewest alternative white-collar opportunities. For women from nonfarmer households, access to teaching did not newly draw them into the labor force, but shifted which occupation they entered in the labor force. Although suggestive, these patterns highlight the disparate impacts of local labor-market conditions and family-compatible institutions in shaping women's career and family outcomes.

Access to teaching also changed women's family formation by both delaying and deterring marriage. Marginal passers were 6.8 percentage points more likely never to have married by 1930 and 5.2 percentage points more likely never to have married by 1940, consistent with marriage delay. They were also 4.4 percentage points less likely to have any children by 1940, and delayed their first birth by approximately 0.6 years. The number of children observed in the household by 1940 declined by 16 percent, with imprecise evidence suggesting that both intensive and extensive margins are contributing factors. Despite these large changes in marriage and fertility, I find little evidence that access to teaching changed the characteristics of whom women married. Husbands did not have higher income, education, or income rank relative to women's fathers. Access to a professional career therefore changed whether and when women formed families, but did not generate gains in the marriage market; instead, the economic gains were concentrated among women themselves.

I connect my estimates to aggregate changes in LFP and in fertility across birth cohorts to show that expanded access to professional work can account for a meaningful share of these broad demographic shifts. Under a very conservative calculation, I show that additional entry into the labor force, driven by marginal exam passers (those within 3 percentage points of the passing threshold), explains roughly 1 percent of the rise in prime-age labor force participation and 0.2 percent of the decline in completed fertility among women in downstate Illinois. Extending my estimates to the growth of teaching in rural America, I find that additional early-career entry into teaching accounts for roughly 8 percent of the rise in rural women's labor force participation and just under 1 percent of the decline in fertility.

My results contribute to a growing literature studying how women's economic opportunities affected labor market and family outcomes during the early-to-mid twentieth century. Seminal work exploits variation in female labor demand generated by World War II mobilization to study the effects of women's employment on wages (Acemoglu, Autor and Lyle, 2004), labor supply (Goldin and Olivetti, 2013), fertility (Doepke, Hazan and Maoz, 2015), and the social preferences of the next generation (Fernández, Fogli and Olivetti, 2004).[3] More closely related to my work is a recent set of papers that use linked census microdata to study women's entry into specific occupations, including women's entry into clerical work as the use of the typewriter expanded (Rashid, 2025), their exit from telephone operating as automation displaced workers (Feigenbaum and Gross, 2024), and their entry into nursing after new nursing schools opened (Bald, 2025). These papers exploit temporal and geographic variation in women's economic opportunities, which can combine individual effects with peer, network, and general equilibrium effects.[4] In contrast, I study a shock that determined women's access to an occupation at the individual level, comparing women with different economic opportunities while holding labor markets and education institutions fixed, as well as the norms and preferences that may evolve with them.

Relative to existing work, my setting shows that being a career woman is not predetermined and can be induced by early career opportunities into sustained labor force participation lasting more than two decades. This is unlike settings in which women left the labor force after an occupational opportunity disappeared (Feigenbaum and Gross, 2024; Li, 2025; Rose, 2018) or upon marriage. It also differs from Bald (2025), who finds that expanded access to nursing created persistent careers but primarily diverted women from other occupations. Instead, access to teaching drew marginal women into the labor force in early adulthood and kept them there in the long run.

In this way, my work speaks to both the literature on women's entry into the workforce and to the forces and institutions that sustained (or limited) women's labor force participation over the life course, especially their persistence after marriage. Prior work emphasizes changes that made married women's employment easier to combine with family life, including exposure to working women (Fernández, Fogli and Olivetti, 2004), improvements in

[3]Recent work has questioned whether World War II mobilization led to women's increased participation in the labor force by 1950. Using more granular data on the geography of war work, Rose (2018) shows that locations with the largest growth in war work did not exhibit sustained increases in female labor force participation by 1950. In a separate exercise, Briskin (2026) revisits Acemoglu, Autor and Lyle (2004) and Goldin and Olivetti (2013) using the full-count census and shows that many of the results, originally derived from the 1 percent census sample, are not robust to using the full-count data.

[4]One might think this distinction is particularly relevant if exposure to working women itself changes perceptions or norms around women's work (Fernández, Fogli and Olivetti, 2004; Fernández, 2013), or for market-mediated outcomes such as marriage.

maternal health and infant feeding (Albanesi and Olivetti, 2016), and the end of marriage bars (Goldin, 1988; Kim and Tsao, 2025; Yang and Brough, 2026). I find that much of the long-run persistence in the labor force came from married women, but only where marriage bars did not restrict their work. This suggests that married women's demand for career-like work was stronger than their participation rates show, in part because discriminatory local policies limited their opportunities.[5]

I also contribute to research on how women's economic opportunities shape family formation. Existing work has considered the effects of contraception (Goldin and Katz, 2000, 2002; Bailey, 2006), education (Black, Devereux and Salvanes, 2008; McCrary and Royer, 2011; Geruso and Royer, 2018; Doxey, Karger and Nencka, 2026), and labor-market changes (Vidart, 2026; Rashid, 2025) on marriage, fertility, and the timing of childbearing. A related literature studies how women's economic opportunities affect whom they marry (Shenhav, 2021; Rashid, 2025). I show that access to teaching reduced marriage and fertility, but did not lead women to marry husbands with more education, higher occupational standing, or higher earnings. Instead, the gains from professional access accrued primarily through women's own employment.

Finally, I contribute to the literature studying teacher labor markets and certification systems. This work often centers on how women's outside options (Corcoran, Evans and Schwab, 2004; Bacolod, 2007) or state certification systems (Angrist and Guryan, 2008; Orellana and Winters, 2025) shape teacher quality and supply. Two notable exceptions study teachers' own outcomes using individual-level certification-exam thresholds. Deneault, Riehl and Zou (2026) find that narrowly failing a Texas exam delays entry into teaching at a cost of roughly $10,000 over a decade, and Tsao (2025) uses Kentucky's entry-exam cutoff to estimate the pay and working-conditions rents teachers earn over their next-best jobs. Relative to these papers, the alternative to teaching in my setting was often not to enter the labor force, highlighting the importance of teaching as a central female occupation in the early twentieth century.

The remainder of the paper is organized as follows. Section 2 provides institutional background on teaching and the Illinois certification exam. Sections 3 and 4 describe the data, record linkage, and sample selection. Section 5 presents the regression discontinuity design and how its identifying assumptions are examined in this setting. Sections 6 and 7 present effects on labor and occupation outcomes and on family formation. Section 8 examines heterogeneity by marriage-bar status and rurality, and Section 9 presents robustness checks. Section 10 considers the extent to which these estimates can account for aggregate changes

[5] I also find suggestive evidence that the marriage bar induced different family-formation responses among affected women, providing new evidence on the non–labor market consequences of the marriage bar.

in LFP and fertility across birth cohorts, and Section 11 concludes.

2 Institutional Background

2.1 Teaching and Women's Work

Teaching was among the first occupations to open to women on a large scale, and had become a female-dominated profession by 1860. It feminized as the spread of public common schools raised the demand for teachers while urbanization and industrialization drew men toward higher-paying work (Ogren, 2005). Reformers in the common-school movement, which advocated for free, taxpayer-funded public schools in the mid-nineteenth century, encouraged hiring women, arguing that women had a natural predisposition to teach because it was in their nature to rear children (Ogren, 2005). Teaching was therefore distinct from other work for women. It offered entry into a respectable white-collar profession at a time when nearly all other white-collar work was closed to women. For the next sixty years, teaching remained the dominant white-collar occupation for women. By 1920, teaching accounted for roughly 25 percent of all women's white-collar work and 10 percent of women's total employment, making it the single largest white-collar occupation for women.[6]

Teaching also paid comparatively well by the standards of women's employment in this era. Using median wage income in the 1940 census, teaching ranks highest among the principal occupations open to women (Figure 2). Teaching's relative pay was more modest in the 1920s, although still reasonable. In 1922, the average downstate Illinois teacher earned $1,046 over a school term of roughly eight months, or about $131 per month in session. To benchmark this against contemporary alternatives, I digitize the pay scales in the 1922 classification of the Illinois State Civil Service Commission, which records monthly salaries alongside stated education requirements for every state job class (Illinois State Civil Service Commission, 1922). A teacher's in-session pay matched or exceeded the starting pay of the clerical classes requiring a high school education—stenographers, department clerks, bookkeepers—though spread over twelve months it fell toward the middle of that range (Appendix Figure B1). However, geography also determined whether this other, potentially comparable white-collar work was available. For rural women, teaching made up 51 percent of white-collar work, relative to 15 percent in urban settings (Appendix Figure B2).

Beyond pay, teaching underwent several changes between 1915 and 1940 that increasingly professionalized the occupation. These changes made teaching uniquely career-like compared

[6] Author's own calculation using the 1920 full-count census, where white-collar work is defined as 1950 occupation codes 0–99 and 200–399.

with other occupations available to women. In 1915, teachers across Illinois outside Chicago gained access to a statewide pension system. School districts also increasingly adopted salary schedules that rewarded experience and additional training, whereas higher teaching credentials were typically associated with higher pay. Teachers were required in their first years, and more senior teachers were encouraged, to update their skills and qualifications through teacher institutes, summer coursework, and additional training in teacher colleges (known as normal schools). Together, these institutions created incentives to remain in teaching and invest in occupation-specific skills.

At the same time, teaching (as well as other white-collar work) was considered incompatible with family life for many. Throughout this period, many school boards operated, either through formal or informal channels, the marriage bar, which limited women's ability to work upon marriage (Goldin, 1988). In Illinois, by 1942, around one-third of school districts responding to a survey reported marriage as potential grounds for dismissal. However, not all places operated the marriage bar, and how it operated depended on settings; for example, often school boards enforced the "hire bar," which prohibited the hiring of new married teachers more strictly than the "retention bar," in which teachers were dismissed upon marriage. Whether school systems operated marriage bars also depended on the local supply of teachers, which was often more constrained in rural areas.

2.2 The Illinois State Examination System

Illinois introduced its statewide examination system in 1914, replacing a fragmented system of county-operated exams (State of Illinois, 1915). The state system was part of a broader effort to improve teacher quality by centralizing and professionalizing teacher qualifications. Officials described the certificating law as the beginning of a "new and a better era" in Illinois education, and the 1914 law as a major step toward "keeping out the incompetent" and attracting competent teachers (Illinois Department of Public Instruction, 1919, 1923, p. 3).

The 1914 law created third-grade, second-grade, and first-grade elementary certificates.[7] The second-grade certificate, which is the primary credential in my empirical setting, was valid for two years and renewable on evidence of successful teaching, professional training,

[7]Certificate "grades" denote tiers of the credential, not the school grade levels taught.

and professional growth (State of Illinois, 1915, p. 8).[8] The exam therefore created a regular, statewide route into teaching, but it was not meant to be a complete substitute for further training.

The examination system remained central between 1916 and 1923 (Figure 3). About half of the new elementary certificates issued to women outside Cook County were obtained through examinations. Officials recognized that raising standards too quickly could restrict the supply of teachers, especially during and after World War I. Ultimately, in 1924, the board of examiners raised the passing threshold for the second-grade certificate from 75 to 80 percent, and in turn the number of new elementary teachers entering by exams fell precipitously.

2.3 Exam Grading and Administration

Examinations were held two to three times each year at each county seat. Applicants were assigned an identification number, which they wrote on their exam papers instead of their names, so graders could grade exams anonymously. Applicants had to pay $1 (equivalent of $19–$30 today) to write. Written exams were mailed to the Illinois State Examining Board, which graded every exam centrally and recorded scores in a ledger that was then shared with counties.[9] The formal examination rules specified sealed envelopes for each half-day's questions, separate desks for applicants, no communication among applicants, and central return of sealed manuscripts to the Examining Board (Illinois State Examining Board, 1940, pp. 7–9).

Candidates sat for exams in 13 subjects (State of Illinois, 1915). Appendix Figure B4

[8]The third-grade certificate was a one-year certificate available to individuals failing the second-grade examination, renewable only once for an additional year; some people applied directly to the third-grade certificate, but this accounted for less than 10 percent of applications. In 1919, this certificate was abolished, leaving those who failed the exam eligible for an "emergency certificate," which was nonrenewable and valid for one year. Neither the third-grade ("provisional certificate") nor the "emergency certificate" offered a career-like pathway to teaching because of their nonrenewable nature.

[9]The board would enlist the help of educational professionals to assist in grading, but additional details on who specifically was recruited to grade exams are not known; though often it seems that these were education professionals in Springfield. How graders were assigned to exams is also unknown. The Department of Public Instruction described the graders and the anonymity of grading as follows: "The grading of the papers is done by persons who have proven; (1) a liberal scholarship in the subject assigned to them; (2) a successful experience in teaching those particular subjects, and (3) a quality of mind that fits them to determine the merit of a written discussion in their subject. It may be well to state further that no grader can have any knowledge of the author of the paper he is grading. He knows it only by the application number and the county from which it comes. There is no possibility of favoring special candidates. There is very little probability of an error in grading for the reason that during the last eight years we have been organizing a force of graders whose work has proven their ability to grade the papers with a minimum number of errors. We admit that an occasional error is possible and will be possible so long as human effort is not infallible" (Illinois Department of Public Instruction, 1923) (Appendix Figure B3).

reproduces sample questions from eight of these subjects. Between 1916 and 1923, to pass, a candidate's cumulative average across all subjects had to meet the passing threshold. In principle, candidates were eligible for a certificate only if they had no subject grade below 60 percent.[10]

Candidates could resit the examination at a later sitting, and retaking was common; roughly 40 percent of test-takers sat the exam more than once. Individual subject scores above 75 percent are carried over to subsequent exam attempts, meaning a returning candidate would not have to retake subject exams in which she already achieved 75 percent. This carry-over system makes it challenging to compare exam results from retakes across candidates because candidates may be taking a different number of exams.

Passing the certification exam was not sufficient to receive credentials. County superintendents received ledgers of exam scores (including the names and scores of all county applicants) and then had discretion in hiring teachers. Often this included imposing additional screens. For example, teaching applicants also had to prove their good "morals" through three character references.

Therefore, because not all passers were offered teaching positions, and some who failed found alternative routes into the profession, obtaining the average on the first attempt is best understood as a *fuzzy* treatment for access to teaching, and my regression discontinuity estimates capture intent-to-treat effects.

2.4 Alternative Pathways to Teaching Credentials

Normal schools and colleges could serve as alternative credentials for students who either did not take or did not pass certification exams. Normal schools were distinct from colleges. They were teacher-training institutions, often offering one- or two-year programs, and some offered more extensive 4-year BA options that combined academic coursework with pedagogy and practice teaching.

In Illinois, students who attended two or more years of normal school or college could receive credit for that attendance and become eligible for the second-grade teaching certificate without taking the teacher exam. Completing normal school could also make candidates eligible for more advanced school positions, such as high school teacher, specialist teacher, and principal, routes not open to candidates who gained credentials through the exam alone.

Provisional and temporary certificates were another route into teaching, but they offered less security. Between 1916 and 1918, the third-grade certificate (or "provisional certificate") existed specifically for people who failed the second-grade examination. It was valid for

[10] In 1924, certification requirements changed, requiring a cumulative average of 80 percent across exams and a minimum of 50.

one year, eligible for renewal for an additional year, but not for a second renewal. In 1919, it was eliminated, leaving only the "emergency" certificate, which was nonrenewable, valid for only one year, and open to those not writing the exam. Officials described these temporary certificates as necessary measures to ensure a sufficient supply of teachers during the transition to higher-quality standards, but measures that would ultimately be phased out as teachers increasingly gained credentials through college routes (Illinois Department of Public Instruction, 1919, 1923, p. 3).

These alternative pathways to teaching provide important context for the RD design. Although failing the exam did not make teaching impossible, it made entry more costly (through normal school) or temporary (through emergency or, for some years, third-grade certificates).

3 Data

3.1 Examination Ledgers

My primary data source is examination ledgers obtained from the Illinois State Archives (Illinois State Archives, 1923) (Appendix Figure B5 shows an example page). These ledgers record results from teacher certification examinations administered across all Illinois counties except Cook County, which was exempt from the system. Every ledger corresponds to one exam sitting (unique at the month-year level), and its entries record the candidate's name, age, city of residence, years of education at the time of the examination, the date and county of the exam, and her exact score on each of the 13 subject exams.

Several information fields are typewritten, including names, ages, residence, and baseline education. The subject grades, however, were handwritten. I digitize both the typed and handwritten fields using a large language model (Gemini 3.0-flash). Because each page is organized into grids with faint row and column boundaries, I first preprocess the page images to align and superimpose the column and row boundaries before passing them to the model; this step substantially improves transcription accuracy on the tabular score data.

3.2 Census Data

I link examination records to the full-count censuses of 1910, 1920, 1930, and 1940 (Ruggles et al., 2024). First, the 1910 and 1920 censuses (depending on the exam year) provide baseline household characteristics for candidates, including urban versus rural residence, the household head's occupational score, the mother's working status, and whether any woman in the household was in the labor force. The 1920 (for exams taken in 1916–1919), 1930, and

1940 censuses provide the relevant short- and long-run labor and family outcome variables. In Section 3.5, I discuss how I link the exam data to the census and link individuals across censuses. Below, I describe key variables in more detail.

- **Labor outcomes:** Teacher (occupation 93); other white-collar occupation (occupations 0–99 excluding 93, and 200–399); other occupation; labor force participation; and own income (reported only in 1940)
- **Marriage:** Marital status (specifically, never married)
- **Fertility:** Presence and age of children in the household (used to estimate age at first birth)
- **Spousal characteristics:** Husband's education, occupational score, income (available in 1940)

3.3 Measuring Marriage and Fertility

Below, I describe in more detail how I construct family formation measures, as this often requires combining information from across censuses.

Marriage: To measure marriage over time, I focus on measures of never married in the 1930 and 1940 censuses. I also explore other marital status outcomes, like widowhood, divorce, and married with or without a partner present. The advantage of focusing on never married is that it captures the extensive margin of whether a woman has ever been married.

Number of observed children: I measure fertility as the estimated number of children in the household in the 1930 and 1940 Censuses, which I call the "number of observed children" for short. To do this, I restrict the analysis to women I observe in both the 1930 and 1940 censuses. I count the number of distinct own children a woman is observed with across the 1930 and 1940 censuses. I identify distinct children from their names and birth years so that a child appearing in both years is counted once.

I limit to women whom I observe in both 1930 and 1940, as measures of fertility from any single census risk miscounting the number of children ever born, since new children enter the household while older children leave. This approach helps overcome this entry versus exit issue with children. However, this measure has limitations. It does not recover children who have died, were born and exited the household before 1930, or who were born after 1940. By 1940, women are on average 38, an age close to, but not at, completed fertility (Appendix Figure B6 plots the age distribution of applicants in 1940).

Childless: I define women as having no children if they are observed in both the 1930 and 1940 censuses and observed with none of their own children in their household. As referenced above, by 1940 women are approximately 38, which, while too young to be complete fertility,

is a reasonable age to assume that women will not have their first child after this age.

Age at first birth: I define women's approximate age at first birth among women I observe in both 1930 and 1940 with a child. I identify the oldest child of all children reported in either the 1930 or 1940 census. I define a woman's age at first birth as her age (in the census in which the oldest child is first observed) minus the age of the oldest child.

3.4 Data on Marriage Bars

I supplement my analysis with data on the geography of teacher marriage bars. I first use data on US cities that operated the marriage bar as reported in my complementary work studying marriage bars (Yang and Brough, 2026). Because our paper tracks the existence of the marriage bar in select large cities (most not in Illinois), I adopt our data-collection approach to the Illinois context. Our original data identified 18 distinct Illinois counties with cities that had the marriage bar. My expanded data build recovers an additional 10 counties with reports of cities adopting the marriage bar. While no systematic data source tracks the prevalence of the marriage bar, a 1942 survey of education in Illinois found that about one-third of reporting school districts operated marriage bars. This is broadly consistent with my finding that 28 of 102 counties (27 percent) operated marriage bars.

3.5 Record Linkage

3.5.1 Matching Back to Baseline Censuses

I match each examination candidate to the most recent full-count census preceding her exam—the 1910 census for exams taken in 1916–1919, and the 1920 census for exams taken in 1920–1923. In cases where I do not find a match in the 1920 census for the 1920–1923 cohort, I attempt to match to the 1910 census. I use individuals' first name, middle initial, last name, year of birth (inferred from reported age at exam), and residence as identifiers. Before matching, I restrict the pool of candidate census records to women enumerated in Illinois (excluding Cook County) in the 1910 or 1920 census.

Deterministic and Probabilistic Matches. I begin with a modified deterministic matcher that follows the ABE–Extra Information (ABE–EI) methodology (Abramitzky et al., 2021, 2025). The matcher cascades through match rounds that become progressively permissive based on name agreement. I start with exact name matches, and then move to phonetically standardized (sound-adjusted) names, and then to names that agree closely under the Jaro–Winkler string-similarity metric. In every round, I use extra information—middle initial and county of residence—to help distinguish (create uniqueness) among same-name candidates

and require uniqueness within a 2-year birth-year window. I keep a candidate whenever a round produces a unique link. Each round attempts only the candidates still unmatched by the stricter rounds before it.

I supplement the deterministic matcher with a probabilistic record linkage using the Python package `splink` (Linacre et al., 2022), which implements Fellegi and Sunter (1969). This approach is useful in settings where the set of available identifiers differs across observations, for example, if data fields are missing, and where measures are continuous. When linking each examination candidate to the 1910 or 1920 census, I use first and last names, middle initial, birth year (inferred from age at exam), and the geographic distance between her county of examination and her census place of residence. This differs from the deterministic approach, which matches based on county exactly (as opposed to geographic distance). To limit false links, I accept a pair only when every nonmissing field agrees to within a set tolerance, and the link is one-to-one across both files. I only keep best matches whose match probability exceeds 0.90. In cases where deterministic and probabilistic matchers disagree, I keep the deterministic matches. Together, these procedures link just over 70 percent of exam candidates to a baseline census record.

3.5.2 Matching Forward to Outcome Censuses

To observe postexam outcomes, I link women across censuses using the Census Tree crosswalks (Buckles et al., 2025), a set of high-coverage links between full-count censuses constructed from machine-learning matching and crowd-sourced genealogical records. The Census Tree project is unique and necessary for this project because it matches women across census waves. Because my records include only maiden names, I rely exclusively on these genealogy-based links to track women into marriage. This is necessary because protected identification keys (PIKs) and Social Security numbers (SSNs) are not available for linking individuals across the censuses at this time.

Two concerns arise with forward matching. First, the pool of women who successfully match to future censuses may differ from the full exam-taking population. Matched women are somewhat more likely to have taken the exam in earlier years, live in rural areas, and pass the exam. To address this, I construct inverse probability weights (IPW) so that the women who match to subsequent censuses resemble the baseline matched pool on observables (estimates using this IPW and other exercises are discussed in Appendix D). Second, the match rate may differ by treatment status. Reassuringly, I do not see that the match rate changes discontinuously at the RD threshold, as discussed in Section 5.3.

4 Sample Construction and Descriptives

In this section, I describe my main sample and characterize exam-takers' individual and childhood circumstances, and how they compare to other groups of similar-aged women in Illinois. This exercise helps me assess the generalizability of my population.

4.1 Sample Construction

My data includes 48,029 unique applicants between 1916 and 1923. I then restrict this set of applicants to those who have 13 valid exam scores (36,527), to those with female names (29,042), and to those aged 17–23 when they take their first exam (26,612). I match this group to a baseline census, achieving a 70 percent match rate.

4.2 Exam Takers Compared to Similar-Aged Women

Table 1 compares exam applicants' childhood-household characteristics to several comparison groups, including Normal School enrollees, general Illinois women age 25 and under, and a pool of women who were eligible to sit the exam based on education.

Relative to Illinois. Relative to the general Illinois population (columns 2 and 6), exam applicants look more rural and lower in socioeconomic status (SES) on household-head occupation score.

Relative to Illinois, Excluding Cook. This comparison, however, is driven almost entirely by Cook County (which includes Chicago). Once Cook is excluded (columns 3 and 7), overall differences decrease significantly. Teacher applicants and downstate Illinois women are identical on household-head occupation score (a difference of 0.00 points), though they remain significantly more rural (+13 percentage points) and more likely to come from farm households (+10 percentage points). Household heads are a bit more concentrated in jobs that require higher education (+0.7 points) and are slightly more likely to have professional jobs (+0.5 percentage points). They are also, unsurprisingly, significantly more likely to be enrolled in school (+12 percentage points).

That teacher applicants are not more different from downstate women is somewhat surprising. To be eligible to teach, women had to complete at least 10 years of schooling (in practice, applicants had on average 11 years of schooling), putting them in the top half (based on 10 years) or top third (based on 11 years) of the education distribution for their birth cohorts.

Relative to Exam-Eligible Groups. Therefore, a second, potentially more natural comparison is to the subset of similar-aged women in downstate Illinois whose reported completed schooling in the 1940 census is at least 10 years (i.e., women who would have been eligible to sit the exam). Because linking 1910 records forward to 1940 is selective, I weight this group by the inverse probability of matching to 1940, estimated from 1910 characteristics to make the matched sample representative of the underlying 1910 population.[11] Against this group of eligible women, exam applicants are much more *negatively* selected on socioeconomic status. Their household heads have occupation scores 2.2 points lower, are 20.6 percentage points more likely to be farmers than the eligible pool, and have heads of households in occupations that require less schooling (-3.6 points).

Together, these facts suggest that the exam-based teaching system drew highly educated women disproportionately from more modest, rural households. This made them somewhat more comparable to non-Cook Illinois women as a whole than expected, but negatively selected relative to non-Cook Illinois women whose education levels would have made them eligible to be teachers.

Relative to Normal School Enrollees. Finally, I compare exam applicants with women who pursued a teaching credential by enrolling at a normal school. To build this comparison group, I digitize enrollment rosters from the Illinois State Normal School covering 1912–1923, which record roughly 4,000 enrollees; I link 2,390 of these women to the 1910 census.

Table 1 reports childhood-household characteristics for both groups (columns 5 and 9), where the difference column is a coefficient on a teacher-applicant indicator from a regression controlling for birth-year fixed effects. Exam applicants and Normal School attendees are similar in age and race. Exam applicants have household heads whose occupation scores are 1.6 points lower than those of Normal School attendees (on a base of 19.9) and are 9.3 percentage points more likely to be farmers (on a base of 40.0 percent); exam applicants are also somewhat less urban, by 9.6 percentage points (relative to 29.5 percent among Normal School attendees). Normal School students were also more than twice as likely to come from households headed by a professional (6.2 percent versus 2.7 percent among exam applicants), and their household heads had education scores 5 points higher than exam applicants. This pattern is consistent with exam-based credentials being a more accessible way into teaching for rural, lower-SES women than Normal School attendance, which required tuition, travel, and accommodations away from home.

[11] Defining eligibility from 1940 schooling is mildly posttreatment for the applicants who reach ten years partly because of teaching; because the comparison is on predetermined childhood-household characteristics rather than on schooling itself, this does not drive the pattern.

5 Methodology

5.1 RD Design around the First-Attempt Cumulative Score

A candidate met the cumulative score threshold if her average score across the 13 subject exams reached 74 percent.[12] My RD design exploits this cutoff to compare women who scored just above and below this threshold.

I measure each candidate's score using her first exam attempt, because scores on the first attempt are more exogenous than scores on subsequent sittings. This is true in this setting for two reasons. First, candidates could preserve, or "carry over," subject scores above 75 percent to subsequent sittings, so the number of exams and the cumulative points a candidate needed on later attempts depended on her performance in earlier ones. Second, the decision to retake the exam is endogenous, and selective retaking can bias estimates (Jepsen, Mueser and Troske, 2016).

I estimate the following equation:

$$Y_i = \alpha + \tau \cdot \mathbf{1}(X_i \geq 0) + \beta_L X_i \cdot \mathbf{1}(X_i < 0) + \beta_R X_i \cdot \mathbf{1}(X_i \geq 0) + \varepsilon_i \tag{1}$$

Here, Y_i is an outcome of interest, for example, whether woman i is observed as a teacher in a subsequent census; X_i is the running variable, measuring the distance of her first-attempt average score from the effective passing threshold of 74 percent, in percentage points (recentered to 0); τ, the parameter of interest, captures the discontinuous jump in the outcome at the threshold; and β_L and β_R capture the slope of the regression function below and above the threshold, fit using a nonparametric local linear function. The main specification includes no additional controls. Because not all passers ultimately taught and some who failed found other routes into teaching, τ is an intent-to-treat (ITT) effect of crossing the passing threshold.

I implement the bias-corrected estimator of Calonico et al. (2017), using a local-linear specification (polynomial order one) (Pei et al., 2022) with a triangular kernel and the MSE-optimal bandwidth (the bandwidth that minimizes the mean squared error). I use a triangular kernel throughout because it down-weights observations far from the threshold, which limits the volatility that can otherwise arise in small-sample RD estimates. In Appendix D, discussed in Section 9, I show that my results are robust to these and other specification choices as well as to the inclusion of covariates.

[12] The statutory passing average was 75 percent. However, in practice, the Examining Board recorded averages as whole numbers and rounded up to the next highest integer, meaning that an average between 74 and 75 was recorded as 75, one between 75 and 76 as 76, and so on. A computed average of 74 percent was therefore sufficient to meet the cumulative score threshold (Appendix Figure B7).

The design relies on two assumptions. First, that there is smoothness in the running variable across the threshold. Several features support this. First, because candidates were scored from 0 to 100 on 13 exams, my scores range from 0 to 1,300, which is a relatively continuous distribution compared with a more discrete setting (for example, integers between 0 and 100). Moreover, candidates wrote anonymized exams that were graded centrally, leaving little scope for one to manipulate her own score. However, there is still the case of grader manipulation, which I address in Section 5.2.

Second, absent passing, the potential outcomes must evolve continuously across this threshold. Although this continuity assumption is untestable, I assess it indirectly by testing whether predetermined characteristics change discontinuously at the cutoff (Section 5.3). The balance table (Table 2) serves both as a diagnostic and as a useful summary statistics table.

5.2 Identifiable Grade Bumping and Implications for Smoothness

I identify one key, but addressable, threat to the RD continuity assumption. Recall that in principle, applicants were required to meet both the average threshold (74 percent) and a minimum threshold—specifically, that no single subject of 13 can fall below 60. This created incentives for examiners to selectively bump one's minimum subject score to exactly 60 for candidates for whom this change would allow them to pass the exam. Because these bumps added points only when they moved candidates past the passing margin, they created excess mass just above the cutoff and a corresponding deficit to the left of the passing margin. As a result, this pattern creates a visible and discontinuous break in the density of first-attempt exam scores. However, this can be corrected by removing individuals whose minimum score is exactly 60.

To observe this, see Figure 4, panel (a). Here, the histogram plots the number of applicants by first-attempt average score relative to the passing threshold (centered to 0, in percentage points). The gray shaded regions represent the share of individuals in each histogram bin whose minimum score (across the 13 exams) is any value other than 60; the purple shaded regions represent the share of individuals whose minimum exam score is exactly 60.

The counterfactual curve (plotted in green) fits a smooth skew-normal distribution of cumulative scores to the data. It is clear that there is significant excess mass occurring just to the right of the passing threshold, which extends relatively far to the right side of the distribution, and a hollowing of mass to the left. In the full sample, the Cattaneo, Jansson and Ma (2018) density test strongly rejects the null of no manipulation ($p < 0.01$; panel (c)). However, once I remove individuals whose minimum score is exactly 60, the distribution

appears smooth (panel (b)). Importantly, this exclusion is applied symmetrically around the cutoff, meaning that the interpretation of the RD is unchanged and the design still compares those who marginally passed versus failed to meet the passing cutoff, among the applicants who remain. This does not break continuity in the running variable, because all cumulative scores around the cutoff are represented regardless of whether I include individuals whose minimum score is 60. Using the Cattaneo, Jansson and Ma (2018) density estimator, which implements a McCrary-style manipulation test (McCrary, 2008) but in a local-polynomial framework, on women aged 17–23, I cannot reject density smoothness ($p = 0.50$).

To interpret what kind of manipulation likely generates this pattern, I simulate several simple grading regimes in Appendix Figure B9. The data are most consistent with a regime in which candidates first receive grades on all 13 exams. Then examiners bump sub-60 scores to exactly 60 only for candidates for whom this change would allow them to pass the exam (panel (a)). This regime produces a density break in the full sample—excess mass just to the right of the cutoff, with a smaller, more spread-out shortfall to the left of it. This is because the regime increases the cumulative scores of individuals only when doing so lets them meet the average. However, this excess mass is identifiable and can be excluded from the distribution by dropping observations with a minimum subject score of exactly 60.[13]

By contrast, direct manipulation of the average score at the cumulative cutoff (panel (b), left) results in a narrow excess mass occurring exactly at the pass threshold, with a corresponding "hollowing" in the distribution to the left of the cutoff. Excluding 60s does not help to achieve a smooth density. Similarly, a regime in which the probability that one's minimum is bumped up to 60 is distributed uniformly (i.e., not targeted to passers) (panel (b), right) is also inconsistent with the true data. In this regime, everyone's grades are uniformly higher in a way that produces no excess mass at the pass threshold, and excluding those with a minimum of 60 does not change the shape of the distribution.

5.3 Summary Statistics and Covariate Balance

Table 2 presents RD estimates for baseline characteristics. First, I find no evidence that women who marginally meet the cumulative score criteria differ from marginal failers in urbanicity, age, or household head occupational scores. Reassuringly, I also do not find any difference in probability of matching to subsequent censuses; the by-census estimates range

[13]Given heaping at the value 60, however, the newly recovered distribution lies below the unmanipulated distribution. Heaping in the exam scores is evident in Figure B10, in which it is clear that 60 is the most common single grade on a subject test. Figure B11 decomposes this further, highlighting that when one removes individuals whose minimum is exactly 60, this removes the excess mass (manipulation)—the region shaded in purple—but it also excludes the additional people who obtained a minimum of 60 without manipulation—the region shaded in gold.

from −0.04 to +0.00, and the pooled measure of "matched to any subsequent census" is −0.006 ($p = 0.71$). The one possible exception is that marginal passers appear to be more likely to be living in their childhood household in their baseline census (observed in 1910 for exams taken in 1916–1919, and in 1920 for exams taken in 1920–1923). However, these magnitudes are relatively small (+2–3 pp on a baseline of 0.91–0.94). In Appendix C, I also show that outcomes predicted from these preexam characteristics do not change discontinuously at the threshold for any of the main outcomes.

This table also helps characterize the marginal population, among whom the effects of access to teaching will be identified. On average, the marginal applicant is young (age 18.2) at the time of her exam and has completed 11 years of schooling. She is disproportionately from a rural household (80 percent) and a farm household (48 percent).

5.4 First Stage Effects

Figure B13 plots the relationship between meeting the cumulative score on one's first attempt and the probability of ever clearing the average for my main estimating sample. Passing the average on the first attempt raises the probability of ever passing by about 47 percentage points (robust $p < 0.001$). Among all women who ever clear the average, roughly 60 percent do so on their first attempt. Throughout, I present ITT estimates; however, to scale the presented ITT estimates accordingly, it is appropriate to multiply by 2.1 to recover the LATE.

6 Labor Supply Outcomes

In this section, I first investigate the relationship between marginally passing and women's employment as teachers as well as their labor force participation as a whole. I also investigate income, which is observed only in 1940.

6.1 Teaching

Figures 5 and 6, together with Table 3, report the immediate and long-run effects on teaching. First, I show that marginally passing the certification exam increases the probability of being observed in subsequent censuses as a teacher by 13.3 percentage points, or 34 percent on a baseline of 39 percent (Table 3, column 1).

This estimate is strong evidence of a large first stage of passing the certification exam on teaching. However, the magnitude needs careful consideration. The decennial nature of the census means that the full trajectory of women's labor market outcomes is not observed.

Ideally, whether a woman was ever employed as a teacher would be observable on an annual basis, whereas the decennial censuses provide only snapshot views in 1920, 1930, and 1940. This measure misses, for example, women who pass the exam in 1923, teach for several years, but exit teaching before the 1930 census.

To investigate how women's teaching evolves over time, Figure 6 plots the effects of certification on teaching in the short, medium, and long run.

To identify the short-run effects, I focus on exam cohorts 1916–1919 and their observed teaching status in 1920. Passing increases the probability of reporting a teaching occupation in the 1920 census by 18.2 percentage points ($p < 0.01$; Table 3, column 3). The response is even larger among the recent 1918–1919 cohorts, whom I observe approximately one to two years after the exam. Passing raises teaching by 20.2 percentage points ($p < 0.01$; column 2).[14]

Over the long run, this initial teaching effect dampens but remains economically meaningful. Passing raises teaching by 6.6 percentage points in 1930 ($p = 0.06$; below-cutoff mean 24.5 percent; column 4) and by 8.9 percentage points in 1940 ($p = 0.01$; mean 14.1 percent; column 5). Pooling the 1930 and 1940 estimates, marginal passers are 8.5 percentage points (33 percent) more likely to be teachers in the long run (Table 8, panel B).

This decay rate is unsurprising given general teaching retention in this period (Appendix Figure B14). Among women who are teachers in 1920, 34 percent are still teachers in 1930, and 29 percent are still teachers in 1940. Thus, although there is significant exit from teaching in the first decade, conditional on still teaching in 1930, the large majority remain teachers in 1940.

6.2 Labor Force Participation

Passing also increases extensive-margin labor force participation. Overall, marginal passers are 10.9 percentage points (18 percent) more likely to ever be observed in the labor force ($p = 0.003$; below-cutoff mean 59.1 percent; Table 4, column 1). This is not an obvious corollary of the teaching estimates. Women who fail could have entered other occupations, but instead it appears that the majority of women on the margin of teaching were also on the margin of labor force entry. Consistent with this, the labor force effect is not statistically distinguishable from the teaching effect (a difference of 3.8 percentage points, $p = 0.13$;

[14]Note that, although 1920 is the best census year to observe short-run effects, it has several limitations. First, there was a teacher shortage in 1919, leading to a large number of emergency certificates being issued in 1919 (40 percent of elementary credentials issued in 1919 were on an emergency basis, which is approximately double that of typical years). This may lead to an inflation in the number of nonpassers observed as teachers in this census. As such, I view 20 percentage points as a likely lower bound on the immediate short-run effects, but it is the best approximation given data constraints.

Appendix Table A1).

As with the teaching effect, I decompose this labor force effect over time. Figures 7 and 8, together with Table 4, show an increase of 18.9 percentage points among the 1918–1919 cohorts and 15.6 percentage points among all 1916–1919 cohorts in 1920 (both $p < 0.01$; columns 2 and 3). Over the longer run, the participation effect is 7.2 percentage points in 1930 ($p = 0.06$; below-cutoff mean 39.6 percent; column 4) and 8.6 percentage points in 1940 ($p = 0.02$; mean 30.9 percent; column 5).

The participation gains also translate to earnings. Figure 9 and the final column of Table 4 (column 6) show that passing raises a woman's own wage income in 1940 by \$91.5 ($p = 0.06$), approximately 35 percent of the below-cutoff mean of \$260; this is driven entirely by extensive margin LFP.

6.3 Labor Force Persistence

Marginal passers remain more likely to be in the labor force and to remain teachers both in the long run—7 to 14 years after the exam—and in the very long run—17 to 24 years after the exam. In this section, I examine factors driving this persistence, focusing on teaching and labor force participation in the 1930 and 1940 censuses. Because both censuses capture the same long-run horizon, I pool them where each outcome equals one if a woman is observed in the indicated state in either 1930 or 1940. For example, the outcome "married teacher" equals one if a woman is observed teaching with a spouse present in either census. Appendix Table A2 reports the estimates separately by census wave.

6.3.1 Teaching Drove Persistence

I first ask which occupations account for the long-run labor force response (Table 8). Pooling the 1930 and 1940 censuses, passing raises labor force participation by 7.1 percentage points (15 percent; panel A, column 1), and the main driver of this response is teaching (panel B, column 1). Passing has no detectable effect on nonteaching white-collar work or on other work (Panels C and D), which makes teaching account for essentially all of the long-run participation response, though this does differ somewhat by marital status, as discussed next. Figure B15 shows the shares of the top 10 occupations observed in the full sample in the 1930–1940 censuses. The leading occupation is teaching at 53.5 percent, followed distantly by stenography (8.1 percent) and bookkeeping (5.6 percent).

6.3.2 Married Women Persisted

I next ask whether the long-run effects are driven by married or never-married women. Because my main estimates in Table 8 are pooled outcomes across the 1930 and 1940 censuses, and because women's marital status can change between censuses, the married and unmarried cells overlap, and their estimates do not sum to the totals because of pooling and to different bandwidths. For census-wave-specific estimates, see Appendix Table A2.

Passing raises the probability of nonmarried LFP by 4 percentage points (12 percent), though imprecisely ($p = 0.19$), and married LFP by 5 percentage points (39 percent) (Table 8, panel A, columns 2 and 3). For unmarried women, this LFP effect is driven by teaching (+6 percentage points; panel B, column 2), which potentially crowded out white-collar work (−3 pp, $p = 0.20$; panel C, column 2). In contrast, married women's LFP effect is driven mostly by teaching (+3 percentage points; panel B, column 3) and potentially by other white-collar work (+1.5 pp; panel C, column 3).

These pooled effects mask some heterogeneity by census wave (Appendix Table A2), in which married women's LFP effects in 1930 are driven by both teaching and other white-collar work, whereas by 1940 married LFP was driven primarily by teaching. In 1930, single LFP is driven by teaching with potentially some crowd out of other white-collar work, and by 1940 driven mainly by teaching. These findings suggest that, although access to teaching may have some intermediate effects on other white-collar work, its main long-run impacts on LFP are driven by those who teach.

To better understand the composition of those in the labor force and as teachers in 1940, Figure 14 plots the entry and exit patterns of women relative to their 1930 occupation and labor force status. I plot this for the full sample, as the RD is not suited to identify granular transition behavior, and so this exercise should be interpreted as descriptive only. I see that 9 percent of teachers in 1940, and 18 percent of women in the labor force in 1940, are women who were married and not teaching (respectively, not in the labor force) in 1930 and who are still married in 1940, suggesting reasonably large rates of reentry into the profession—and even more so into the labor force—for married women. This suggests that, although for many teaching was a sustained and continuous career, acquiring a teacher certificate in early adulthood may have allowed women to enter, exit, and then reenter the labor force later on. This is consistent with the observation that 1940 and 1950 married LFP was driven by older married women, precisely those with past work experience (Goldin, 1989).

6.3.3 Education Reinforced Persistence

Next, I investigate how labor force persistence interacts with human capital accumulation. In Table 9 (columns 1–2), I show that marginal passers have slightly more years of schooling in 1940 (+0.3 years, $p = 0.07$) and are significantly more likely to have 14 or more years of schooling (+7 pp). Columns 1–2 report years of completed schooling in 1940 and an indicator for 14 or more years.

To understand who drives these effects, I split the 14+ indicator by whether women were observed teaching in 1930 or 1940. The effect comes from women who were teachers (+6.0 percentage points, $p = 0.06$; column 3), with no effect among women who were not (+0.8 percentage points, $p = 0.74$; column 4). Although these joint outcomes do not establish that education causally produces persistence, they highlight that education and labor force persistence are complementary.

7 Family Effects

I then turn to family outcomes where, unlike the teaching or LFP outcomes, never-married and fertility measures are absorbing in nature, which facilitates better measurement of these outcomes. This also means, for example, that one should not try to "scale" family outcomes by LFP or teaching outcomes, because the LFP and teaching measures are incomplete. Finally, I examine whether changes in income translate to changes in household resources.

7.1 Marriage

Access to teaching also changed women's marriage outcomes. Table 5 and Figure 11 show that passing raises the probability of never having married by 1930 by 6.8 percentage points ($p = 0.07$), against a below-cutoff mean of 34.4 percent (column 1). The effect attenuates to 5.2 percentage points by 1940 ($p = 0.06$; mean 19.3 percent; column 2), which displaces marriage with a spouse present (which falls by 5.7 percentage points, $p = 0.07$; column 3). Divorce and widowhood are unchanged (columns 4 and 5). This pattern is consistent with at least part of the response operating through delayed marriage.

7.2 Fertility

Next, I consider the number of observed children (Table 6 and Figure 12). To analyze fertility, I limit the sample to women observed in both the 1930 and 1940 censuses. Because

the census does not record children ever born,[15] I measure fertility as the number of distinct children ever observed in a woman's household across the two waves. Limiting the sample to women who appear in both censuses allows me to capture children who appear in 1930 but have left the household by 1940, and children who arrive in 1940 but were not present in 1930. Section 3 provides details. On average, women are 38 years old in 1940, so these measures are close to, but likely not exactly, completed fertility.

Overall, marginal passing lowers estimated fertility by 0.32 children ($p = 0.03$), about 16 percent of the below-cutoff mean of 2.02 (column 1). Though imprecise, there is evidence that this effect is driven by both the extensive and intensive margins of fertility. The probability of having no children rises by 4.4 percentage points ($p = 0.25$; column 2). At the same time, fertility conditional on being positive falls by 0.26 children ($p = 0.16$; column 3). This second effect is possibly driven in part by a delay in family formation; passing is also associated with possible delays in age at first birth by 0.56 years ($p = 0.21$; column 4).

7.3 Women's Contribution to Household Wage Income

Table 7 shows a negative but statistically insignificant effect on total household wage income (−\$41.5; $p = 1.00$; column 1). Passing raises a woman's own wage income by \$110.6 ($p = 0.08$; column 2),[16] concentrated among women without a coresident spouse (\$89.9; column 5) rather than among those with one (\$20.7; column 4), though neither component is individually precise.

Husband-contributed wage income contributes negatively to the household income (−\$155.3; $p = 0.38$; column 3), both due to women being less likely to be married at all (and therefore, having husbands' contributing 0 to household income) and to the fact that conditional on marriage, women marry husbands who earn if anything less (\$86 less; Appendix Table A9). Beyond earnings, I find no evidence that passing changed whom women married; husbands are no better matched on education, occupational standing, or income rank relative to the woman's father (Appendix Table A3). Appendix Table A5 shows no discontinuity in income-data availability.

8 Heterogeneity

Although my sample is not large enough to support extensive heterogeneity analysis, I present suggestive analyses describing how effects differed across locations with and without the

[15] The 1940 census asked children ever born only of the sample-line subsample.

[16] This estimate uses a common bandwidth across the household-income components and so differs from the \$91.5 in Table 4, which uses the outcome-specific MSE-optimal bandwidth.

marriage bar, and across women from farm versus nonfarm households. In each heterogeneity analysis, I estimate group-specific RD effects with jointly chosen, group-specific bandwidths and test the statistical difference across groups using Calonico et al. (2025). Although no individual subgroup difference is statistically significant at the 5 percent level, I emphasize the pattern of point estimates.

8.1 By Marriage-Bar County

Whether married women could keep working depended in part on whether they lived in places that barred married women from teaching. Many districts operated such marriage bars, either formally or informally (Section 2). Figure 13 splits the married-women employment effects by whether a woman's county of residence has a documented marriage bar. The married-teacher effect is concentrated in counties with no documented bar; in counties with a documented bar, it is close to zero. There is not strong evidence that women in marriage-bar counties shifted into other white-collar work (though the estimate is 2.1 percentage points, $p = 0.30$).

Appendix Table A6 extends this heterogeneity analysis to the full outcome set. The point estimates suggest the family response also differed by bar status. Extensive-margin effects are often larger (ever married by 1930 falls 10.5 percentage points; childlessness rises 7.0 points) in locations with marriage bars, whereas intensive-margin effects (first births are delayed by 1.0 years) are larger in locations without the marriage bar.

8.2 By Farm versus Nonfarm Household

Roughly half of fathers are farmers, who share a single occupation score. Therefore, instead of splitting my sample by household occupation score, the most appropriate heterogeneity analysis by childhood characteristics is whether a woman's father was a farmer, as measured in her 1910 childhood household (Section 3).[17] The split is reported in Appendix Table A7.

I find that passing increases teaching similarly in both groups (1940: +9.2 percentage points farm, +8.2 nonfarm; both $p < 0.10$). However, labor force effects, and other downstream outcomes, are consistently larger among farmers' daughters, who had the fewest local white-collar alternatives. Own wage income rises \$166 ($p < 0.05$) among farm daughters against \$21 ($p = 0.74$) among nonfarm daughters; long-run labor force participation, schooling (+0.44 years, $p = 0.10$, vs. +0.17), and the fertility decline (−0.47 children, $p = 0.02$, vs. −0.00) are all concentrated in the farm group. For nonfarm daughters, who more often

[17] Taking the father from the 1910 household, recovered via the woman's own childhood census record, raises coverage to about 93 percent of the sample; the split is close to even (49.6 percent farm).

had access to other white-collar work, passing shifted women into teaching without an accompanying rise in labor force participation. This is consistent with teaching playing a more central role in women's lives when outside options were scarce, and provides a new context to study women's work relative to existing work, which focuses on clerical work or relies disproportionately on white-collar shocks in urban settings.

9 Robustness

I show robustness of my results to several different empirical design decisions. The results of these exercises are reported across Appendices D and E.

Selected bandwidths: The main estimates use MSE-optimal bandwidths chosen outcome-by-outcome. I reestimate every outcome at common fixed windows around the cutoff (1.9, 3.8, and 5.8 percentage points of the average score, or 25, 50, and 75 raw points), so that differences across outcomes cannot reflect differences in the estimation sample. A related concern is that the few women sitting right at the threshold drive the estimates. To address this, I also implement a donut, dropping women within 0.4 percentage points (five raw points) of the threshold. Across the majority of these adjustments, I find that estimates keep their sign and are of a similar magnitude, though at times less precise and possibly sensitive.

Covariate adjustment: A second concern is that the design is sensitive to covariates. I add predetermined covariates (age at exam, years of education at exam, and birth year) to the main specifications; covariate adjustment should affect precision but not point estimates if the design is valid. I also reweight the linked sample to the baseline composition by the inverse of the census-match propensity (Appendix D). Finally, I predict each outcome from preexam characteristics alone and estimate the RD on the prediction (Appendix C). Estimates are similar with covariates and with reweighting, and no predicted outcome jumps at the threshold.

Estimation choices: A third concern is that the decisions I made about the RD specification drive the results, including decisions on weighting and inference. The main estimates use a triangular kernel; I reestimate with a uniform kernel, which increases the weight given to observations that lie further from the cutoff. I also report conventional, robust bias-corrected, and alternative clustering choices for the main outcomes, and honest confidence intervals for a discrete running variable (Kolesár and Rothe, 2018; Armstrong and Kolesár, 2020). The results are similar.

Placebo cutoffs: Finally, I show that the results I observe are unlikely to be driven by chance. I reestimate the main specifications at artificial cutoffs placed away from the true passing

threshold; the full placebo distributions for every outcome are reported in Appendix E.

10 The Role of Teaching in the Rise of Rural Women's Work

In this section, I provide back-of-the-envelope estimates to translate the observed RD estimates into overall changes in LFP and fertility in Illinois (excluding Cook County). To do this, I define the share of an aggregate change explained as the number of women who additionally entered a career, multiplied by the effect per entrant, divided by the aggregate change itself. Because the regression discontinuity estimates apply to women near the certification threshold, I multiply my RD estimates by the size of that marginal population and compare the result to the aggregate changes in fertility and LFP over relevant age cohorts. I compute this only among women in Illinois, excluding Cook. This is a conservative approach because I do not assume anything about how the effect, identified on marginal women, translates to different contexts. However, given that many women in the relevant birth cohorts (11 percent) are observed taking the teacher-certification exam, this exercise produces nontrivial aggregate estimates.[18]

In this exercise, I consider women marginal if their cumulative scores fell between 72 and 78 percent and if they passed, which made up 10.2 percent of test-takers. I count only the passers because my estimate is the effect of crossing the threshold. This then implies that roughly 1 percent of downstate women in the 1897–1902 birth cohorts were on the margin of access.

$$\underbrace{0.113}_{\text{sat the exam}} \times \underbrace{0.102}_{\text{passed within } \pm 3 \text{ of cutoff}} \approx 0.012,$$

I multiply this 1.2 percent by each reduced-form effect (Table 4, column 5 for 1940 LFP and Table 6, column 1 for fertility) and divide the result by the corresponding aggregate change among downstate-Illinois women to estimate the share of each aggregate change attributable to entry at the certification margin.

I calculate the aggregate change in LFP among women aged 35–44, comparing the 1940 census to the same age group in the 1900 census. To estimate changes in fertility, I instead compute the difference in completed fertility among women aged 45–54 in the 1950 census

[18] 19,805 unique women applicants born 1897–1902 in the exam registers, divided by 174,724 Illinois women in those birth cohorts in the 1920 full-count census, excluding Cook County.

relative to women aged 45–54 in the 1900 census.[19]

For women aged 35–44 between 1900 and 1940, labor force participation rose from 10.5 to 21.5 percent. For women aged 45–54, completed fertility fell from 4.83 children (1900 full count) to 2.67 (1950 sample-line children-ever-born, with never-married women assigned zero).

Under a very limited set of assumptions (only that the effect is constant among women scoring 72–78 percent), the certification margin accounts for a small but nontrivial share—roughly 1 percent—of the cohort rise in downstate-Illinois women's labor force participation, and 0.2 percent of the fertility decline.

$$\text{Female labor force (1940):} \quad \frac{0.012 \times 8.6 \text{ pp}}{21.5\% - 10.5\%} \approx 1\%,$$

$$\text{Completed fertility:} \quad \frac{0.012 \times 0.32 \text{ children}}{4.8 - 2.7 \text{ children}} \approx 0.2\%.$$

Because my estimates come from a predominantly rural population for whom teaching was the main white-collar opening, the most natural extension is to the growth of teaching in rural America. Among rural women aged 18–24, the share employed as teachers rose from 4.1 percent in 1910 to 5.2 percent in 1930, or by 1.1 percentage points. Because the relevant entrants here are teachers, I scale my estimates by the short-run first stage on teaching (20 percentage points), which implies an effect per entrant of 42.6 percentage points on 1940 participation and −1.58 children. However, not all of this growth represents new work. Across rural counties, each additional young woman employed as a teacher is associated with 0.67 to 0.90 additional young women in nonfarm work, with the remainder drawn largely from other white-collar jobs.[20] I therefore scale by this net-employment factor. Under these assumptions, the growth of teaching positions accounts for roughly 8 percent of the rise in rural women's prime-age labor force participation between 1900 and 1940, and just under 1 percent of the decline in rural completed fertility.[21]

[19] Fertility among women aged 45–54 is effectively completed fertility, measured as children ever born. The question was asked only of ever-married women, so never-married women are counted as zero. All samples are women in non-Cook County, Illinois.

[20] County-level regression of the change in young rural women's nonfarm employment share on the change in their teaching share, 1910–1930 and 1920–1930, from the full-count censuses. State fixed effects; weighted by the number of young women; standard errors clustered by state. The coefficients are 0.67 (0.16) and 0.90 (0.07). Nonfarm employment avoids the 1910 overcount of women on family farms.

[21] I measure the aggregate changes among rural women. Labor force participation of women aged 35–44 rose from 12.7 percent in 1900 to 17.4 percent in 1940, and completed fertility of women aged 45–54 fell from 5.15 children in 1900 to 3.24 in 1950. Completed fertility is children ever born. In 1950 the question was asked only of ever-married women on the sample line, so never-married women are counted as zero. Labor force participation in 1900 is proxied by reporting a gainful occupation.

$$\text{Rural prime-age labor force:} \quad \frac{0.011 \times (0.67 \text{ to } 0.90) \times 42.6 \text{ pp}}{17.4\% - 12.7\%} \approx 7\% \text{ to } 9\%,$$

$$\text{Rural completed fertility:} \quad \frac{0.011 \times (0.67 \text{ to } 0.90) \times 1.58 \text{ children}}{5.15 - 3.24 \text{ children}} \approx 0.6\% \text{ to } 0.8\%.$$

Appendix A.1 reports a much more liberal calculation that extends my estimates to other forms of white-collar work, where I find that if one is willing to assume that my per-entrant estimates apply to all new entrants into white-collar work, then this additional entry can account for up to 37 percent of the rise in prime-age LFP and 8 percent of the decline in fertility.

11 Conclusion

In this paper, I provide estimates of the long-run effects of women's access to professional work during a period of substantial change in women's employment, education, and family formation. I exploit individual-level variation in access to teaching generated by the passing threshold on Illinois teacher-certification examinations. Using newly digitized examination records linked to the full-count census, I show that women who marginally gained access to teaching were significantly more likely to become teachers and to enter the labor force. These effects persisted for more than two decades. Marginal passers remained 33 percent more likely to teach and 15 percent more likely to participate in the labor force in the long run.

Women's long-run persistence in the labor force was driven almost entirely by continued employment in teaching. A substantial share of the effect was also driven by women who continued working after marriage. Finally, marginal passers had more schooling by 1940, with these gains concentrated among women who remained teachers. Together, these findings suggest that access to teaching created sustained careers and encouraged further investment in education that reinforced LFP persistence.

Access to teaching also changed women's family formation. Marginal passers are more likely to forgo (and possibly delay) marriage, and to have fewer children. Though imprecise, my estimates are consistent with both extensive and intensive fertility responses driving the overall decline in fertility. Access to teaching did not, however, change the characteristics of partners whom women married, measured across a variety of indicators including income, education, or occupational standing. Instead, I observe that economic gains associated with teaching were concentrated among women themselves.

Extending these estimates to aggregate settings, I show that additional early-career entry can explain meaningful shares of the rise in LFP and nontrivial shares of the fertility decline. Assuming my estimates extend to the growth in teaching in other rural settings, additional early-career entry into teaching accounts for roughly 8 percent of the rise in rural women's prime-age labor force participation and just under 1 percent of the decline in completed fertility. Extending them instead to all white-collar work, additional early-career entry can explain up to 37 percent of the rise in prime-age labor force participation and 8 percent of the decline in fertility.

These findings show that career women were not a predetermined group. Women could be induced into sustained careers when career opportunities became available. At a time when few professions were open to women, access to teaching had lasting effects on whether and how long they worked, including after marriage; how much education they pursued; and their decisions about marriage and children. Together with the aggregate estimates, these results suggest that early-career access to white-collar work was a key force in the transformation of women's economic and family lives.

References

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez. 2021. "Automated Linking of Historical Data." *Journal of Economic Literature*, 59(3): 865–918.

Abramitzky, Ran, Leah Platt Boustan, Harriet M. Brookes Gray, Katherine Eriksson, Santiago Pérez, Hannah M. Postel, Myera Rashid, and Noah Simon. 2025. "Finding John Smith: Using Extra Information for Historical Record Linkage." National Bureau of Economic Research NBER Working Paper 33999.

Acemoglu, Daron, David H. Autor, and David Lyle. 2004. "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury." *Journal of Political Economy*, 112(3): 497–551.

Albanesi, Stefania, and Claudia Olivetti. 2016. "Gender Roles and Medical Progress." *Journal of Political Economy*, 124(3): 650–695.

Angrist, Joshua D., and Jonathan Guryan. 2008. "Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements." *Economics of Education Review*, 27(5): 483–503.

Armstrong, Timothy B., and Michal Kolesár. 2020. "Simple and Honest Confidence Intervals in Nonparametric Regression." *Quantitative Economics*, 11(1): 1–39.

Bacolod, Marigee P. 2007. "Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality." *Review of Economics and Statistics*, 89(4): 737–751.

Bailey, Martha J. 2006. "More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply." *Quarterly Journal of Economics*, 121(1): 289–320.

Bald, Anthony. 2025. "The Birth of an Occupation: Professional Nursing in the Era of Public Health." Working paper.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2008. "Staying in the Classroom and out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births." *Economic Journal*, 118(530): 1025–1054.

Briskin, Michael. 2026. "Women, War, and Weak Instruments: A Re-Reassessment of the Role of World War II on Women's Labor Supply." Working paper, Boston University.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley E. B. Wilbert. 2025. "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project." *Explorations in Economic History*, 98: 101717.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2017. "rdrobust: Software for Regression-Discontinuity Designs." *The Stata Journal*, 17(2): 372–404.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, Filippo Palomba, and Rocío Titiunik. 2025. "rdhte: Conditional Average Treatment Effects in Regression Discontinuity Designs." arXiv:2507.01128.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2018. "Manipulation Testing Based on Density Discontinuity." *The Stata Journal*, 18(1): 234–261.

Cook, Katherine M. 1921. "State Laws and Regulations Governing Teachers' Certificates." U.S. Department of the Interior, Bureau of Education Bulletin 1921, No. 22, Washington, DC.

Corcoran, Sean P., William N. Evans, and Robert M. Schwab. 2004. "Changing Labor-Market Opportunities for Women and the Quality of Teachers, 1957–2000." *American Economic Review*, 94(2): 230–235.

Deneault, Christa, Evan Riehl, and Jian Zou. 2026. "Disparate Impacts of Teacher Certification Exams." Federal Reserve Bank of Dallas Working Paper 2602.

Doepke, Matthias, Moshe Hazan, and Yishay D. Maoz. 2015. "The Baby Boom and World War II: A Macroeconomic Analysis." *Review of Economic Studies*, 82(3): 1031–1073.

Doxey, Alison, Ezra Karger, and Peter Nencka. 2026. "High Schools and the Uneven Rise in American Opportunity." National Bureau of Economic Research Working Paper 35068.

Feigenbaum, James J., and Daniel P. Gross. 2024. "Answering the Call of Automation: How the Labor Market Adjusted to Mechanizing Telephone Operation." *Quarterly Journal of Economics*, 139(3): 1879–1939.

Fellegi, Ivan P., and Alan B. Sunter. 1969. "A Theory for Record Linkage." *Journal of the American Statistical Association*, 64(328): 1183–1210.

Fernández, Raquel. 2013. "Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century." *American Economic Review*, 103(1): 472–500.

Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti. 2004. "Mothers and Sons: Preference Formation and Female Labor Force Dynamics." *Quarterly Journal of Economics*, 119(4): 1249–1299.

Geruso, Michael, and Heather Royer. 2018. "The Impact of Education on Family Formation: Quasi-Experimental Evidence from the UK." National Bureau of Economic Research Working Paper 24332.

Goldin, Claudia. 1988. "Marriage Bars: Discrimination against Married Women Workers, 1920's to 1950's." National Bureau of Economic Research NBER Working Paper 2747.

Goldin, Claudia. 1989. "Life-Cycle Labor-Force Participation of Married Women: Historical Evidence and Implications." *Journal of Labor Economics*, 7(1): 20–47.

Goldin, Claudia. 1990. *Understanding the Gender Gap: An Economic History of American Women.* New York, NY:Oxford University Press.

Goldin, Claudia. 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review*, 96(2): 1–21.

Goldin, Claudia, and Claudia Olivetti. 2013. "Shocking Labor Supply: A Reassessment of the Role of World War II on Women's Labor Supply." *American Economic Review*, 103(3): 257–262.

Goldin, Claudia, and Lawrence F. Katz. 2000. "Career and Marriage in the Age of the Pill." *American Economic Review*, 90(2): 461–465.

Goldin, Claudia, and Lawrence F. Katz. 2002. "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy*, 110(4): 730–770.

Greenwood, Jeremy, Ananth Seshadri, and Mehmet Yorukoglu. 2005. "Engines of Liberation." *Review of Economic Studies*, 72(1): 109–133.

Illinois Department of Public Instruction. 1919. "Educational Press Bulletin." *Educational Press Bulletin*, , (139). Springfield, IL. Digitized: HathiTrust, https://hdl.handle.net/2027/coo.31924066940648.

Illinois Department of Public Instruction. 1923. "Educational Press Bulletin." *Educational Press Bulletin*, , (178). Springfield, IL. Digitized: HathiTrust, https://hdl.handle.net/2027/coo.31924066940648.

Illinois State Archives. 1923. "Reports of the Illinois State Examining Board: Teachers' Examination Ledgers, 1916–1923." *Springfield, IL. Digitized by the author*, Record series number to be added.

Illinois State Civil Service Commission. 1922. "Classification of Offices and Places of Employment Under the Jurisdiction of the Illinois State Civil Service Commission." Job classes with monthly salary ranges, duties, and education requirements.

Illinois State Examining Board. 1940. "Illinois State Examining Board, Circular No. 316." Compiled by C. H. Engle; issued by John A. Wieland, Superintendent of Public Instruction. Digitized: HathiTrust, https://hdl.handle.net/2027/mdp.39015074837801.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics*, 142(2): 615–635.

Jepsen, Christopher, Peter Mueser, and Kenneth Troske. 2016. "Labor market returns to the GED using regression discontinuity analysis." *Journal of Political Economy*, 124(3): 621–649.

Kim, Amy, and Carolyn Tsao. 2025. "The Effects of Prohibiting Marriage Bars: The Case of U.S. Teachers." Working paper, Princeton University.

Kolesár, Michal, and Christoph Rothe. 2018. "Inference in Regression Discontinuity Designs with a Discrete Running Variable." *American Economic Review*, 108(8): 2277–2304.

Linacre, Robin, Sam Lindsay, Theodore Manassis, Zoe Slade, Tom Hepworth, Ross Kennedy, and Andrew Bond. 2022. "Splink: Free Software for Probabilistic Record Linkage at Scale." *International Journal of Population Data Science*, 7(3).

Li, Sophie. 2025. "The Effect of a Woman-Friendly Occupation on Employment: U.S. Postmasters Before World War II." Working paper.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics*, 142(2): 698–714.

McCrary, Justin, and Heather Royer. 2011. "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth." *American Economic Review*, 101(1): 158–195.

Ogren, Christine A. 2005. *The American State Normal School: An Instrument of Great Good.* New York:Palgrave Macmillan.

Orellana, Alexis, and Marcus A. Winters. 2025. "Licensure Tests and Teacher Supply." *Journal of Human Resources*. Forthcoming.

Pei, Zhuan, David S. Lee, David Card, and Andrea Weber. 2022. "Local Polynomial Order in Regression Discontinuity Designs." *Journal of Business & Economic Statistics*, 40(3): 1259–1267.

Rashid, Myera. 2025. "Keys to Upward Mobility: Typewriter Adoption and Women's Economic Outcomes." Working paper; revise and resubmit, American Economic Review.

Rose, Evan K. 2018. "The Rise and Fall of Female Labor Force Participation during World War II in the United States." *Journal of Economic History*, 78(3): 673–711.

Ruggles, Steven, Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren. 2024. "IPUMS Ancestry Full Count Data: Version 4.0 [dataset]."

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, Renae Rogers, Jonathan Schroeder, and Kari C. W. Williams. 2025. "IPUMS USA: Version 16.0 [dataset]."

Shenhav, Na'ama. 2021. "Lowering Standards to Wed? Spouse Quality, Marriage, and Labor Market Responses to the Gender Wage Gap." *Review of Economics and Statistics*, 103(2): 265–279.

State of Illinois. 1915. "The School Law of Illinois, Enacted by the Forty-Ninth General Assembly." Certification of teachers: act approved June 28, 1913, in force July 1, 1914; sec. 6 amended June 23, 1915, pp. 7–9.

Tsao, Carolyn. 2025. "It's Not (Just) About the Money: Pay and the Value of Working Conditions in Teaching." Working paper.

Vidart, Daniela. 2024. "Human Capital, Female Employment, and Electricity: Evidence from the Early 20th-Century United States." *Review of Economic Studies*, 91(1): 560–594.

Vidart, Daniela. 2026. "Revisiting the Link between Electrification and Fertility: Evidence from the Early 20th Century United States." *Explorations in Economic History*.

Yang, Wenni, and Rebecca Brough. 2026. "Teacher Marriage Bars." Working paper.

Figures

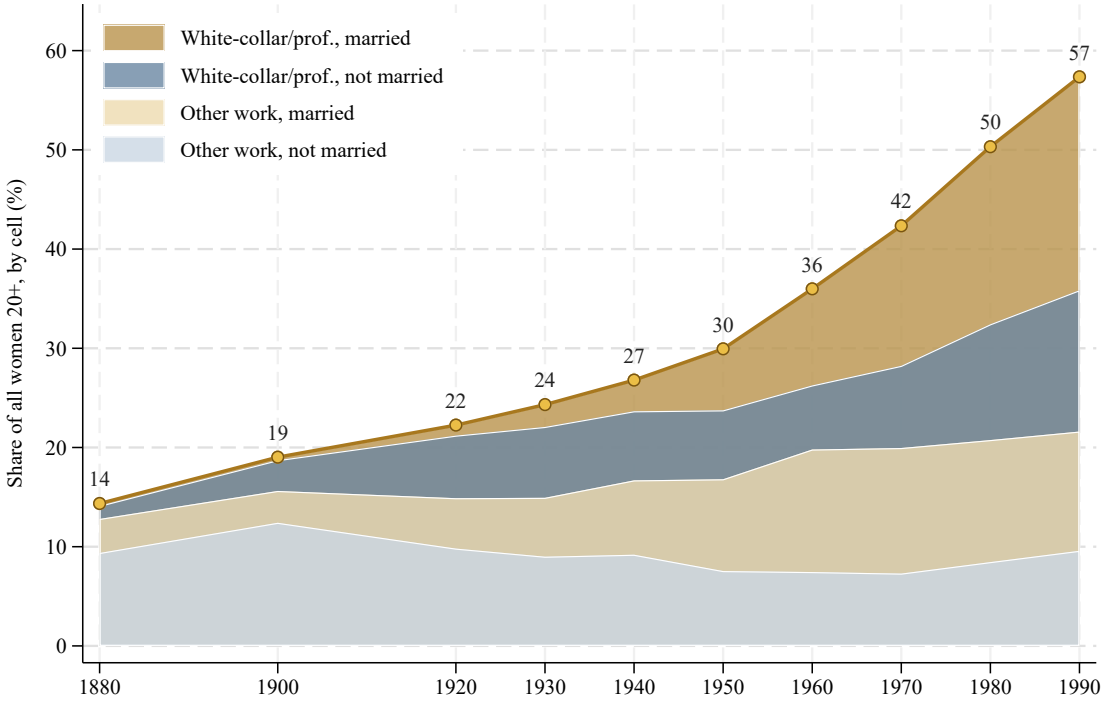


Figure 1: The Transformation of Women's Work, 1880–1990

Notes: The figure decomposes female labor force participation into four mutually exclusive groups defined by marital status (currently married versus not currently married) and occupation type (white-collar/professional versus other work). The gold line denotes total participation. White-collar/professional work is 1950 occupation codes 1–99 (professional and technical) and 200–399 (managerial and clerical). Data are the IPUMS 1 percent samples of the decennial censuses for 1880–1990 (the 1910 sample is excluded because its enumeration overstates women's gainful employment); estimates are weighted by person weights. Appendix Figure B16 plots the same series for professional and technical work alone. *Source:* IPUMS USA 1 percent samples of the decennial censuses (Ruggles et al., 2025).

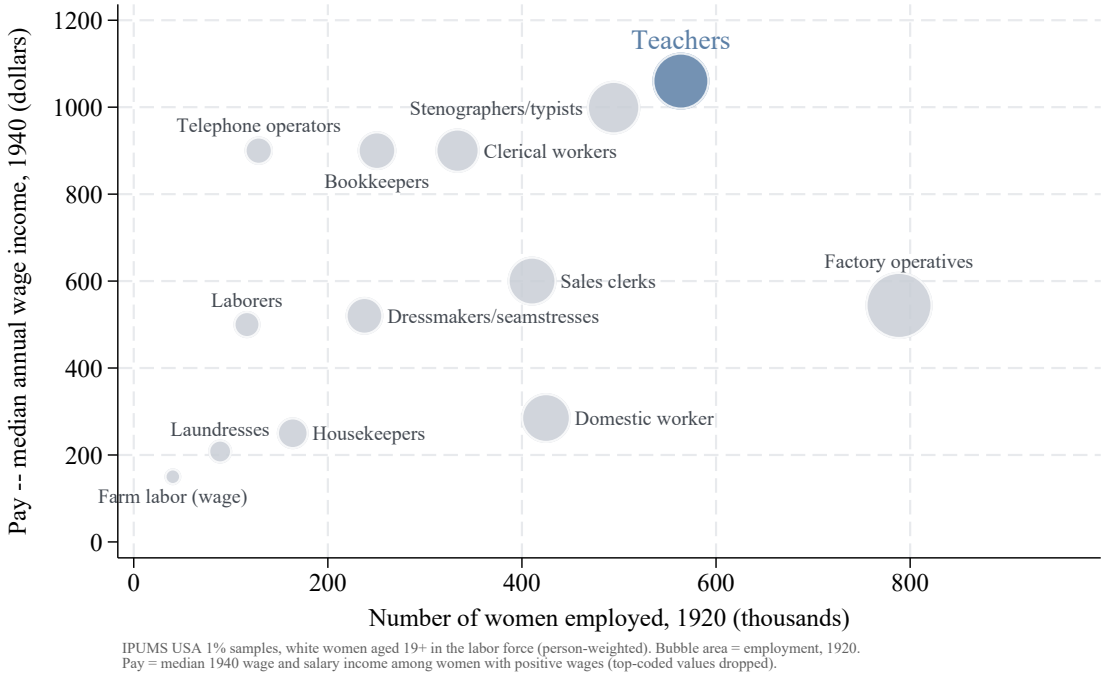


Figure 2: Teaching as a Large, High-Paying Occupation for Women

Notes: The figure plots major occupations for white women aged 19 and older in the 1920 census labor force. The x-axis gives the number of women employed in each occupation in 1920, and the y-axis gives the median annual wage income in 1940 among white women aged 19 and older in the labor force in that occupation with positive wage income (top-coded values dropped). Bubble size is proportional to the number of women employed in 1920. Farmers and unpaid family farm workers are omitted because their work was not paid as wages. *Source:* IPUMS USA 1 percent samples of the 1920 and 1940 censuses (Ruggles et al., 2025).

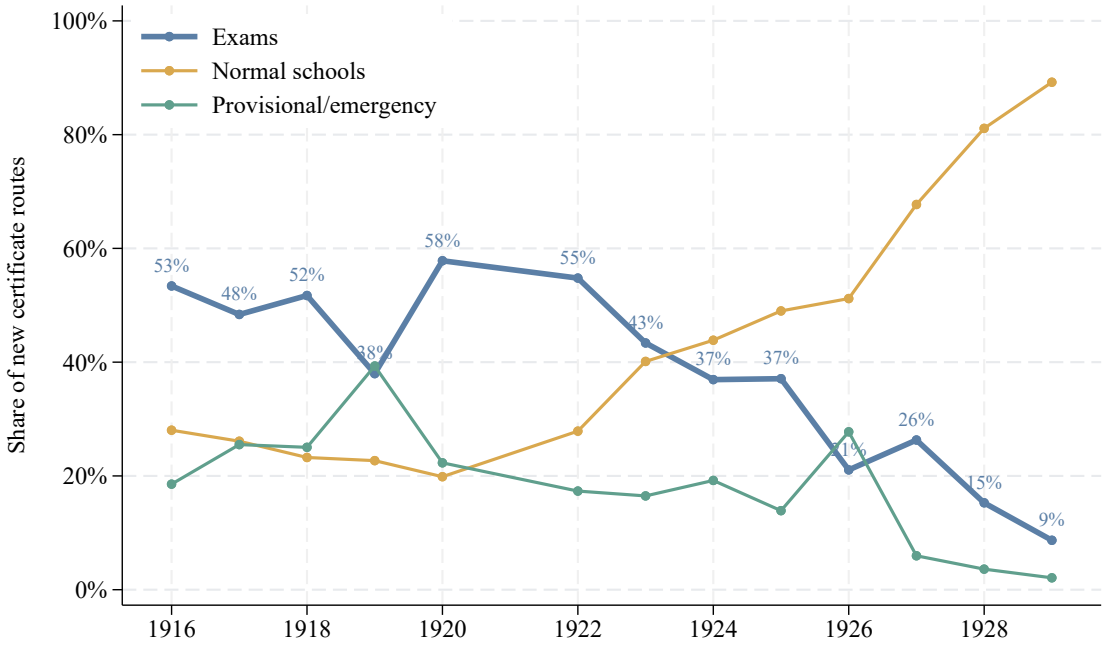
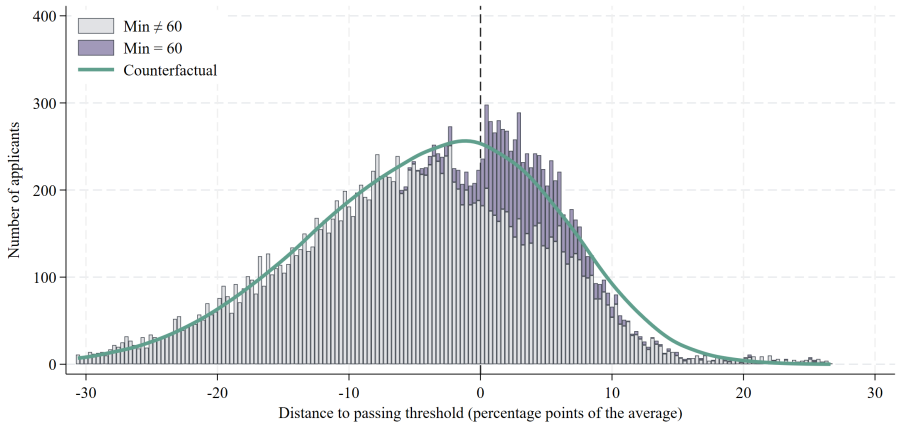
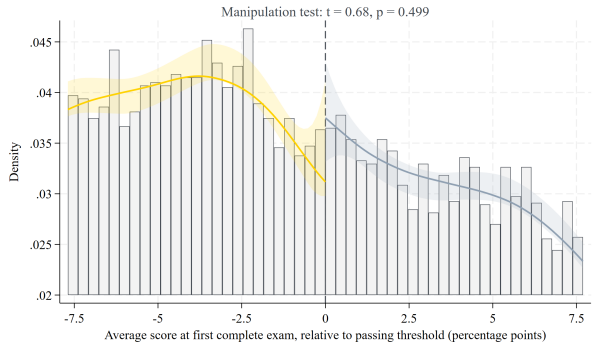


Figure 3: Routes to Elementary Teaching Certificates for Women outside Cook County

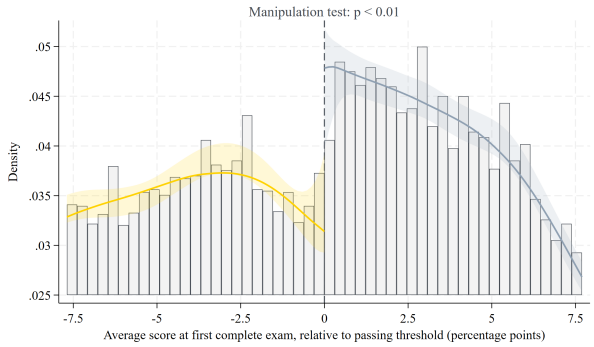
Notes: The figure plots new second-grade elementary certificate routes for women in Illinois, excluding Cook County. Normal schools are certificates issued based on credits filed. Provisional/emergency combines provisional and emergency certificates.



(a) Excess mass of exactly-60 records above the counterfactual



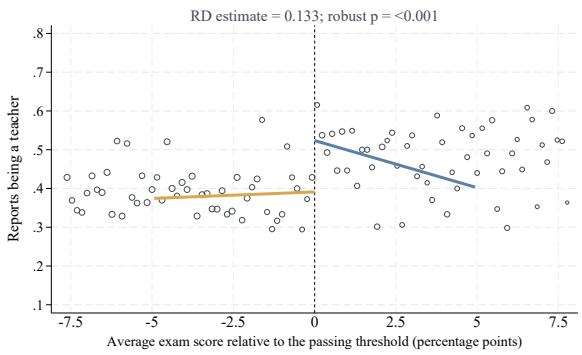
(b) Excluding bumped records (min score $\neq 60$)



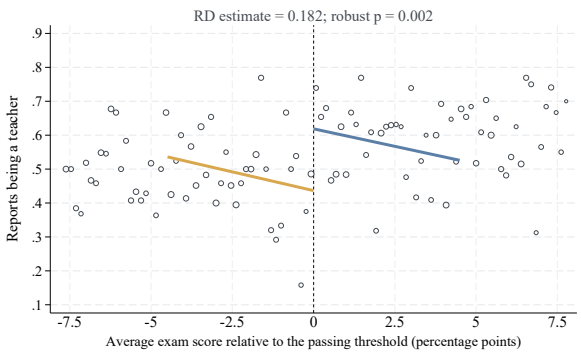
(c) Including bumped records (min score $= 60$)

Figure 4: Recovering Continuity in Passing Threshold

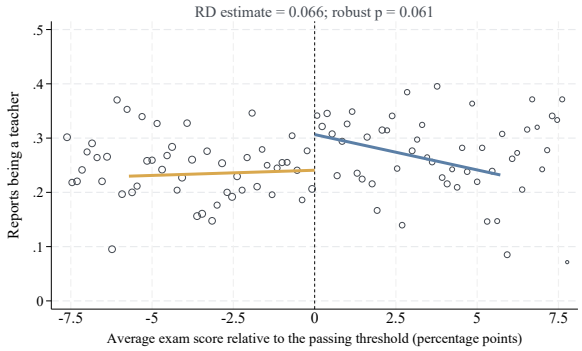
Notes: Exam scores of women aged 17–23, in percentage points of the average score, recentered so the passing threshold falls at 0. Panel (a) uses first-attempt scores. Records with a minimum subject score of exactly 60 (purple) are stacked as excess mass above records with a minimum score not equal to 60 (gray). The counterfactual (green) is the unmanipulated distribution from the grading simulation in Appendix Figure B9, scaled to the same number of applicants. Panels (b) and (c) report Cattaneo, Jansson and Ma (2018) manipulation tests for women aged 17–23 at their first complete 13-subject exam, with the score in percentage points of the average and the MSE-optimal bandwidth. Removing the exactly-60 records leaves the density continuous at the threshold ($t = 0.68$, $p = 0.50$), whereas including them produces a sharp discontinuity ($p < 0.01$).



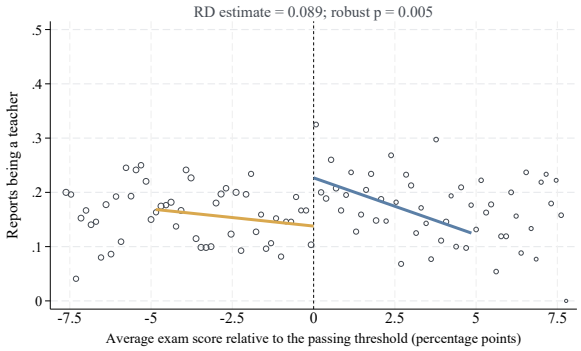
(a) Ever observed teaching by 1940



(b) Teaching in 1920



(c) Teaching in 1930



(d) Teaching in 1940

Figure 5: Effects on Teaching

Notes: Hollow circles are outcome means in bins of two raw points (0.15 percentage points of the average), with marker area proportional to the number of observations. Linear fits use the MSE-optimal bandwidth, and bins are shown within 100 raw points (7.7 percentage points) of the threshold. Each panel reports the RD estimate and its robust p-value. The 1920 panel uses the 1916–1919 exam cohorts; the remaining panels use all 1916–1923 cohorts.

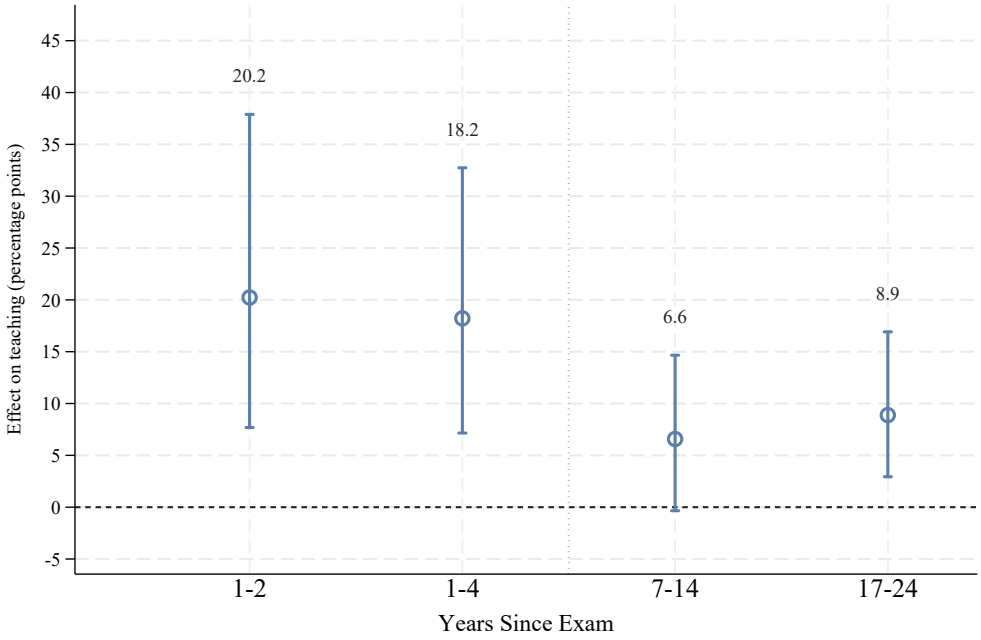
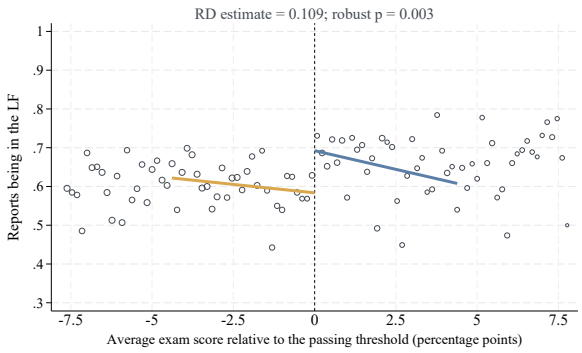
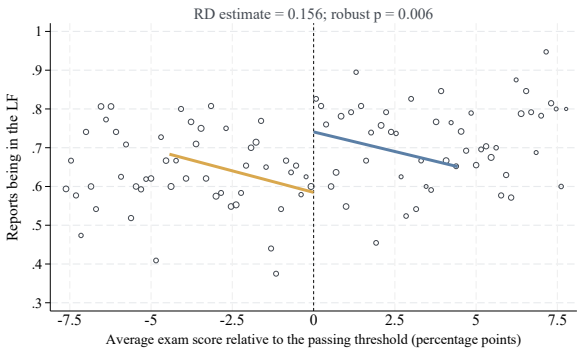


Figure 6: Immediate and Long-Run Effects on Teaching

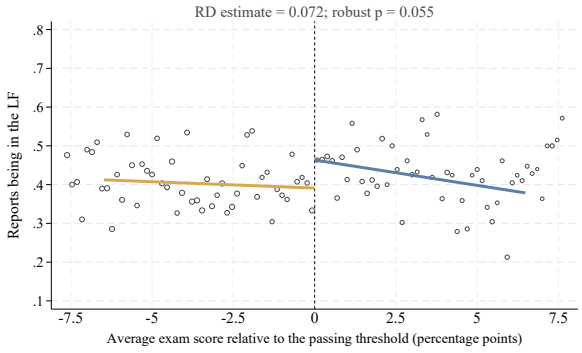
Notes: Markers show conventional point estimates, and vertical bars are 95 percent bias-corrected confidence intervals. Points 1 and 2 are measured in the 1920 census, point 3 in the 1930 census, and point 4 in the 1940 census. The one-to-two-year estimate uses the 1918–1919 cohorts, and the one-to-four-year estimate uses the 1916–1919 cohorts. The 1930 and 1940 estimates pool all 1916–1923 exam cohorts.



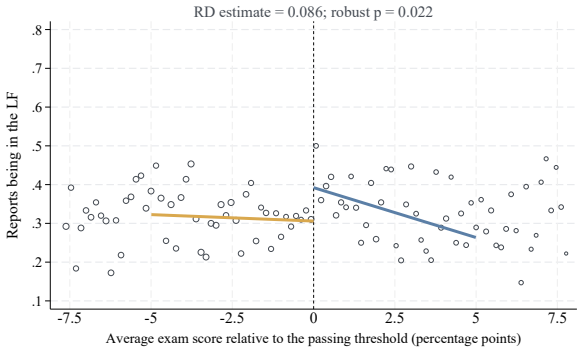
(a) Ever observed in the labor force by 1940



(b) Labor force participation in 1920



(c) Labor force participation in 1930



(d) Labor force participation in 1940

Figure 7: Effects on Labor Force Participation

Notes: Hollow circles are outcome means in bins of two raw points (0.15 percentage points of the average), with marker area proportional to the number of observations. Linear fits use the MSE-optimal bandwidth, and bins are shown within 100 raw points (7.7 percentage points) of the threshold. Each panel reports the RD estimate and its robust p-value. The 1920 panel uses the 1916–1919 exam cohorts; the remaining panels use all 1916–1923 cohorts.

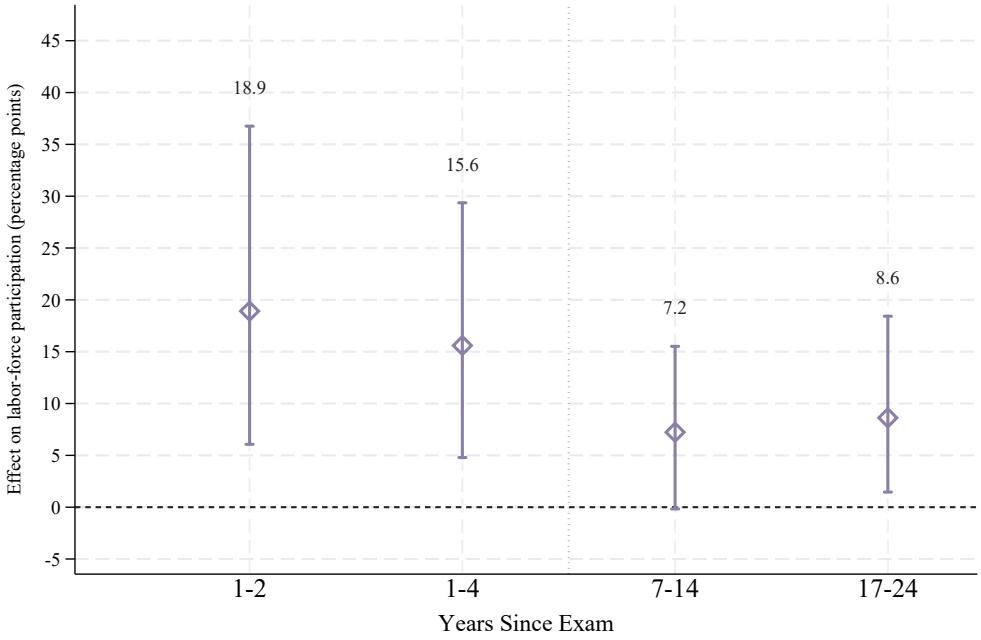


Figure 8: Immediate and Long-Run Effects on Labor Force Participation

Notes: Markers show conventional point estimates, and vertical bars are 95 percent bias-corrected confidence intervals. Points 1 and 2 are measured in the 1920 census, point 3 in the 1930 census, and point 4 in the 1940 census. The one-to-two-year estimate uses the 1918–1919 cohorts, and the one-to-four-year estimate uses the 1916–1919 cohorts. The 1930 and 1940 estimates pool all 1916–1923 exam cohorts.

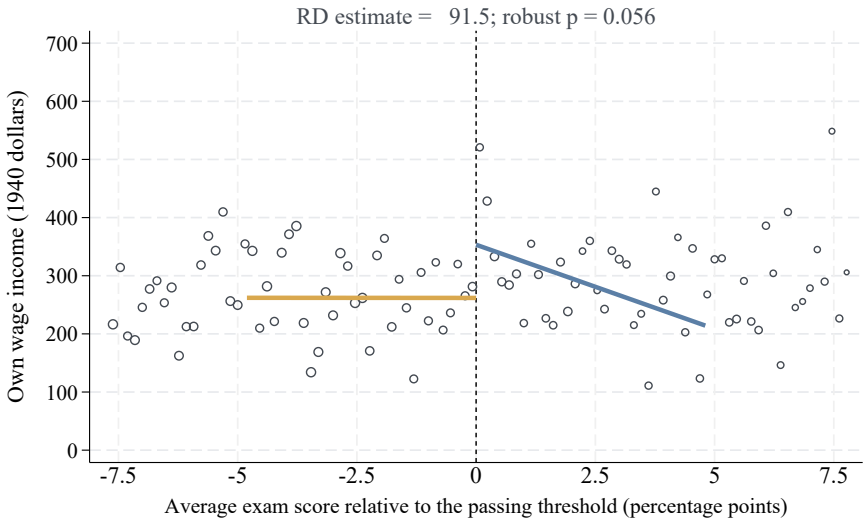
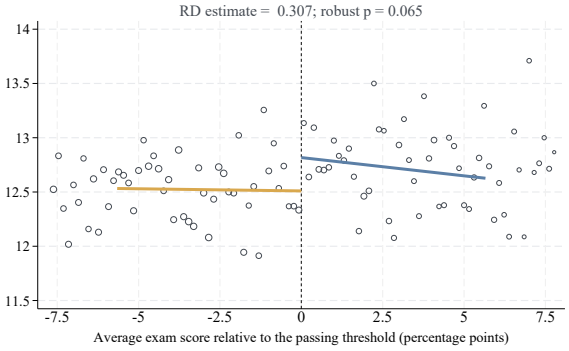
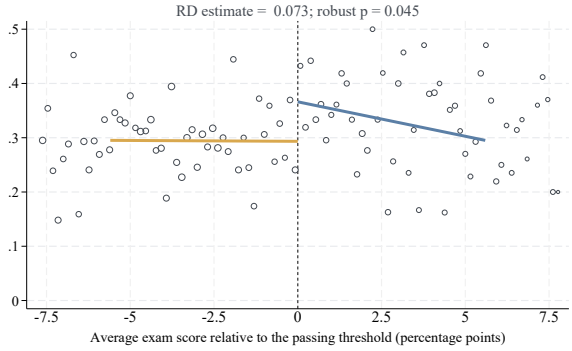


Figure 9: Effect on Own Wage Income in 1940

Notes: Hollow circles are outcome means in bins of two raw points (0.15 percentage points of the average), with marker area proportional to the number of observations. Linear fits use the MSE-optimal bandwidth, and bins are shown within 100 raw points (7.7 percentage points) of the threshold. The figure reports the RD estimate and its robust p-value.



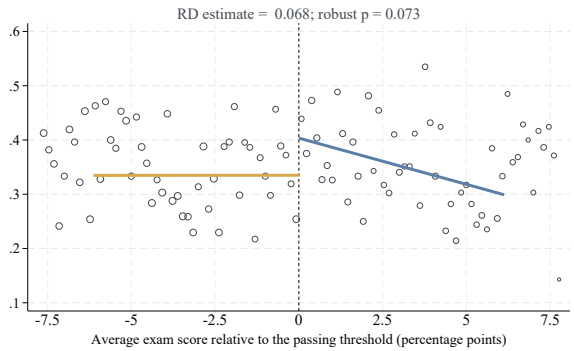
(a) Years of schooling in 1940



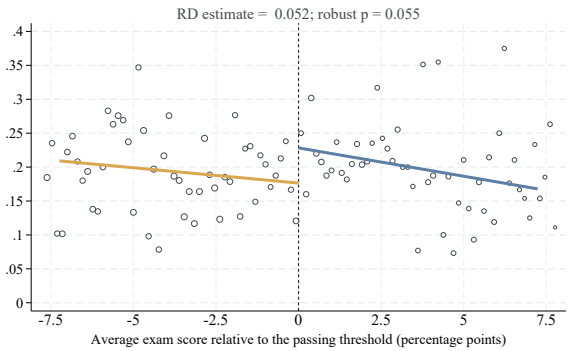
(b) Has 14+ years of schooling

Figure 10: Education in 1940

Notes: In panel (a) the outcome is completed years of schooling reported in 1940; in panel (b) the outcome is an indicator for reaching at least fourteen years of completed schooling. Hollow circles are outcome means in bins of two raw points (0.15 percentage points of the average), with marker area proportional to the number of observations. Linear fits use the MSE-optimal bandwidth, and bins are shown within 100 raw points (7.7 percentage points) of the threshold. Each panel reports the RD estimate and its robust p-value.



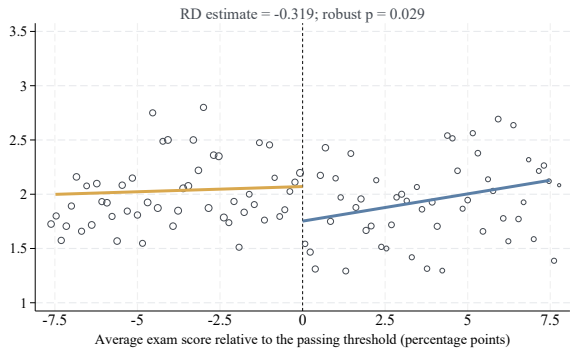
(a) Never married, 1930



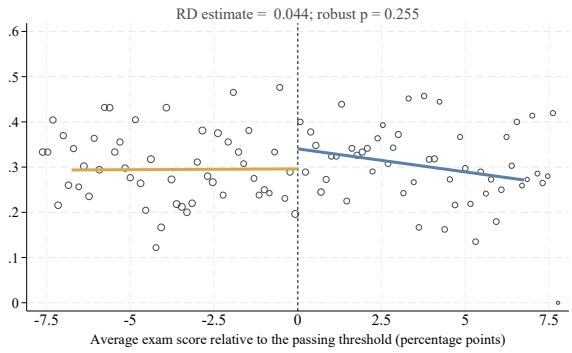
(b) Never married, 1940

Figure 11: Marriage

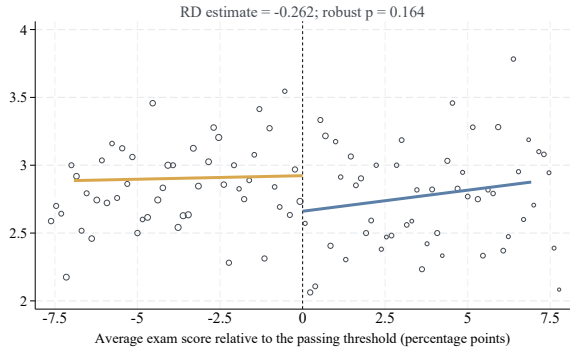
Notes: Hollow circles are outcome means in bins of two raw points (0.15 percentage points of the average), with marker area proportional to the number of observations. Linear fits use the MSE-optimal bandwidth, and bins are shown within 100 raw points (7.7 percentage points) of the threshold. Each panel reports the RD estimate and its robust p-value.



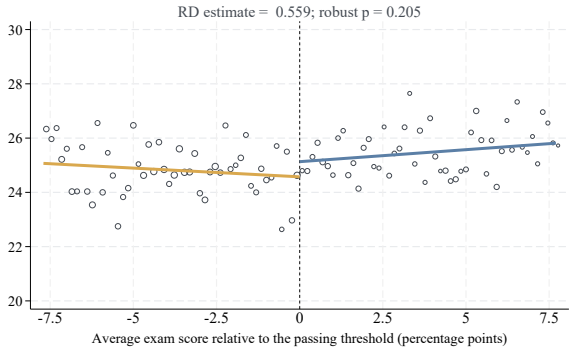
(a) Estimated fertility



(b) No children by 1940



(c) Fertility conditional on positive



(d) Age at first birth

Figure 12: Fertility

Notes: Every panel is restricted to women linked to both the 1930 and 1940 censuses. Hollow circles are outcome means in bins of two raw points (0.15 percentage points of the average), with marker area proportional to the number of observations. Linear fits use the MSE-optimal bandwidth, and bins are shown within 100 raw points (7.7 percentage points) of the threshold. Each panel reports the RD estimate and its robust p-value.

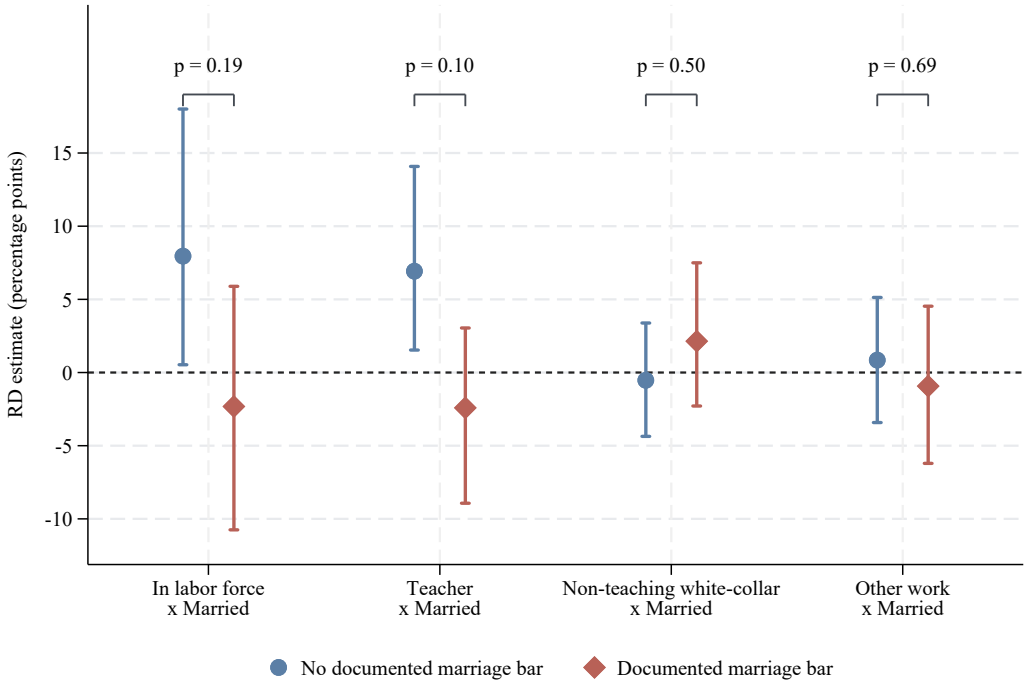


Figure 13: Employment Effects by Current-Residence Marriage-Bar Status

Notes: Markers show conventional point estimates, and vertical bars are 95 percent bias-corrected confidence intervals. Each outcome equals one if a woman is observed working in the indicated occupation while married with her spouse present, in either 1930 or 1940. Nonteaching white-collar work includes occupation codes 1–99, excluding teaching, and 200–399; other work is nonteaching, non–white-collar employment. The bracket above each pair shows the p-value for the difference between the two estimates (bar minus no-bar), following Calonico et al. (2025).

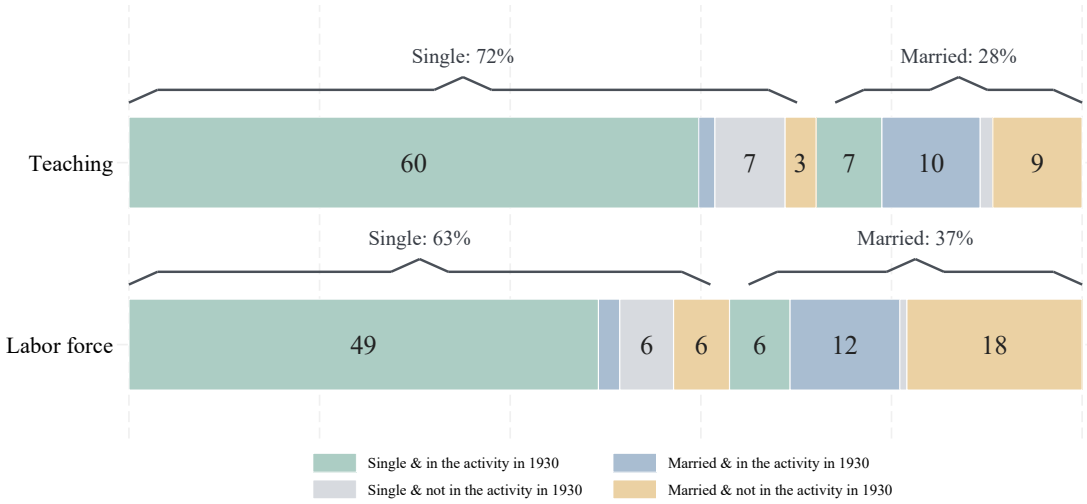


Figure 14: Composition of 1940 Teachers and Labor Force Participants, by 1930 Status

Notes: Each bar describes the composition of women observed in the indicated activity (teaching; any labor force participation) in the 1940 census, among applicant women linked to both the 1930 and 1940 censuses. Braces split each bar by 1940 marital status. Within each block, segments report the woman's 1930 status (single or married), crossed with whether she was observed in the activity in 1930. Values are percentages of all 1940 participants in the activity; segments below 3 percent are unlabeled. Married includes spouse present and spouse absent.

Tables

Table 1: Female Teacher Applicants Compared to Other Groups in Illinois

	Mean					Difference from teacher applicants			
Variable	Teacher applicants (1)	Illinois general (2)	Illinois excluding Cook (3)	Eligible women (4)	Normal School (5)	− Illinois general (6)	− Illinois excluding Cook (7)	− Eligible women (8)	− Normal School (9)
Age	13.22	13.21	13.21	13.34	13.23	–	–	–	–
Nonwhite (%)	0.4	1.6	1.8	1.6	1.3	-1.15	-1.41	-1.16	-0.85
						(0.05)	(0.05)	(0.12)	(0.25)
Urban (%)	19.9	59.0	33.2	41.1	29.5	-39.12	-13.29	-21.29	-9.57
						(0.30)	(0.30)	(0.37)	(1.10)
Foreign-born (%)	0.5	6.4	2.3	1.1	0.3	-5.92	-1.75	-0.62	0.23
						(0.06)	(0.06)	(0.09)	(0.13)
Number of siblings	3.19	3.37	3.46	3.14	3.03	-0.18	-0.26	0.11	0.16
						(0.02)	(0.02)	(0.02)	(0.06)
Mother present (%)	92.9	86.1	85.2	92.6	92.2	6.84	7.71	1.03	0.67
						(0.20)	(0.20)	(0.22)	(0.61)
Father present (%)	89.4	81.0	81.4	87.7	89.1	8.40	8.03	2.54	0.33
						(0.23)	(0.24)	(0.27)	(0.72)
Head in labor force (%)	93.0	92.2	92.8	90.9	90.8	0.77	0.19	2.18	2.23
						(0.19)	(0.19)	(0.23)	(0.68)
Head professional/technical (%)	2.7	2.6	2.2	4.5	6.2	0.13	0.53	-1.73	-3.44
						(0.12)	(0.12)	(0.15)	(0.61)
Head occupation score	18.30	20.53	18.30	20.51	19.89	-2.23	0.00	-2.16	-1.59
						(0.09)	(0.09)	(0.11)	(0.36)
Head education score	9.67	10.17	8.96	13.27	14.76	-0.52	0.70	-3.63	-5.10
						(0.12)	(0.12)	(0.15)	(0.58)
Head is farmer (%)	49.3	23.9	39.0	28.5	40.0	25.43	10.24	20.63	9.29
						(0.37)	(0.37)	(0.43)	(1.20)
In school (%)	90.0	74.5	78.3	91.4	89.3	15.53	11.71	3.25	0.78
						(0.22)	(0.21)	(0.21)	(0.58)
Observations	18,691	1,445,186	829,431	71,826	2,390				

Notes: Childhood-household characteristics measured in the 1910 census (1920 as a fallback where no 1910 link exists), for Illinois teacher-exam applicants, Normal School enrollees, and general Illinois women age ≤25. The column titled "Illinois excluding Cook" excludes Cook County. The column titled "Eligible women" includes women whose education level (at least 10 years of schooling by 1940) made them eligible to sit the exam, who reside outside of Cook County, and who are aged 8–18 in 1910; means are weighted by the inverse probability of matching to the 1940 census to account for nonrepresentativeness in matching between 1910 and 1940. Comparison columns are reweighted to the applicants' baseline-age distribution. Number of siblings is coresident children minus one; head variables refer to the household head, with "professional/technical" occupation codes 1–99 and education score EDSCOR50. Difference columns are coefficients on a teacher-applicant indicator from a weighted regression with age fixed effects; robust standard errors in parentheses. No difference is reported for age, the reweighting variable.

Table 2: RD Balance: Predetermined Covariates

Variable	Left Mean (1)	Jump (2)	p-value (3)
Age at exam	18.230	+0.021	0.690
		(0.076)	
Years of education at exam	11.312	+0.046	0.620
		(0.076)	
Urban (pre-exam residence)	0.216	−0.033	0.191
		(0.027)	
Mother present in household	0.936	+0.027	0.068
		(0.018)	
Father present in household	0.915	+0.019	0.284
		(0.022)	
Mother in labor force	0.076	+0.011	0.485
		(0.020)	
Father in labor force	0.968	−0.009	0.368
		(0.012)	
Any working woman in HH	0.215	+0.028	0.277
		(0.033)	
Number of siblings	3.042	−0.047	0.705
		(0.151)	
Household-head occ. score	18.756	+0.122	0.835
		(1.135)	
Household head is a farmer	0.477	−0.005	0.818
		(0.044)	
Birth year	1902.0	−0.057	0.717
		(0.201)	
Exam year	1919.9	+0.019	0.950
		(0.155)	
Matched to 1920 census (exam $\leq$ 1919)	0.867	−0.014	0.481
		(0.033)	
Matched to 1930 census	0.668	+0.004	0.919
		(0.027)	
Matched to 1940 census	0.653	−0.038	0.194
		(0.034)	
Matched to any subsequent census	0.811	−0.006	0.709
		(0.027)	

Notes: Each row is a separate local-linear RD regression of the listed predetermined covariate on the recentered running variable, using the MSE-optimal bandwidth. Robust bias-corrected standard errors are reported in parentheses. The last four rows test whether marginally passing the cumulative average threshold on the first attempt affected the probability of being linked to the 1920 census (for exams taken in 1916–1919), the 1930 census, the 1940 census, or any subsequent census. The 1920 row is restricted to exam cohorts through 1919; the pooled measure captures linkage to the 1920, 1930, or 1940 census for exam cohorts through 1919 and linkage to the 1930 or 1940 census for exam cohorts after 1919.

Table 3: Effects on Teaching

	Ever a Teacher (1)	1920 Census 1918–1919 (2)	1920 Census 1916–1919 (3)	1930 Census (4)	1940 Census (5)
RD Estimate	+0.133	+0.202	+0.182	+0.066	+0.089
	(0.041)	(0.077)	(0.065)	(0.038)	(0.036)
p-value	< 0.001	0.003	0.002	0.061	0.005
Left Mean	0.392	0.497	0.471	0.245	0.141
Effective N	3,975	1,021	1,532	3,738	3,114
N	11,185	2,564	4,698	9,221	8,816

Notes: Each column reports the RD estimate of the effect of passing the certification exam on the probability of being observed as a teacher, by local linear regression around the recentered cumulative score in the main specification with a triangular kernel and the MSE-optimal bandwidth (Calonico et al., 2017); robust bias-corrected standard errors in parentheses and bias-corrected p-values. "Ever a Teacher" equals one if a woman is observed teaching in any available census through 1940; the 1920 measures are available only for the early cohorts. Column 2 restricts to the 1918–1919 exam cohorts, observed one to two years after the exam; column 3 uses all 1916–1919 cohorts observed in 1920. "Left Mean" is the outcome mean just below the cutoff; "Effective N" is the number of observations within the bandwidth.

Table 4: Effects on Labor Force Participation

	Ever in the Labor Force (1)	1920 Census 1918–1919 (2)	1920 Census 1916–1919 (3)	1930 Census (4)	1940 Census (5)	Own Income 1940 ($) (6)
RD Estimate	+0.109	+0.189	+0.156	+0.072	+0.086	+91.5
	(0.041)	(0.078)	(0.063)	(0.040)	(0.043)	(55.8)
p-value	0.003	0.006	0.006	0.055	0.022	0.056
Left Mean	0.591	0.590	0.615	0.396	0.309	259.8
Effective N	3,567	957	1,503	4,228	3,158	3,045
N	11,185	2,564	4,698	9,221	8,816	8,718

Notes: See notes to Table 3. "Ever in the Labor Force" equals one if a woman is observed in the labor force in any available census through 1940; the 1920 measures are available only for the early cohorts. Column 2 restricts to the 1918–1919 exam cohorts, observed one to two years after the exam; column 3 uses all 1916–1919 cohorts observed in 1920. The next two outcomes are indicators for being in the labor force in the relevant census. The final column reports the woman's own wage income in 1940.

Table 5: Marriage

	Never Married 1930 (1)	Never Married 1940 (2)	Married, Spouse Present, 1940 (3)	Divorced 1940 (4)	Widowed 1940 (5)
RD Estimate	+0.068	+0.052	−0.057	−0.007	+0.015
	(0.040)	(0.031)	(0.036)	(0.011)	(0.012)
p-value	0.073	0.055	0.072	0.538	0.120
Left Mean	0.344	0.193	0.749	0.025	0.020
Effective N	3,986	4,541	3,923	4,442	3,502
N	9,221	8,816	8,816	8,816	8,816

Notes: See notes to Table 3. Outcomes are indicators for the listed marital status in the relevant census. Married, spouse present corresponds to marital-status code 1.

Table 6: Effects on Fertility

	Number of Observed Children (1)	No Children by 1940 (2)	Fertility Conditional on Positive (3)	Age at First Birth (4)
RD Estimate	−0.319	+0.044	−0.262	+0.559
	(0.163)	(0.040)	(0.192)	(0.424)
p-value	0.029	0.255	0.164	0.205
Left Mean	2.024	0.303	2.903	24.679
Effective N	4,079	3,706	2,663	2,813
N	7,784	7,784	5,460	5,237

Notes: See notes to Table 3. Every column is restricted to women linked to both the 1930 and 1940 censuses. Number of observed children is the number of own children observed across the linked census panel; no children equals one when the number of observed children is zero. The conditional-fertility column is restricted to women with at least one observed child. Age at first birth is the earliest inferred first-birth age across the two census waves.

Table 7: Women's Earnings and Household Wage Income

	Household Wage Income Accounting			Women's Wage Income by Spouse Status	
	Household Total (1)	Woman's Own Wage Income (2)	Husband-Contributed Wage Income (3)	Spouse Present (4)	No Spouse Present (5)
RD Estimate	−41.5	+110.6	−155.3	+20.7	+89.9
	(138.1)	(72.2)	(135.6)	(47.1)	(61.9)
p-value	0.999	0.078	0.377	0.417	0.150
Mean below cutoff	1,109.5	312.3	791.0	63.0	249.3
Effective N	3,316	3,316	3,316	3,316	3,316
N	7,059	7,059	7,059	7,059	7,059

Notes: See notes to Table 3. Wage income is measured in the 1940 census. Household wage income equals the woman's own wage income plus her coresident husband's wage income. Women's own-wage income is defined for all women regardless of marital status. Appendix Table A9 decomposes husband-contributed income into spouse-presence and conditional-income margins. Each outcome presented uses the optimal bandwidth selected for the household-income outcome.

Table 8: Persistent Employment and Occupation, Overall and by Marital Status

	All 1930–1940 (1)	Ever No Spouse 1930–1940 (2)	Ever Spouse Present 1930–1940 (3)
Panel A: Labor Force Participation			
RD Estimate	+0.071	+0.044	+0.050
	(0.042)	(0.038)	(0.032)
p-value	0.052	0.191	0.067
Left Mean	0.471	0.363	0.127
Effective N	3,917	4,619	3,536
N	10,234	10,234	10,234
Panel B: Teaching			
RD Estimate	+0.085	+0.062	+0.034
	(0.038)	(0.035)	(0.021)
p-value	0.018	0.068	0.060
Left Mean	0.257	0.217	0.045
Effective N	4,075	4,192	3,863
N	10,234	10,234	10,234
Panel C: Other Nonteaching White-Collar Work			
RD Estimate	−0.011	−0.027	+0.015
	(0.027)	(0.024)	(0.017)
p-value	0.596	0.205	0.307
Left Mean	0.138	0.108	0.036
Effective N	4,464	4,134	4,891
N	10,234	10,234	10,234
Panel D: Other Nonteaching, Non–White-Collar Work			
RD Estimate	+0.020	+0.012	+0.000
	(0.023)	(0.017)	(0.016)
p-value	0.313	0.427	0.985
Left Mean	0.093	0.048	0.050
Effective N	4,670	4,192	5,503
N	10,234	10,234	10,234

Notes: See notes to Table 3. Column 1 equals one if a woman is observed in the listed occupation in either the 1930 or 1940 census, regardless of spouse status. Columns 2 and 3 equal one if she is observed in the listed occupation and spouse-status cell in either census. Columns are not mutually exclusive because women's marital status may change between censuses. "Spouse Present" requires a coresident spouse; "No Spouse" includes never-married, divorced, widowed, and married women whose spouse is not recorded in the household. Other nonteaching white-collar work includes occupation codes 1–99 (excluding teachers, code 93) and 200–399. Other nonteaching, nonwhite-collar work is defined as occupations not already listed in the occ codes above. Appendix Table A2 reports estimates by census wave.

Table 9: Schooling in 1940 and Long-Run Employment Persistence

	Years of Schooling, 1940 (1)	14+ Years Schooling (2)	14+ Years × Teacher 1930–1940 (3)	14+ Years × Not a Teacher 1930–1940 (4)
RD Estimate	+0.307	+0.073	+0.060	+0.008
	(0.185)	(0.042)	(0.036)	(0.031)
p-value	0.065	0.045	0.064	0.742
Left Mean	12.520	0.294	0.151	0.144
Effective N	3,370	3,322	3,148	3,852
N	8,228	8,228	8,228	8,228

Notes: See notes to Table 3. Column 1 reports completed years of schooling in 1940. "14+ Years Schooling" equals one for women with at least 14 completed years of schooling. Columns 3–4 interact having 14+ years of schooling with whether a woman is observed as a teacher in the 1930 and/or 1940 census.

A Appendix Tables

Table A1: Equality of the Teaching and Labor-Force Effects

Outcome	RD estimate (pp)	p-value
Ever observed teaching by 1940	+13.3	<0.001
Ever observed in the labor force by 1940	+10.9	0.003
Difference (teaching − labor force)	+3.8	0.128
95% confidence interval	[−1.1, 9.1]	

Notes: Conventional local-linear RD estimates (MSE-optimal bandwidth, triangular kernel, heteroskedasticity-robust HC3 standard errors). "Ever observed teaching (in the labor force) by 1940" equals one if the woman is observed as a teacher (in the labor force) in any subsequent census. The difference row is the RD estimate on the difference outcome (ever teaching minus ever in the labor force). The reported p-value is the robust p-value testing the null that the effects (ever in the labor force and ever teacher) are equal.

Table A2: Employment and Occupation, by Marital Status and Census Wave

	No Spouse 1920 (1)	Spouse Present 1920 (2)	No Spouse 1930 (3)	Spouse Present 1930 (4)	No Spouse 1940 (5)	Spouse Present 1940 (6)
Panel A: Labor Force Participation × Marital Status						
RD Estimate	+0.156	−0.001	+0.045	+0.027	+0.049	+0.051
	(0.063)	(0.010)	(0.039)	(0.020)	(0.032)	(0.032)
p-value	0.007	0.822	0.193	0.209	0.074	0.055
Left Mean	0.608	0.006	0.328	0.068	0.216	0.093
Effective N	1,532	2,644	3,986	4,730	4,356	2,882
N	4,698	4,698	9,221	9,221	8,816	8,816
Panel B: Teaching × Marital Status						
RD Estimate	+0.179	+0.001	+0.051	+0.012	+0.057	+0.029
	(0.065)	(0.008)	(0.037)	(0.013)	(0.030)	(0.019)
p-value	0.003	0.917	0.130	0.362	0.033	0.080
Left Mean	0.468	0.003	0.213	0.032	0.111	0.030
Effective N	1,532	3,002	3,738	5,256	3,403	3,645
N	4,698	4,698	9,221	9,221	8,816	8,816
Panel C: Other Nonteaching White-Collar Work × Marital Status						
RD Estimate	−0.009	+0.000	−0.026	+0.012	−0.011	+0.002
	(0.033)	(0.000)	(0.021)	(0.011)	(0.021)	(0.015)
p-value	0.796	0.597	0.144	0.285	0.490	0.774
Left Mean	0.080	0.000	0.094	0.016	0.070	0.028
Effective N	2,157	544	4,499	4,963	3,596	4,484
N	4,698	4,698	9,221	9,221	8,816	8,816
Panel D: Other Nonteaching, Non–White-Collar Work × Marital Status						
RD Estimate	−0.011	−0.001	+0.013	+0.000	+0.000	+0.012
	(0.025)	(0.004)	(0.013)	(0.011)	(0.014)	(0.018)
p-value	0.695	0.939	0.252	0.823	0.861	0.323
Left Mean	0.064	0.003	0.023	0.019	0.035	0.038
Effective N	2,027	2,054	4,584	3,639	4,309	3,300
N	4,698	4,698	9,221	9,221	8,816	8,816

Notes: See notes to Table 8. Columns report separate outcomes for the 1920, 1930, and 1940 censuses. The 1920 columns are restricted to the 1916–1919 exam cohorts.

Table A3: Husband Characteristics and Intergenerational Income Rank

Outcome	RD Estimate	p-value	Mean Below Cutoff	Effective N	N
Husband's years of education	+0.067	0.923	10.145	2,844	6,294
	(0.355)				
Husband has 14+ years of education	+0.006	0.889	0.161	3,279	6,294
	(0.035)				
Husband's occupation score	+1.489	0.132	26.846	2,374	6,431
	(1.220)				
Husband is a teacher	+0.014	0.352	0.027	3,260	6,554
	(0.016)				
Age gap: husband minus wife	+0.179	0.598	2.632	2,832	6,554
	(0.461)				
Father's income-score rank	−0.730	0.815	45.643	5,485	10,129
	(2.003)				
Husband's wage-income rank	−2.386	0.317	49.293	3,022	5,864
	(3.032)				
Husband rank minus father rank	−0.568	0.746	5.968	2,099	4,387
	(4.290)				
Husband rank exceeds father rank	+0.024	0.585	0.500	2,300	4,387
	(0.055)				

Notes: Each row reports a local-linear RD estimate using an outcome-specific MSE-optimal bandwidth. Robust bias-corrected standard errors in parentheses. Husband characteristics are measured in the 1940 census and are conditional on a linked coresident husband and observation of the relevant characteristic. Father's income-score rank uses his positive IPUMS occupation score in the woman's preexam census and ranks him against white fathers in the full-count 1910 or 1920 census who had a child born in the same year as the woman. Husband's wage-income rank uses actual 1940 wage income and ranks him against White married, spouse-present men born in the same year in the full-count 1940 census; valid zero income is retained. Rank differences are reported only if both ranks are observed. "Effective N" is the number of observations within the outcome-specific bandwidth.

Table A4: Husband Employment and Wage Income, Conditional on Spouse Presence

Outcome	RD Estimate	p-value	Mean Below Cutoff	Effective N	N
Husband in labor force	−0.031	0.007	0.993	2,497	6,554
	(0.013)				
Husband is a farmer	−0.018	0.480	0.259	2,058	6,554
	(0.048)				
Husband has positive wage income	−0.073	0.091	0.667	2,265	5,519
	(0.052)				
Husband wage income	−124.6	0.190	1,157.6	3,066	5,519
	(112.2)				
Husband wage income, conditional on positive	−53.4	0.584	1,736.4	1,664	3,643
	(137.8)				

Notes: Each row reports a local-linear RD estimate conditional on having a linked coresident husband and observing the relevant outcome. Robust bias-corrected standard errors in parentheses. "Farmer" includes farmers and farm managers (1950 occupation codes 100 and 123). Husband wage income includes valid zeros; the final row is additionally restricted to positive wage income. "Effective N" is the number of observations within the outcome-specific MSE-optimal bandwidth.

Table A5: Income-Data Availability at the Threshold

	Husband Income Missing Unconditional	Husband Income Missing Spouse Present	Household Income Observed
RD Estimate	+0.009	+0.018	−0.015
	(0.021)	(0.028)	(0.032)
p-value	0.834	0.660	0.635
Left Mean	0.067	0.090	0.814
Effective N	3,413	2,541	3,968
N	8,718	6,502	8,816

Notes: Each column reports a local-linear RD estimate at the certification threshold using a triangular kernel and an outcome-specific MSE-optimal bandwidth. Robust bias-corrected standard errors in parentheses. Column 1 indicates that a woman has a coresident husband whose wage income is unavailable, among women with valid own wage income. Column 2 reports the probability that husbands' income is missing, conditional on having a coresident husband and a valid own-income (0 or positive). "Effective N" is the number of observations used within the optimally selected estimation bandwidth.

Table A6: Heterogeneity by Marriage-Bar County

	No Bar		Bar County		
Outcome	Estimate	N	Estimate	N	Difference p-value
Teacher, 1930	+0.083	4,815	+0.055	3,127	0.78
	(0.096)		(0.312)		
Teacher, 1940	+0.122	4,425	+0.085	2,864	0.72
	(0.016)		(0.083)		
In labor force, 1930	+0.059	4,815	+0.070	3,127	0.92
	(0.269)		(0.287)		
In labor force, 1940	+0.095	4,425	+0.091	2,864	0.98
	(0.061)		(0.096)		
Years of schooling	+0.628	4,122	−0.179	2,696	0.10
	(0.004)		(0.398)		
Ever married, 1930	−0.062	4,815	−0.105	3,127	0.69
	(0.256)		(0.048)		
Ever married, 1940	−0.055	4,425	−0.057	2,864	0.98
	(0.263)		(0.344)		
Estimated fertility	−0.309	3,887	−0.327	3,010	0.97
	(0.249)		(0.354)		
Estimated fertility | positive	−0.355	2,702	−0.281	2,095	0.88
	(0.164)		(0.355)		
Age at first birth	+1.014	2,605	−0.017	1,989	0.36
	(0.062)		(0.778)		
No children (1930–40)	+0.011	3,887	+0.070	3,010	0.61
	(0.996)		(0.348)		
Own wage income	+109.4	4,369	+124.7	2,836	0.91
	(0.073)		(0.044)		
Married × Teacher, 1940	+0.057	4,425	−0.000	2,864	0.31
	(0.045)		(0.970)		
Unmarried × Teacher, 1940	+0.059	4,425	+0.089	2,864	0.73
	(0.126)		(0.052)		

Notes: Subgroup RD estimates (MSE-optimal bandwidths, HC3 standard errors). Robust bias-corrected p-values in parentheses. A county is a "bar county" if any city in it ever implemented or was reported operating a married-teacher bar. Counties are classified by women's census county of residence in either 1930 or 1940. The married- and unmarried-teacher rows split teaching in 1940 by spouse presence in 1940. "Difference p-value" is the p-value for the difference between the two subgroup effects, following Calonico et al. (2025). N is per subgroup.

Table A7: Heterogeneity by Farm versus Nonfarm Household

	Nonfarm		Farm		
Outcome	Estimate	N	Estimate	N	Difference p-value
Teacher, 1930	+0.075	4,240	+0.045	4,529	0.74
	(0.099)		(0.412)		
Teacher, 1940	+0.082	4,040	+0.092	4,337	0.90
	(0.058)		(0.043)		
In labor force, 1930	+0.052	4,240	+0.074	4,529	0.81
	(0.305)		(0.107)		
In labor force, 1940	+0.064	4,040	+0.102	4,337	0.72
	(0.250)		(0.050)		
Years of schooling	+0.170	3,754	+0.438	4,074	0.55
	(0.374)		(0.100)		
Ever married, 1930	−0.024	4,240	−0.090	4,529	0.51
	(0.860)		(0.084)		
Ever married, 1940	−0.007	4,040	−0.063	4,337	0.52
	(0.977)		(0.071)		
Estimated fertility	−0.001	3,552	−0.471	3,865	0.25
	(0.818)		(0.025)		
Estimated fertility | positive	−0.061	2,413	−0.328	2,819	0.53
	(0.946)		(0.147)		
Age at first birth	+0.790	2,309	+0.417	2,707	0.75
	(0.211)		(0.567)		
No children (1930–40)	−0.016	3,552	+0.083	3,865	0.34
	(0.737)		(0.049)		
Own wage income	+21.0	3,990	+166.4	4,291	0.30
	(0.742)		(0.013)		

Notes: Subgroup RD estimates (MSE-optimal bandwidths, HC3 standard errors). Robust bias-corrected p-values in parentheses. The sample is split the sample by whether the woman's father was a farmer (occupation code 100) or held any other occupation. Father's occupation is taken from the woman's 1910 household—recovered via her own childhood census record where her baseline census link falls later—covering roughly 93 percent of the sample; the split is close to even (49.6 percent farm). "Difference p-value" is the p-value for the difference between the two subgroup effects, following Calonico et al. (2025). N is per subgroup.

A.1 Alternative Accounting Calculations

A liberal calculation using all white-collar work. To gauge what the estimates imply for white-collar work more broadly, I extend them to the era's other new white-collar entrants.[22] Among American women aged 18–24, employment in white-collar occupations rose from 3.4 percent in 1880 to 19.7 percent in 1920, accounting for the entire net growth in young women's employment. Scaling my estimates to this population requires two changes. Because passing raised the probability of ever being certified by 47 percentage points, the effect per certified woman is about twice the reduced form (18.3 percentage points on 1940 participation and −0.68 children). The ideal data would allow us to identify what fraction of women are induced to ever be employed as a teacher as a result of certification. However, in the absence of this, we can assume that the likely first stage is somewhere between the short-run teaching margin (20 percentage points)[23] and the first stage on certification (47 percentage points),[24] which imply an effect per entrant of 18.3 to 42.6 percentage points on 1940 participation and −0.68 to −1.58 children. And because my estimates come from women whose alternative was not working, I count only the 61 percent of new entrants who came from nonemployment.[25] Under the assumption that the persistence effects of early-career entry extend to these occupations, additional early-career entry accounts for roughly 16 to 37 percent of the rise in American women's prime-age labor force participation between these cohorts, and roughly 3 to 8 percent of the national decline in completed fertility. The two bounds differ only in the first stage.

$$\text{Prime-age labor force:}\quad \frac{0.164 \times 0.61 \times 18.3 \text{ pp}}{27.1\% - 15.7\%} \approx 16\% \quad \text{to} \quad \frac{0.164 \times 0.61 \times 42.6 \text{ pp}}{27.1\% - 15.7\%} \approx 37\%,$$

$$\text{Completed fertility:}\quad \frac{0.164 \times 0.61 \times 0.68 \text{ children}}{4.71 - 2.76 \text{ children}} \approx 3\% \quad \text{to} \quad \frac{0.164 \times 0.61 \times 1.58 \text{ children}}{4.71 - 2.76 \text{ children}} \approx 8\%.$$

[22] Appendix Table A8 reports the same calculation for all teachers, rural teachers, all professional occupations, and all white-collar work.

[23] This is the 1920 effect for exam takers in 1918–1919.

[24] This assumes that all women who gained teaching credentials entered.

[25] The 61 percent follows from where the entrants came from. Among women aged 18–24, nonemployment fell from 71.4 to 61.4 percent and other gainful work from 25.2 to 18.9 percent. Restricting the exercise to downstate Illinois—the setting of my estimates—yields nearly identical results, with white-collar entry rising from 4.2 to 18.9 percent. Within Illinois, the persistence rate is nearly identical whether women began as teachers (36.1 percent), other professionals (37.9 percent), or clerical workers (36.2 percent).

Table A8: Share of the Aggregate Change Explained by Additional Early-Career Entry

Additional early-career entry into	Growth in entry (pp)	Share of the aggregate change: Prime-age LFP	Share of the aggregate change: Completed fertility
Certification margin only, downstate IL (no scaling)	1.2	1%	0.2%
All teachers, US	1.5	1–3%	0.3–0.8%
All teachers, rural US	2.4	6–13%	0.5–1.2%
All professional occupations, US	2.5	2–6%	0.5–1.2%
All white-collar work, US (if clerical ≈ teaching)	16.4	16–37%	3–8%

Notes: Each share is (growth in entry) × (effect per entrant) ÷ (aggregate change). Row 1 uses the reduced-form estimates (+8.6 pp on 1940 labor force participation; −0.32 children) and counts the 1.2 percent of the downstate 1897–1902 birth cohorts who sat the exam and passed within 3 percentage points of the threshold (0.113×0.102); its denominators are downstate-Illinois women (participation 10.5 to 21.5 percent at ages 35–44; completed fertility 4.83 to 2.67 children at ages 45–54). The remaining rows scale per certified woman, dividing the reduced form by the 0.47 first stage on certification (+18.3 pp; −0.68 children) for the lower bound and by the 0.202 first stage on teaching in the 1920 census (1918–1919 exam cohorts; Table 3, column 2; +42.6 pp; −1.58 children) for the upper bound, and count the growth in entry between women aged 18–24 in 1880 and in 1920: teaching 2.3 to 3.8 percent (rural 3.0 to 5.4), professional occupations 2.7 to 5.2 percent, and all white-collar work 3.4 to 19.7 percent. Their denominators are national women aged 35–44 in 1900 and 1940 (participation 15.7 to 27.1 percent; rural 12.7 to 17.4) and aged 45–54 (completed fertility 4.71 to 2.76 children; rural 5.15 to 3.24). Each scaled row applies the per-entrant effect to the 61 percent of new entrants coming from nonemployment.

Table A9: Decomposition of Husband-Contributed Wage Income

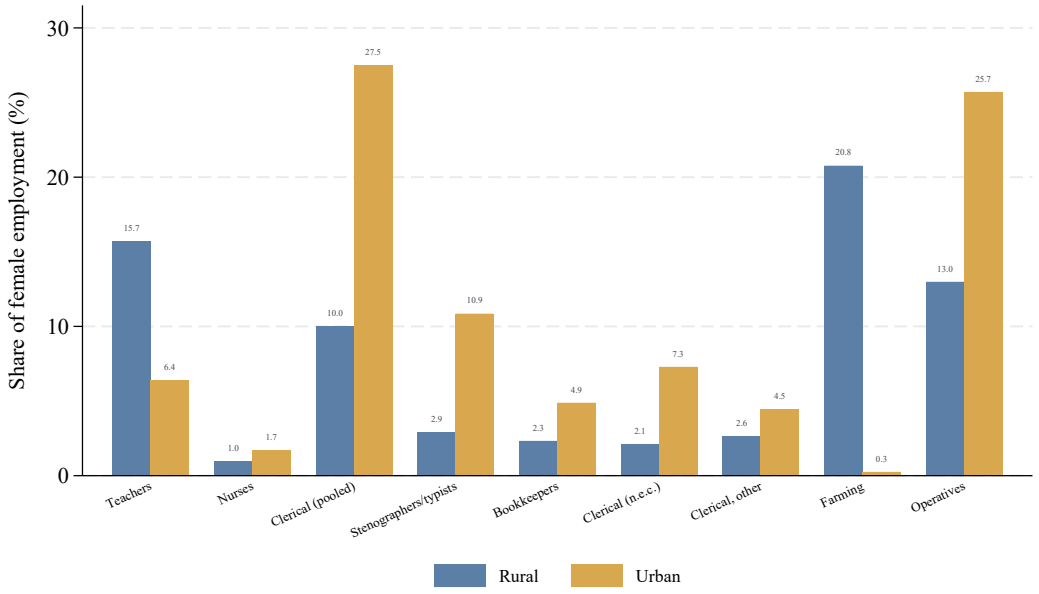
Component	Change in Underlying Margin	Contribution to RD Estimate (dollars)
Total husband-contributed wage income		−155.3
Change in spouse presence	−0.085	−98.9
Change in conditional husband wage income	−86.0	−63.7
Interaction		+7.3

Notes: The table decomposes the RD estimate for husband-contributed wage income into an extensive spouse-presence margin, an intensive conditional-husband-income margin, and their interaction. The decomposition uses a common bandwidth based on the selected bandwidth for the household-income sample, so the coefficients sum to −155.3 ($-98.9 - 63.7 + 7.3 = -155.3$).

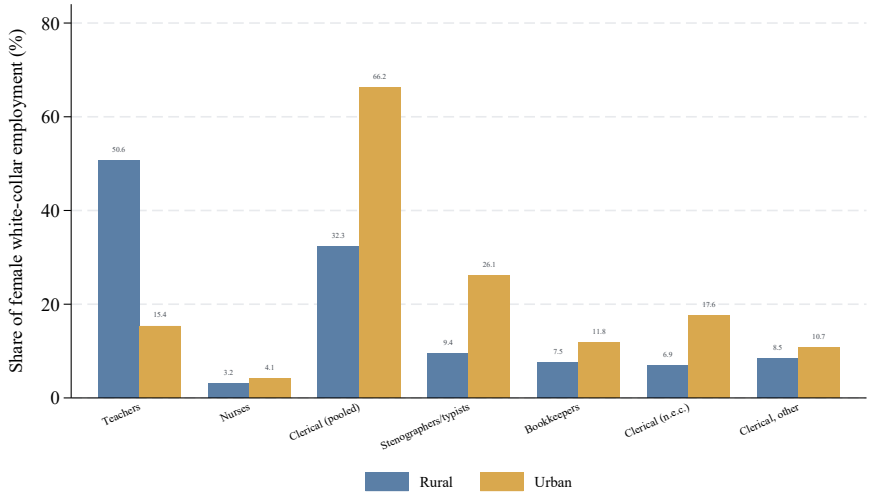
B Appendix Figures

Figure B1: Teaching Pay versus State Civil-Service Pay, 1922

Notes: Bars show monthly salary ranges of women-relevant Illinois civil-service job classes, from the 1922 classification of the Illinois State Civil Service Commission, colored by the stated education requirement; (+ board) indicates the class included maintenance (board and lodging). Civil-service salaries are paid year-round. Vertical lines show the average teacher salary in 1922 (men and women pooled, as the civil-service classes report no sex breakdown) from the State Superintendent's biennial reports. Blue shows Illinois excluding Cook County ($1,046 per year; approximately eight-month term), and purple shows Illinois including Cook County ($1,343 per year; teacher-weighted term of roughly 8.5 months, as Chicago's term ran 9.5 months). The solid line divides annual salary by months in session, and the dashed line divides salary by twelve months.



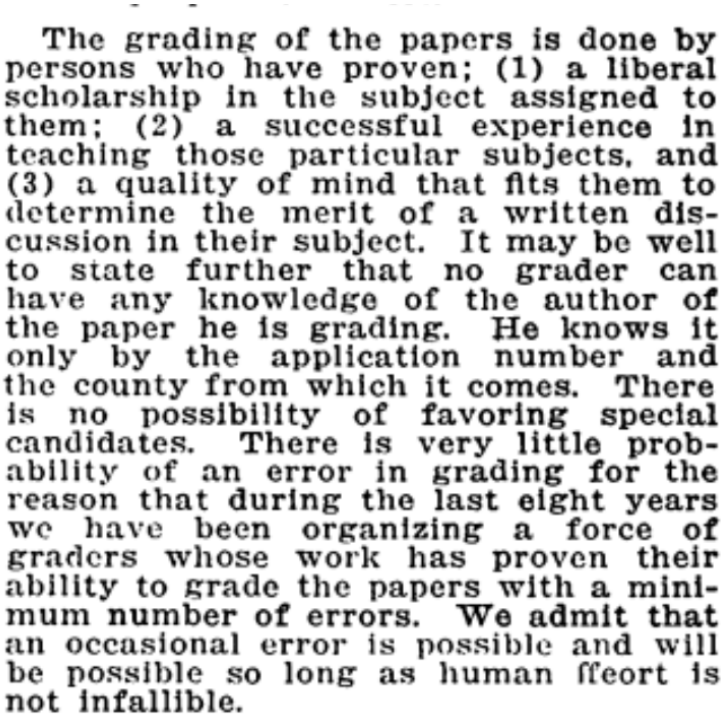
(a) Share of all female employment



(b) Share of female white-collar employment

Figure B2: Women's Occupation Shares by Rural and Urban Residence, 1920

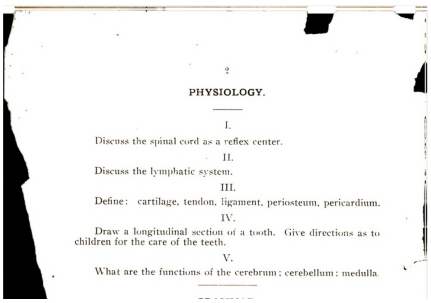
Notes: The sample is white women aged 19 and older in the labor force with a gainful occupation in the full-count 1920 census. Clerical (pooled) is defined as occ1950 codes 300–390, which is a cumulative of other standalone clerical occupations that are reported (stenographers/typists, bookkeepers, clerical n.e.c., other). White collar (panel b) is occ1950 1–99 and 200–399 (professional, managerial, and clerical occupations); farming and operative occupations are therefore shown only in panel (a).



The grading of the papers is done by persons who have proven; (1) a liberal scholarship in the subject assigned to them; (2) a successful experience in teaching those particular subjects, and (3) a quality of mind that fits them to determine the merit of a written discussion in their subject. It may be well to state further that no grader can have any knowledge of the author of the paper he is grading. He knows it only by the application number and the county from which it comes. There is no possibility of favoring special candidates. There is very little probability of an error in grading for the reason that during the last eight years we have been organizing a force of graders whose work has proven their ability to grade the papers with a minimum number of errors. We admit that an occasional error is possible and will be possible so long as human ffeort is not infallible.

Figure B3: Note on Grading from the Department of Public Instruction

Notes: Excerpt from the Illinois Department of Public Instruction's *Educational Press Bulletin* (Illinois Department of Public Instruction, 1923).



2

PHYSIOLOGY.

I.

Discuss the spinal cord as a reflex center.

II.

Discuss the lymphatic system.

III.

Define: cartilage, tendon, ligament, periosteum, pericardium.

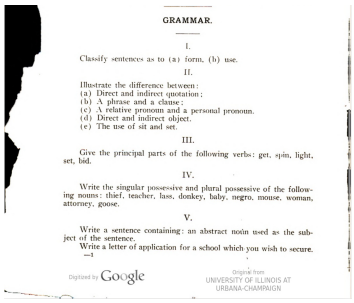
IV.

Draw a longitudinal section of a tooth. Give directions as to children for the care of the teeth.

V.

What are the functions of the cerebrum; cerebellum; medulla.

(a) Physiology



GRAMMAR.

I.

Classify sentences as to (a) form, (b) use.

II.

Illustrate the difference between:
(a) Direct and indirect quotation;
(b) A phrase and a clause;
(c) A relative pronoun and a personal pronoun.
(d) Direct and indirect object.
(e) The use of sit and set.

III.

Give the principal parts of the following verbs: get, spin, light, set, bid.

IV.

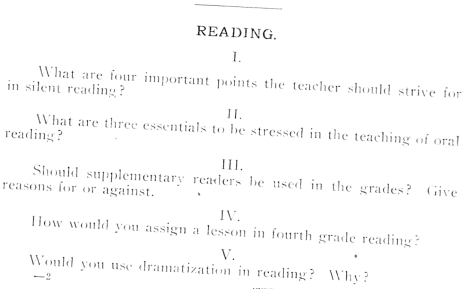
Write the singular possessive and plural possessive of the following nouns: thief, teacher, lass, donkey, baby, negro, mouse, woman, attorney, goose.

V.

Write a sentence containing: an abstract noun used as the subject of the sentence.
Write a letter of application for a school which you wish to secure.
—1

Digitized by Google
Original from UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

(b) Grammar



READING.

I.

What are four important points the teacher should strive for in silent reading?

II.

What are three essentials to be stressed in the teaching of oral reading?

III.

Should supplementary readers be used in the grades? Give reasons for or against.

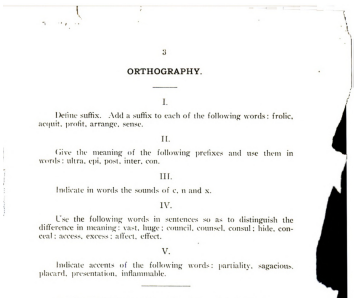
IV.

How would you assign a lesson in fourth grade reading?

V.

Would you use dramatization in reading? Why?
—2

(c) Reading



3

ORTHOGRAPHY.

I.

Define suffix. Add a suffix to each of the following words: frolic, acquit, profit, arrange, sense.

II.

Give the meaning of the following prefixes and use them in words: ultra, epi, post, inter, con.

III.

Indicate in words the sounds of c, n and x.

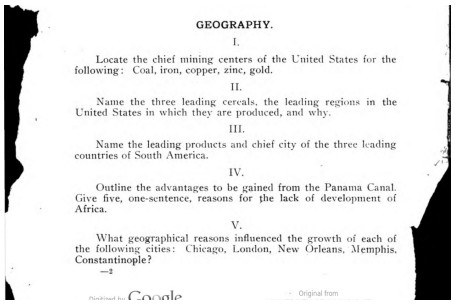
IV.

Use the following words in sentences so as to distinguish the difference in meaning: vast, huge; council, counsel, consul; hide, conceal; access, excess; affect, effect.

V.

Indicate accents of the following words: partiality, sagacious, placard, presentation, inflammable.

(d) Orthography



GEOGRAPHY.

I.

Locate the chief mining centers of the United States for the following: Coal, iron, copper, zinc, gold.

II.

Name the three leading cereals, the leading regions in the United States in which they are produced, and why.

III.

Name the leading products and chief city of the three leading countries of South America.

IV.

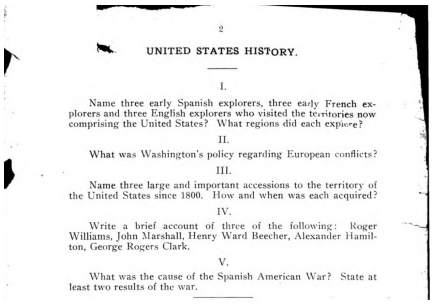
Outline the advantages to be gained from the Panama Canal. Give five, one-sentence, reasons for the lack of development of Africa.

V.

What geographical reasons influenced the growth of each of the following cities: Chicago, London, New Orleans, Memphis, Constantinople?
—2

Digitized by Google
Original from

(e) Geography



2

UNITED STATES HISTORY.

I.

Name three early Spanish explorers, three early French explorers and three English explorers who visited the territories now comprising the United States? What regions did each explore?

II.

What was Washington's policy regarding European conflicts?

III.

Name three large and important accessions to the territory of the United States since 1800. How and when was each acquired?

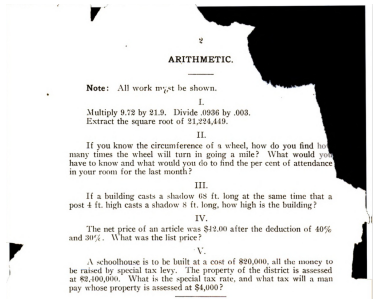
IV.

Write a brief account of three of the following: Roger Williams, John Marshall, Henry Ward Beecher, Alexander Hamilton, George Rogers Clark.

V.

What was the cause of the Spanish American War? State at least two results of the war.

(f) United States History



2

ARITHMETIC.

Note: All work must be shown.

I.

Multiply 9.72 by 21.9. Divide .0936 by .003. Extract the square root of 21,224,449.

II.

If you know the circumference of a wheel, how do you find how many times the wheel will turn in going a mile? What would you have to know and what would you do to find the per cent of attendance in your room for the last month?

III.

If a building casts a shadow 68 ft. long at the same time that a post 4 ft. high casts a shadow 8 ft. long, how high is the building?

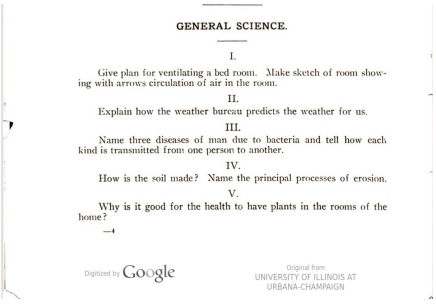
IV.

The net price of an article was $12.00 after the deduction of 40% and 20%. What was the list price?

V.

A schoolhouse is to be built at a cost of $20,000, all the money to be raised by special tax levy. The property of the district is assessed at $2,400,000. What is the special tax rate, and what tax will a man pay whose property is assessed at $4,000?

(g) Arithmetic



GENERAL SCIENCE.

I.

Give plan for ventilating a bed room. Make sketch of room showing with arrows circulation of air in the room.

II.

Explain how the weather bureau predicts the weather for us.

III.

Name three diseases of man due to bacteria and tell how each kind is transmitted from one person to another.

IV.

How is the soil made? Name the principal processes of erosion.

V.

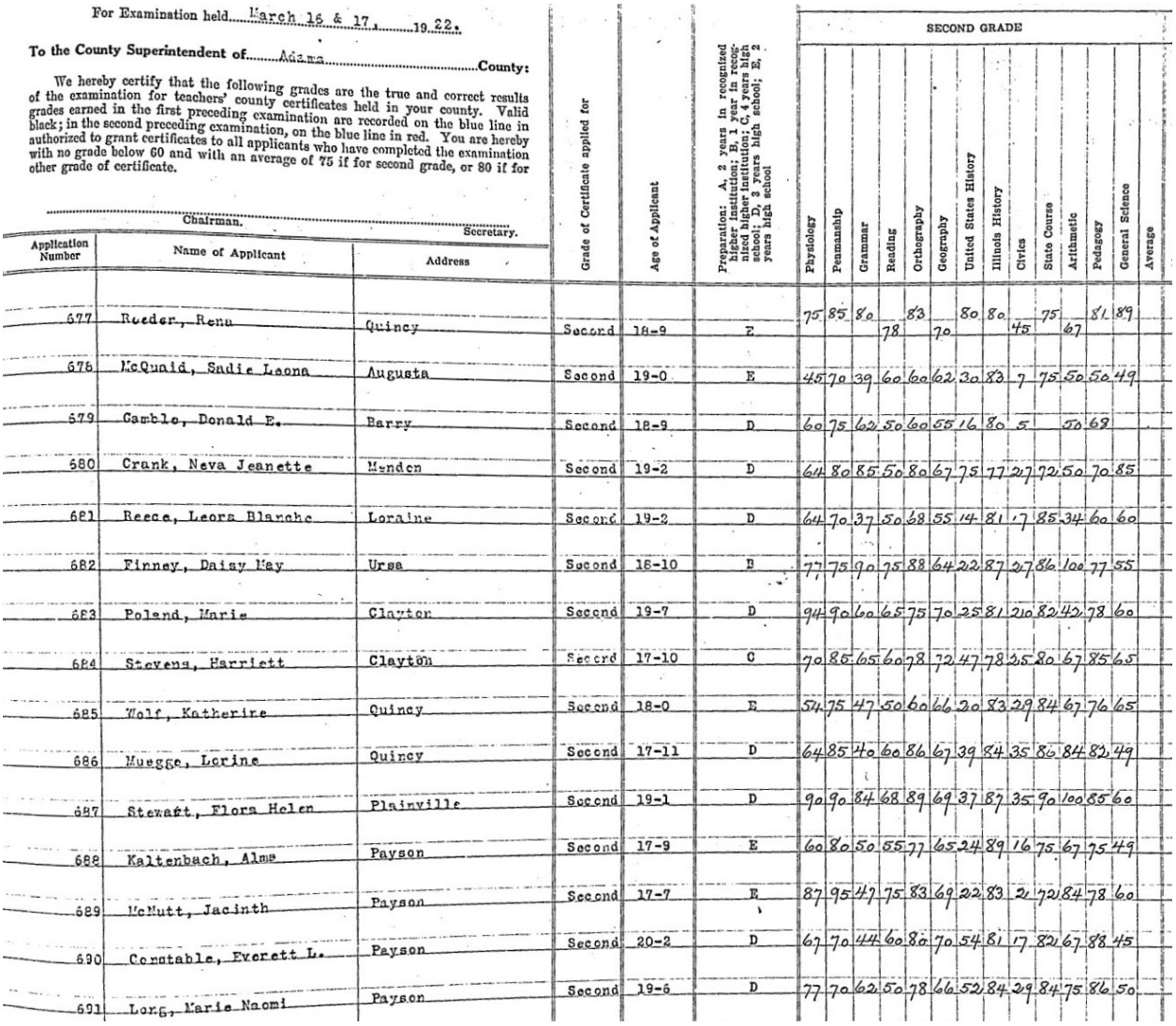
Why is it good for the health to have plants in the rooms of the home?
—4

Digitized by Google
Original from UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

(h) General Science

Figure B4: Sample Examination Questions

Notes: Questions from eight of the 13 subjects. Each subject had five questions.



For Examination held March 16 & 17, 1922.

To the County Superintendent of Adams County:

We hereby certify that the following grades are the true and correct results of the examination for teachers' county certificates held in your county. Valid grades earned in the first preceding examination are recorded on the blue line in black; in the second preceding examination, on the blue line in red. You are hereby authorized to grant certificates to all applicants who have completed the examination with no grade below 60 and with an average of 75 if for second grade, or 80 if for other grade of certificate.

Chairman. Secretary.

Application Number	Name of Applicant	Address	Grade of Certificate applied for	Age of Applicant	Preparation: A, 2 years in recognized higher institution; B, 1 year in recognized higher institution; C, 4 years high school; D, 3 years high school; E, 2 years high school	SECOND GRADE													
						Physiology	Penmanship	Grammar	Reading	Orthography	Geography	United States History	Illinois History	Civics	State Course	Arithmetic	Pedagogy	General Science	Average
677	Reeder, Rena	Quincy	Second	18-9	E	75	85	80	78	83	70	80	80	45	75	67	81	89	
678	McQuaid, Sadie Leona	Augusta	Second	19-0	E	45	70	39	60	60	62	30	83	7	75	50	50	49	
679	Gamble, Donald E.	Barry	Second	18-9	D	60	75	62	50	60	55	16	80	5		50	69		
680	Crank, Neva Jeanette	Mendon	Second	19-2	D	64	80	85	50	80	67	75	77	27	72	50	70	85	
681	Reece, Leora Blanche	Loraine	Second	19-2	D	64	70	37	50	68	55	14	81	17	85	34	60	60	
682	Finney, Daisy May	Ursa	Second	18-10	B	77	75	90	75	88	64	22	87	27	86	100	77	55	
683	Poland, Marie	Clayton	Second	19-7	D	94	90	60	65	75	70	25	81	20	82	42	78	60	
684	Stevens, Harriett	Clayton	Second	17-10	C	70	86	65	60	78	72	47	78	25	80	67	85	65	
685	Wolf, Katherine	Quincy	Second	18-0	E	54	75	47	50	60	66	20	83	29	84	67	76	65	
686	Muegge, Lorine	Quincy	Second	17-11	D	64	85	40	60	86	67	39	84	35	86	84	82	49	
687	Stewart, Flora Helen	Plainville	Second	19-1	D	90	90	84	68	89	69	37	87	35	90	100	85	60	
688	Kaltenbach, Alma	Payson	Second	17-9	E	60	80	50	65	77	65	24	89	16	75	67	75	49	
689	McNutt, Jacinth	Payson	Second	17-7	E	87	95	47	75	83	69	22	83	2	72	84	78	60	
690	Constable, Everett L.	Payson	Second	20-2	D	67	70	44	60	80	70	54	81	17	82	67	88	45	
691	Long, Marie Naomi	Payson	Second	19-6	D	77	70	62	50	78	66	52	84	29	84	75	86	50	

Figure B5: Example Examination Ledger Page

Notes: A page from the Report of the Illinois State Examining Board (Adams County, examination held March 1922). Each row records an applicant's application number, name, address, grade of certificate applied for, age, preparation code, and score on each of the 13 subject examinations. *Source:* Illinois State Archives (Illinois State Archives, 1923).

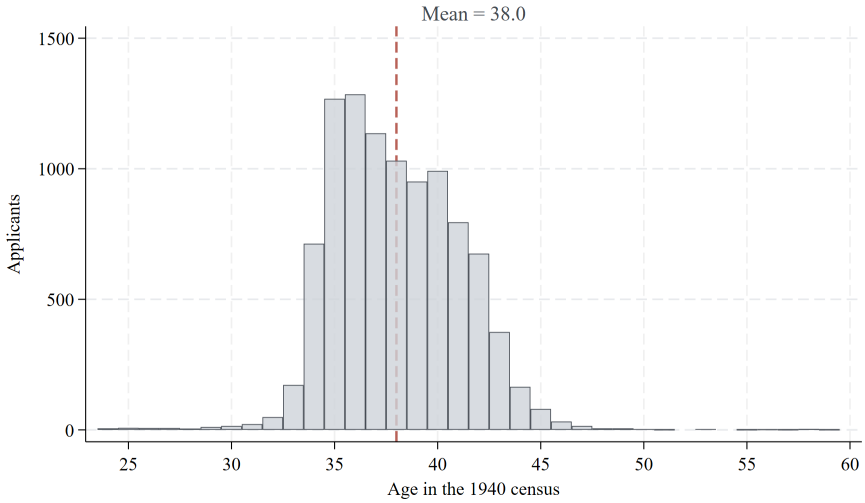


Figure B6: Age Distribution of Applicants in the 1940 census

Notes: Distribution of applicants' ages in the 1940 census, among sample women linked to 1940. The dashed line marks the mean (38.0).

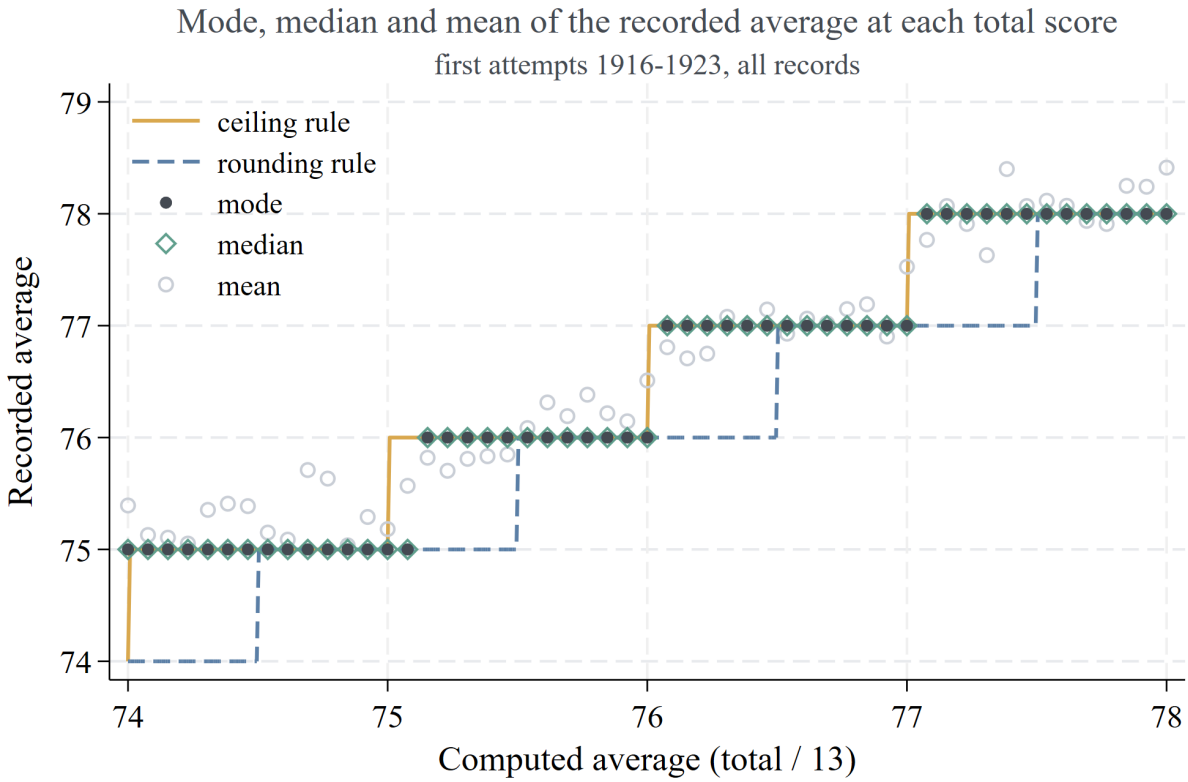


Figure B7: Rounding Regime: Recorded versus Computed Averages

Notes: This figure plots the calculated average versus the recorded average for test-takers' first complete 13-subject attempts, 1916–1923. The horizontal axis shows the calculated average (total score divided by 13). The vertical axis represents the average recorded by the Examining Board; recorded averages are always integers. Markers denote the mode, median, and mean recorded average at each total score. The solid line represents how a computed average would be recorded if graders followed a "ceiling" regime. The dashed line shows how a computed average would be recorded if graders rounded to the nearest whole number. The plot is limited to grades of 74 percent or higher because graders did not always record their calculated average, and they did so disproportionately when the applicant met the passing average, and often not even then (Appendix Figure B8).

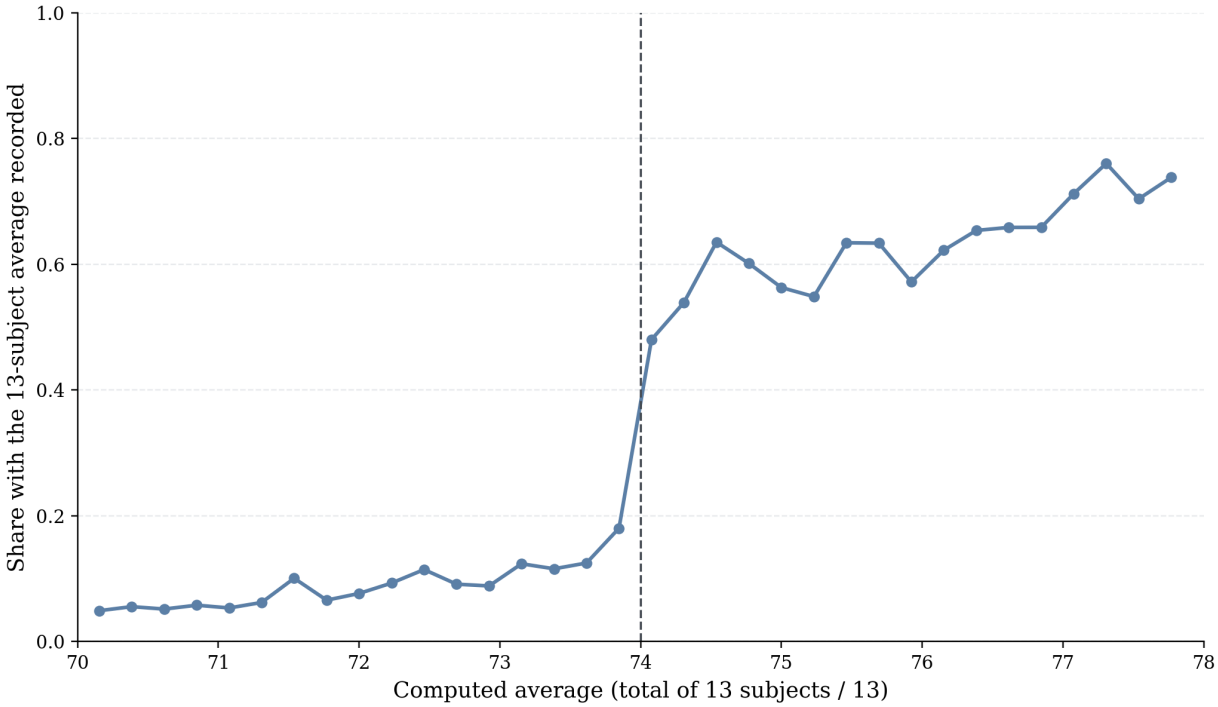
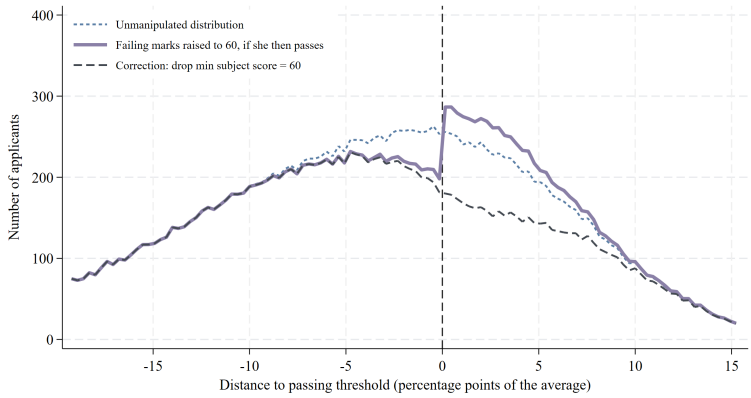
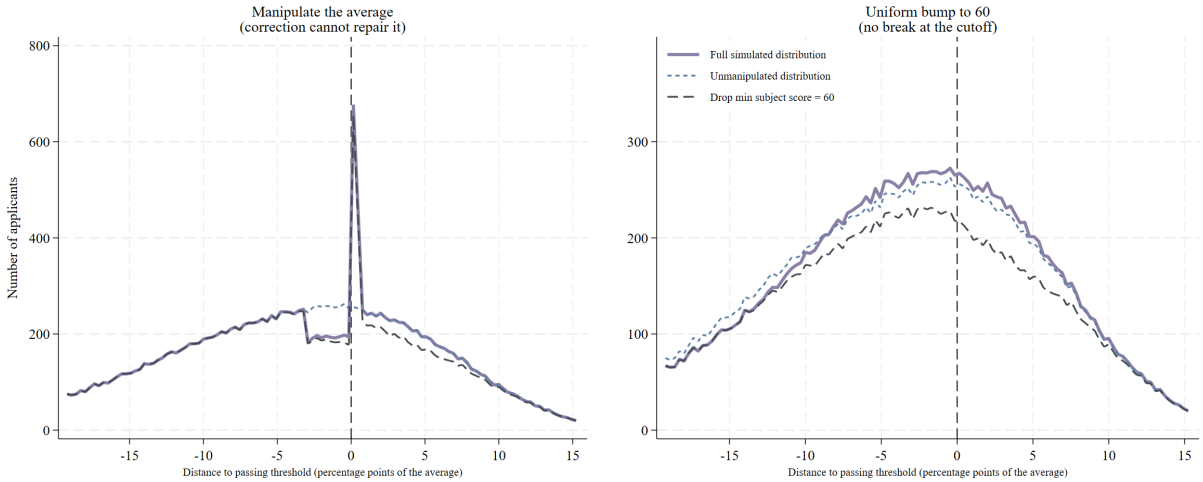


Figure B8: Share of Applicants with the Average Recorded in the Ledger

Notes: This figure plots the probability that 13-subject scores are reported alongside a recorded average, in bins of three total-score points, for all sittings, 1916–1923. The horizontal axis presents the calculated average (total score divided by 13). The dashed line marks an average of 74.



(a) Minimum bumped to 60 when this lets the candidate pass



(b) Manipulating the average (left) and a uniform bump to 60 (right)

Figure B9: Simulated Grading Behavior

Notes: Each panel plots a simulated distribution of the exam score, in percentage points from the average, recentered so the passing threshold falls at zero, under a different grading regime. Each panel shows the full distribution (solid), the distribution after dropping candidates whose minimum subject score equals exactly 60 (dashed), and the true unmanipulated distribution (dotted). The three regimes are (a) a subject-60 bump applied only to candidates for whom this change would allow them to pass the exam; (b, left) direct manipulation of the average across the cutoff; and (b, right) a subject-60 bump applied to a share of candidates regardless of their average. Dropping minimum-score 60s recovers a smooth density under regime (a) and leaves the already-smooth density of the uniform bump smooth, but does not repair direct manipulation of the average. Because some candidates score a minimum of 60 on their own merit, the exclusion rule also removes genuine, unmanipulated records, which results in additional depletion of the recovered unmanipulated distribution (dotted lines); for more information, see Figure B11.

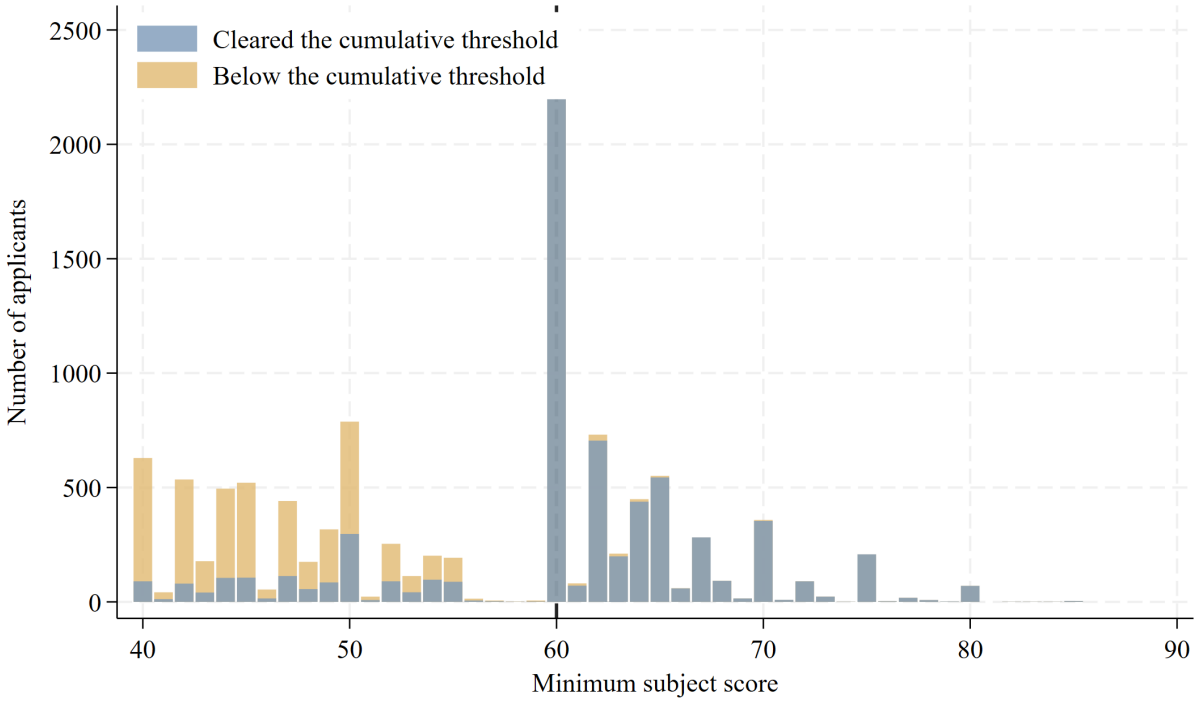


Figure B10: Minimum Subject Score and Grade Bumping

Notes: Distribution of the minimum subject score among first-attempt applicants with complete 13-subject exams, split by whether the candidate cleared the cumulative passing threshold on that attempt.

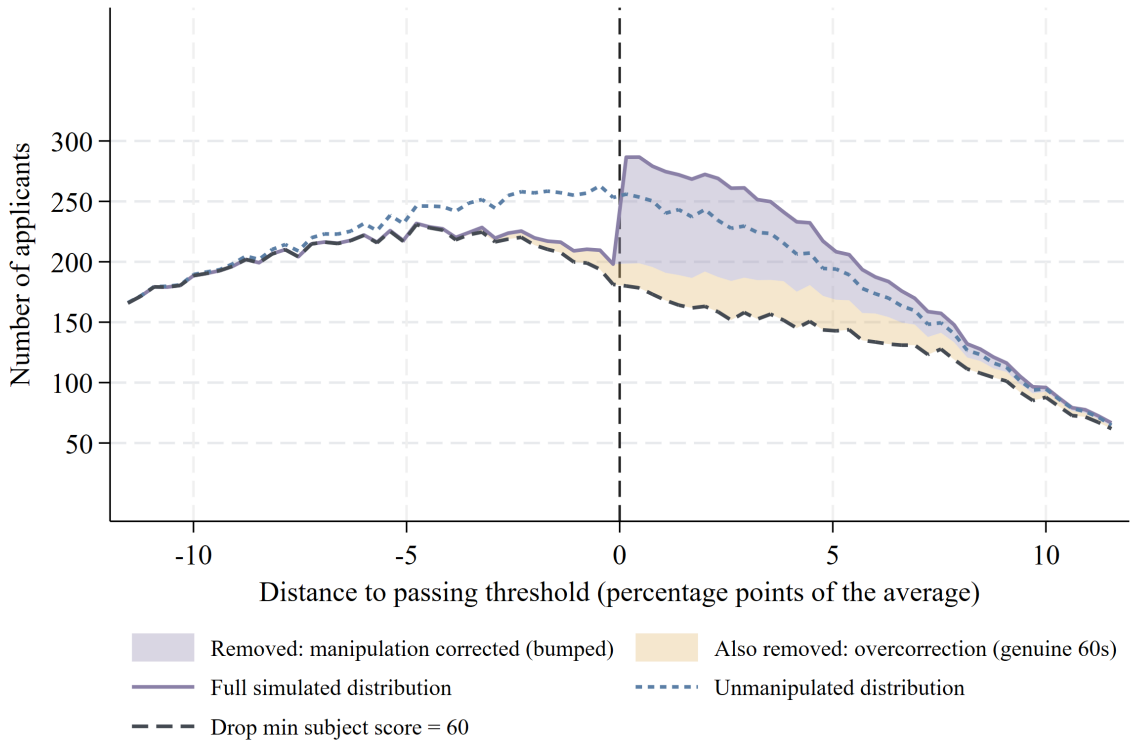


Figure B11: Correcting Manipulation versus Overcorrection

Notes: Panel (a) of Figure B9, enlarged around the passing threshold. Dropping candidates with a minimum subject score of exactly 60 removes the manipulated records bumped up to 60 (purple) but also excludes applicants who scored a minimum of 60 on their own merit (gold).

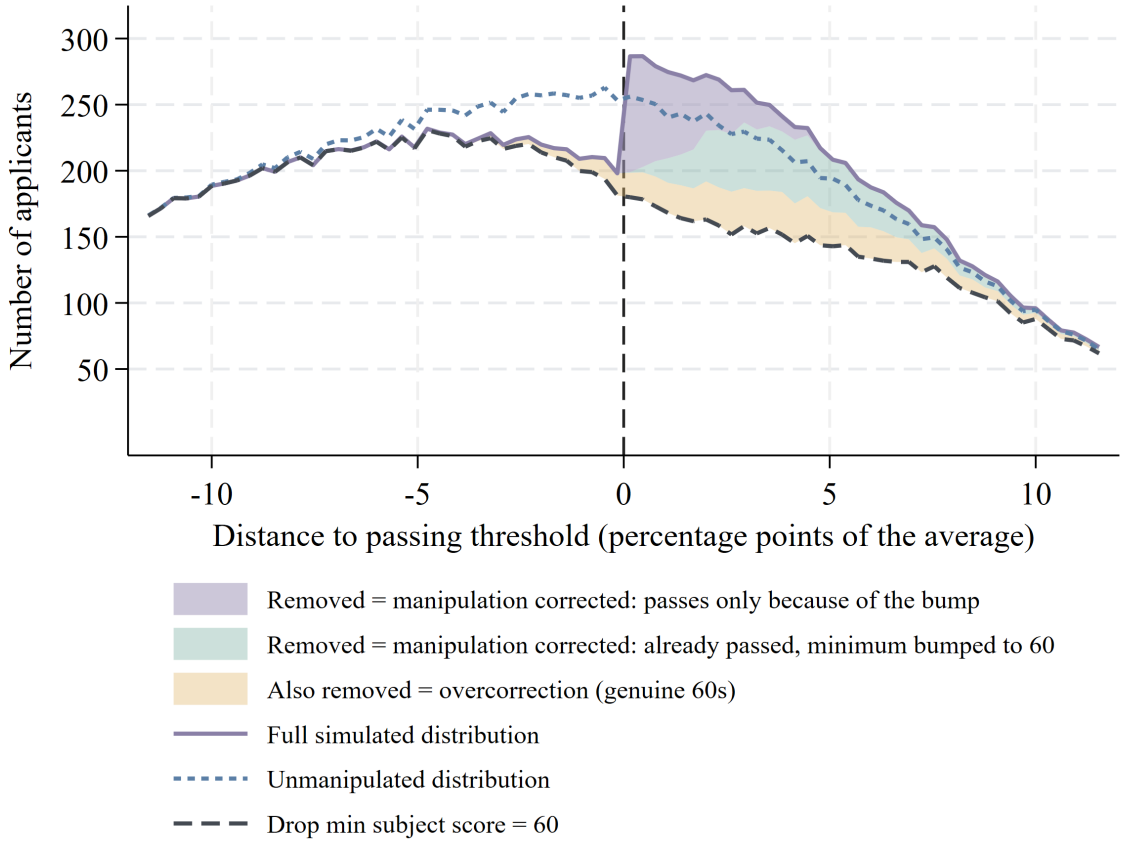


Figure B12: Correcting Two Forms of Manipulation versus Overcorrection

Notes: Panel (a) of Figure B9, enlarged around the passing threshold. Excluding candidates with a minimum subject score of exactly 60 removes three groups. These are candidates whose average passes only because a failing subject score is raised to 60 (purple), candidates who had already cleared the average and whose failing subject score is raised to 60 (teal), and candidates who genuinely scored a minimum of 60 (gold).

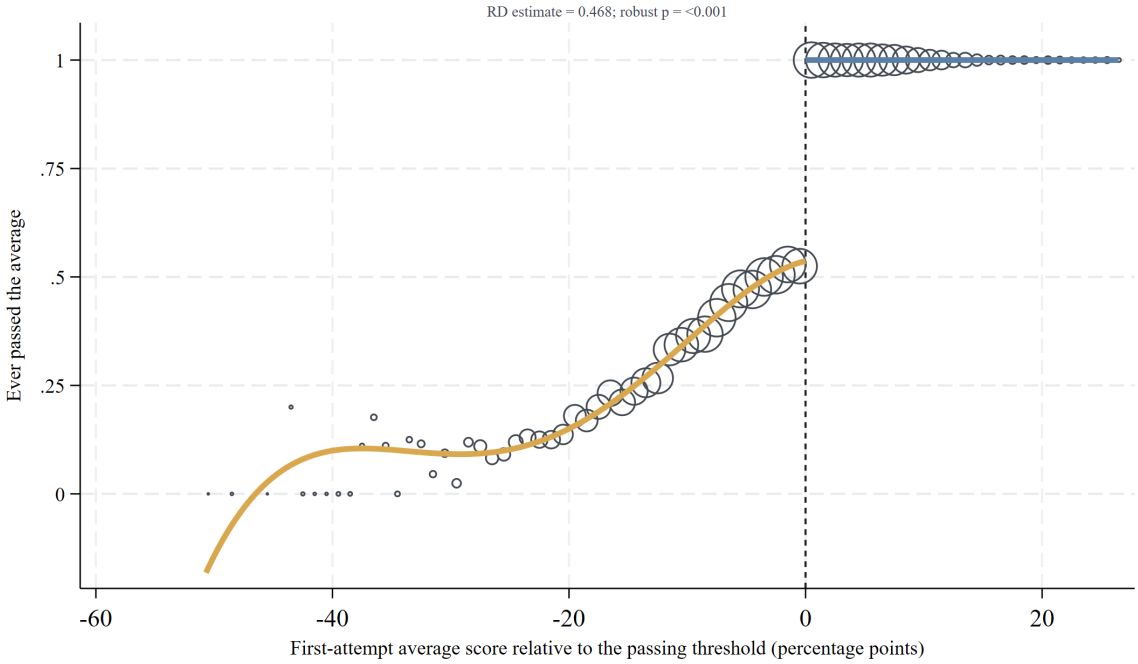
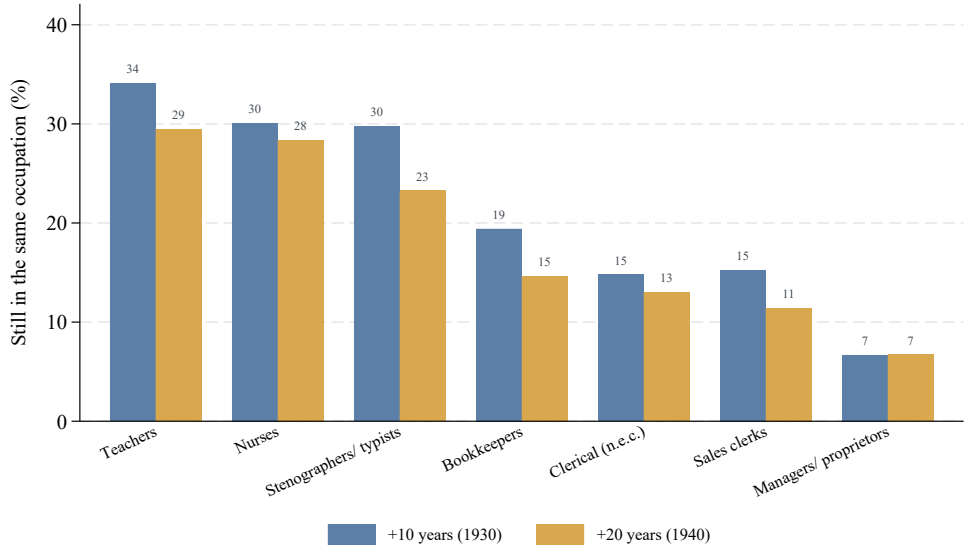
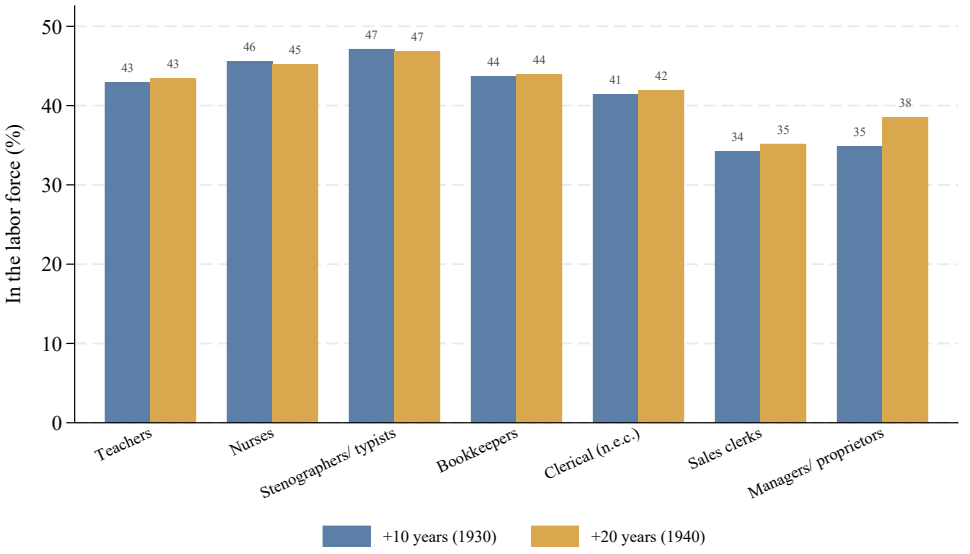


Figure B13: First Stage: Passing the Average on the First Attempt and Ever Passing

Notes: Regression discontinuity plot of ever passing the average on the first-attempt average score relative to the passing threshold, in percentage points, among women aged 17–23 whose first-attempt minimum subject score is not exactly 60. "Ever passed" equals one if the candidate's average cleared the passing threshold in any of her first four sittings (the first attempt plus the next three). Hollow circles are binned means (bins of one percentage point) with area proportional to the number of applicants in the bin; the curves are fourth-order polynomial fits estimated separately on each side of the cutoff. Candidates scoring above the passing threshold on their first attempt passed, whereas roughly half of the marginal failures eventually cleared it (first stage of 0.47; robust $p < 0.001$). Of all candidates who ever pass the average, about 60 percent do so on the first attempt.



(a) Still in the same occupation



(b) In the labor force

Figure B14: Occupational Retention: Teaching versus Other White-Collar Occupations

Notes: The 1920 cohort consists of white women aged 18–30 observed in the occupation in the 1920 census and linked to both the 1930 and 1940 censuses (national linked full-count panel). Panel (a) plots the share of women employed in a given occupation in 1920 who are still recorded as in that same occupation ten and twenty years later (in 1930 and in 1940); panel (b) plots the share of women starting in a given occupation who remain in the labor force (in any occupation).

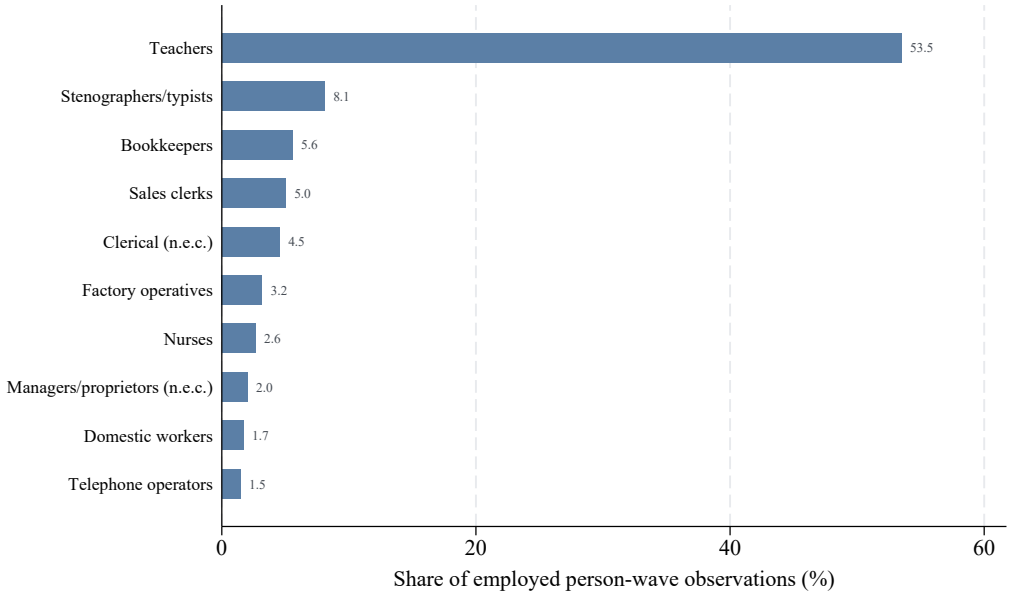


Figure B15: Occupations of Exam-Sample Women, Pooled 1930–1940

Notes: Top ten occupations among exam-sample women observed in the labor force with a gainful occupation, pooling person-wave observations across the 1930 and 1940 censuses. Shares are of all employed person-wave observations.

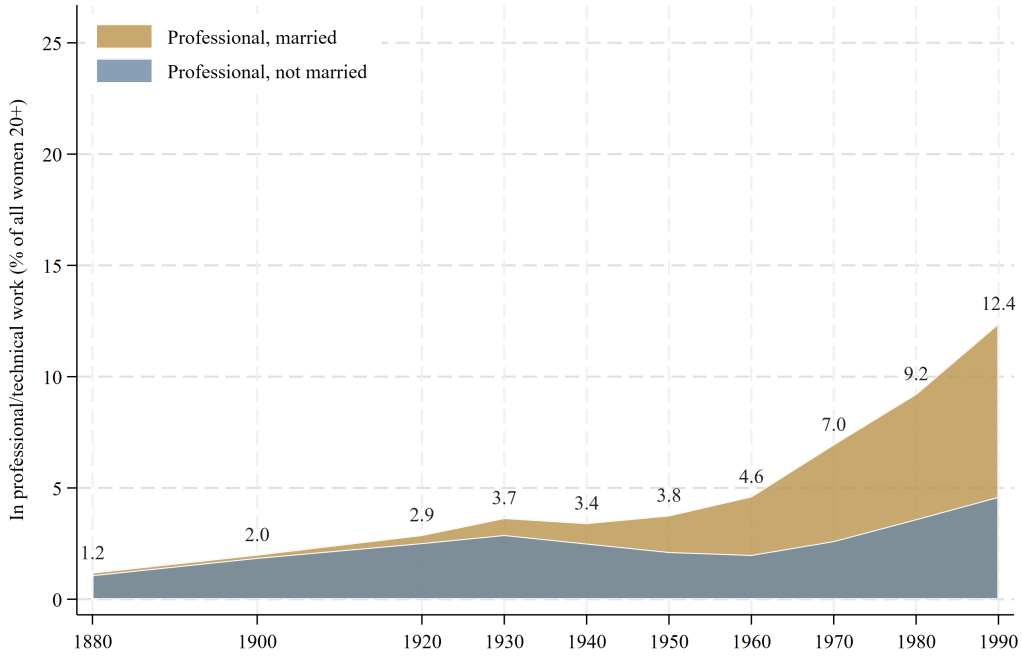


Figure B16: The Rise of Professional and Technical Work among Women, by Marital Status

Notes: The figure plots the share of all women aged 20 and older employed in professional and technical work (1950 occupation codes 1–99), stacked by marital status (currently married versus not currently married); value labels give the total. Data are the IPUMS 1 percent samples of the decennial censuses for 1880–1990 (the 1910 sample is excluded because its enumeration overstates women's gainful employment); estimates are weighted by person weights. *Source:* IPUMS USA 1 percent samples of the decennial censuses (Ruggles et al., 2025).

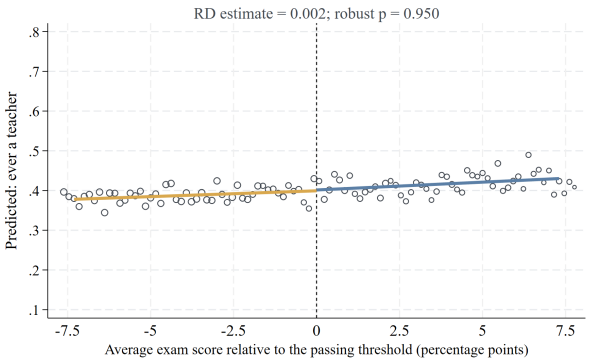
C Predicted Outcomes

This appendix assesses the validity of the continuity in potential outcomes assumption by testing whether preexamination characteristics alone can explain the differences in outcomes observed at the cutoff. To do this, I predict outcomes (teaching, in the labor force, etc.) using age and years of education at the exam, urban residence, whether each parent is present, each parent's labor force status, whether there is any working woman in the household, number of siblings, the household head's occupational score and farm status, birth year, and exam-year fixed effects. I then examine whether the predicted outcome differs at the threshold. Table A10 reports actual and predicted RD estimates for every main outcome. Figures B17–B18 plot the predictions.

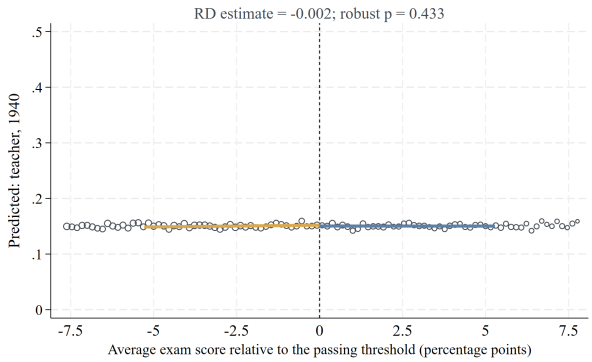
Table A10: Actual and Predicted RD Estimates

	Actual RD estimate (1)	Predicted RD estimate (2)	R^2 of prediction (3)	N (4)
Ever a teacher	+0.133	+0.002	0.086	11,185
	(0.000)	(0.950)		
Teacher, 1920	+0.182	+0.007	0.028	4,698
	(0.002)	(0.371)		
Teacher, 1930	+0.066	−0.001	0.019	9,221
	(0.061)	(0.845)		
Teacher, 1940	+0.089	−0.002	0.004	8,816
	(0.005)	(0.433)		
Ever in the labor force	+0.109	−0.001	0.057	11,185
	(0.003)	(0.721)		
In the labor force, 1920	+0.156	−0.010	0.016	4,698
	(0.006)	(0.115)		
In the labor force, 1930	+0.072	−0.002	0.032	9,221
	(0.055)	(0.747)		
In the labor force, 1940	+0.086	−0.003	0.009	8,816
	(0.022)	(0.428)		
Own wage income, 1940	+91.5	−2.6	0.017	8,718
	(0.056)	(0.529)		
Never married, 1930	+0.068	−0.004	0.042	9,221
	(0.073)	(0.584)		
Never married, 1940	+0.052	−0.001	0.011	8,816
	(0.055)	(0.754)		
Married, spouse present, 1940	−0.057	+0.002	0.010	8,816
	(0.072)	(0.615)		
Number of observed children	−0.319	+0.004	0.022	7,784
	(0.029)	(0.859)		
No children by 1940	+0.044	+0.000	0.015	7,784
	(0.255)	(0.984)		
Age at first birth	+0.559	+0.030	0.025	5,237
	(0.205)	(0.607)		
Household wage income, 1940	−41.5	−4.3	0.028	7,059
	(0.583)	(0.806)		
In the labor force with spouse present, 1930/40	+0.050	+0.000	0.004	10,234
	(0.067)	(0.655)		
In the labor force with no spouse, 1930/40	+0.044	−0.001	0.025	10,234
	(0.191)	(0.881)		
14+ years of schooling, 1940	+0.073	+0.002	0.021	8,228
	(0.045)	(0.647)		

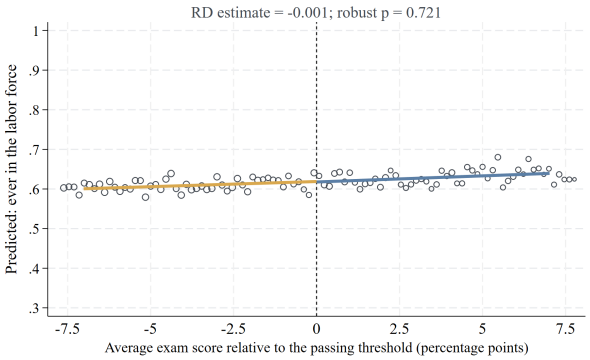
Notes: Column (1) is the main RD estimate. Column (2) is the RD estimate on the outcome predicted from preexam characteristics and exam-year fixed effects (ordinary least squares, OLS, on the outcome's own sample; missing characteristics set to zero with an indicator). Column (3) is the R^2 of the prediction. Robust bias-corrected p-values in parentheses. Local-linear, triangular kernel, MSE-optimal bandwidth.



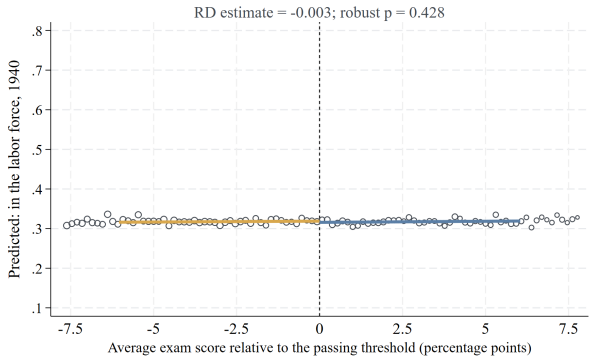
(a) Ever a teacher



(b) Teacher, 1940



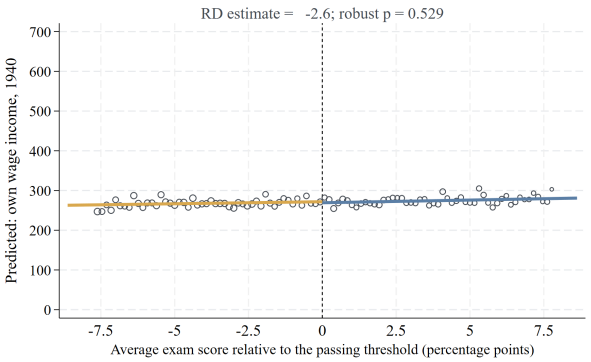
(c) Ever in the labor force



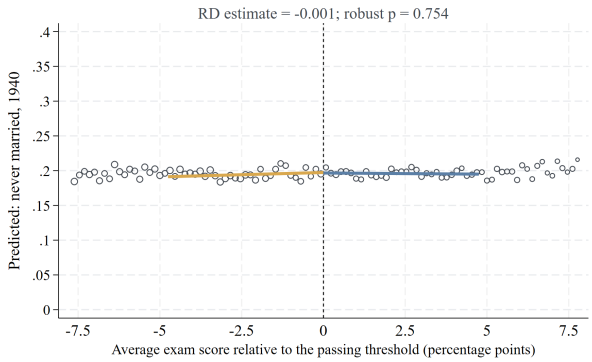
(d) In the labor force, 1940

Figure B17: Predicted Labor Outcomes

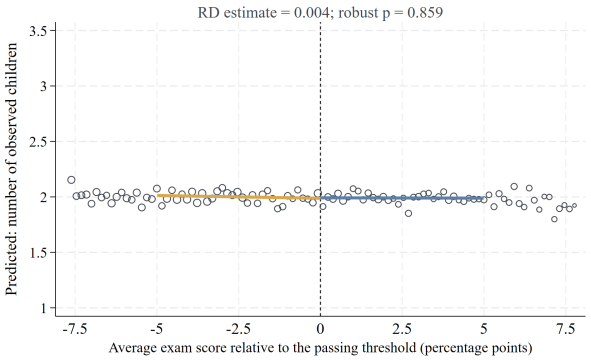
Notes: Each panel plots the predicted outcome (Table A10) against the running variable, on the same vertical scale as the main figure. Markers are bin means; lines are local-linear fits on each side of the threshold.



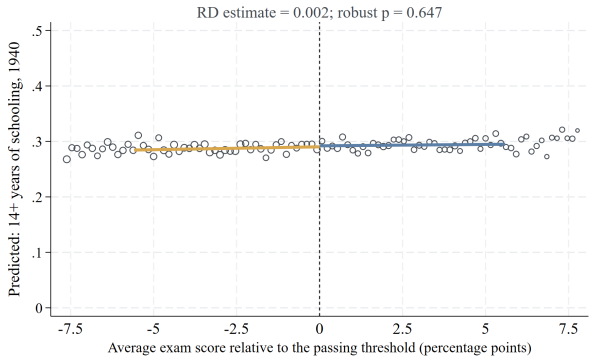
(a) Own wage income, 1940



(b) Never married, 1940



(c) Number of observed children



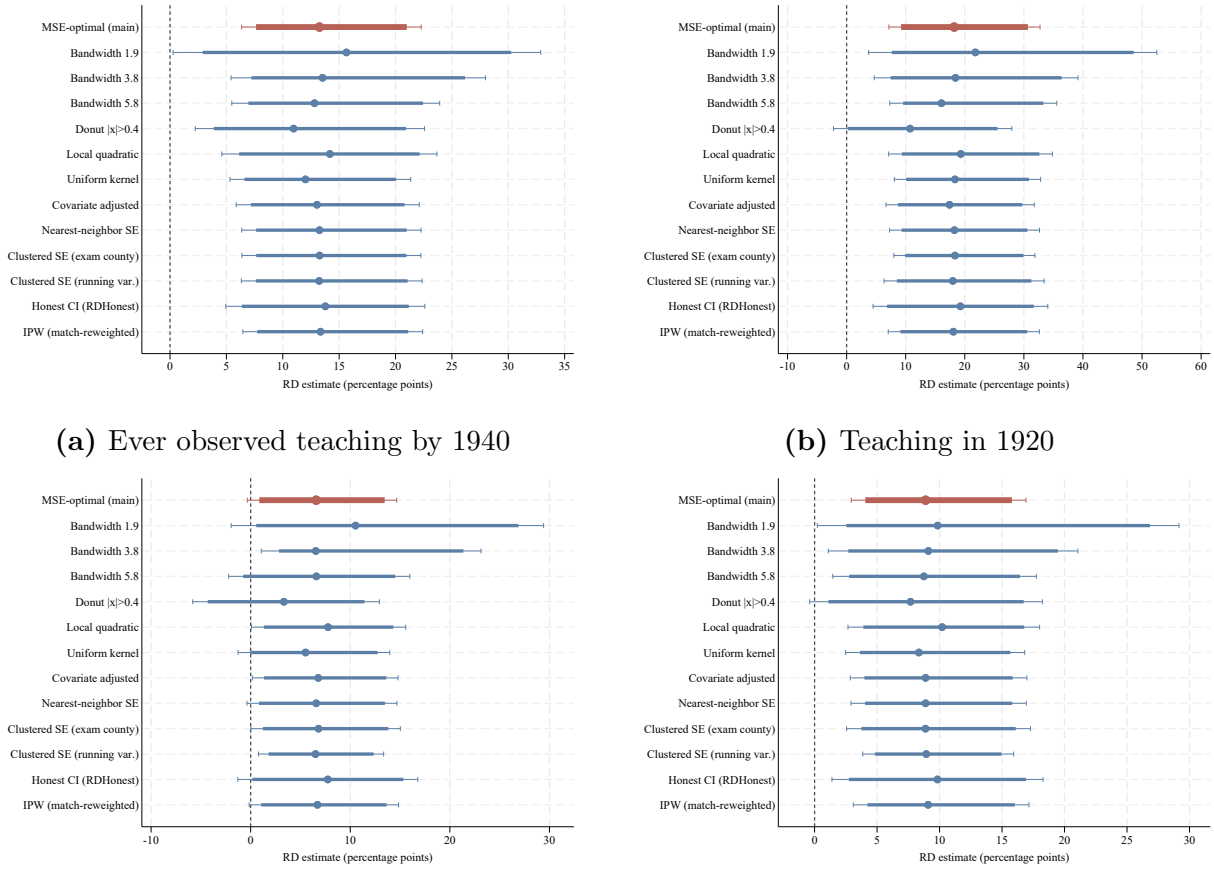
(d) 14+ years of schooling, 1940

Figure B18: Predicted Income, Marriage, Fertility, and Education Outcomes

Notes: See notes to Figure B17.

D Robustness Across Specifications

Figures B19–B24 compare the main estimate with twelve alternative specifications, including standard errors clustered by exam county and by the value of the running variable, honest confidence intervals for a discrete running variable (Kolesár and Rothe, 2018; Armstrong and Kolesár, 2020), and an estimate that reweights the linked sample to its baseline composition.



(a) Ever observed teaching by 1940

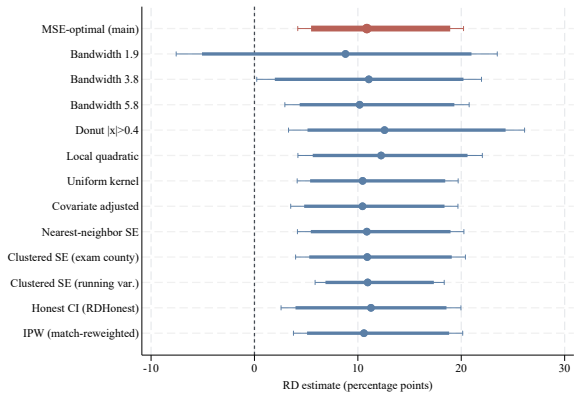
(b) Teaching in 1920

(c) Teaching in 1930

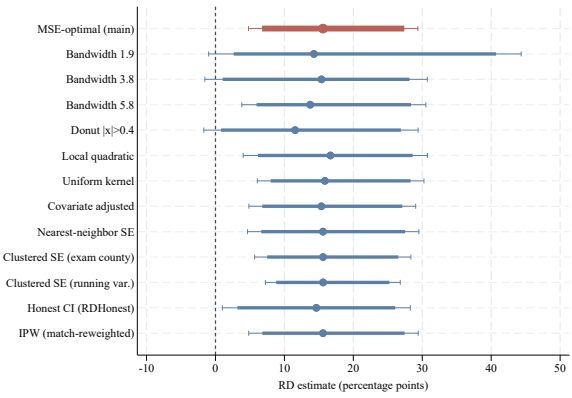
(d) Teaching in 1940

Figure B19: Robustness: Teaching

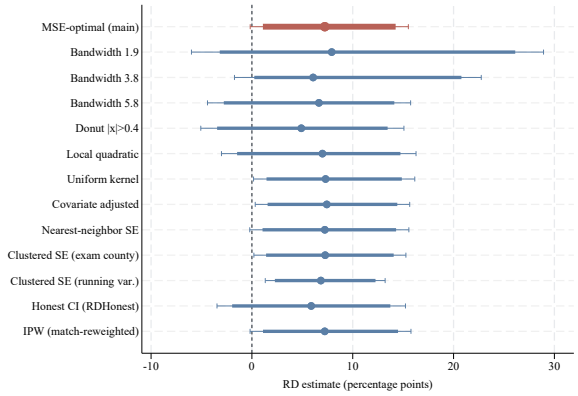
Notes: The red marker is the main estimate, using the MSE-optimal bandwidth. Blue markers denote the specifications for the twelve alternative specifications. These are bandwidths of 1.9, 3.8, and 5.8 percentage points; a donut that drops observations within 0.4 percentage points of the cutoff; a local-quadratic fit; a uniform kernel; controls for age at exam, education at exam, and birth year; nearest-neighbor standard errors; standard errors clustered by exam county; standard errors clustered by the value of the running variable; honest confidence intervals (RDHonest, triangular kernel, MSE-optimal bandwidth); and an estimate that reweights the linked sample to match characteristics of the sample who match to a baseline census. Thick bars are 90 percent and thin capped lines 95 percent confidence intervals, bias-corrected for the local-polynomial specifications and honest for the RDHonest estimate.



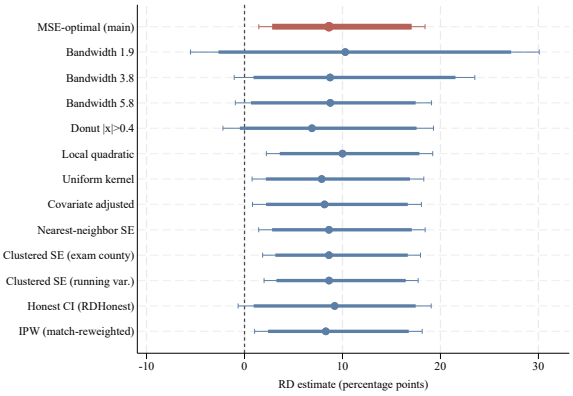
(a) Ever observed in the labor force by 1940



(b) Labor force participation in 1920



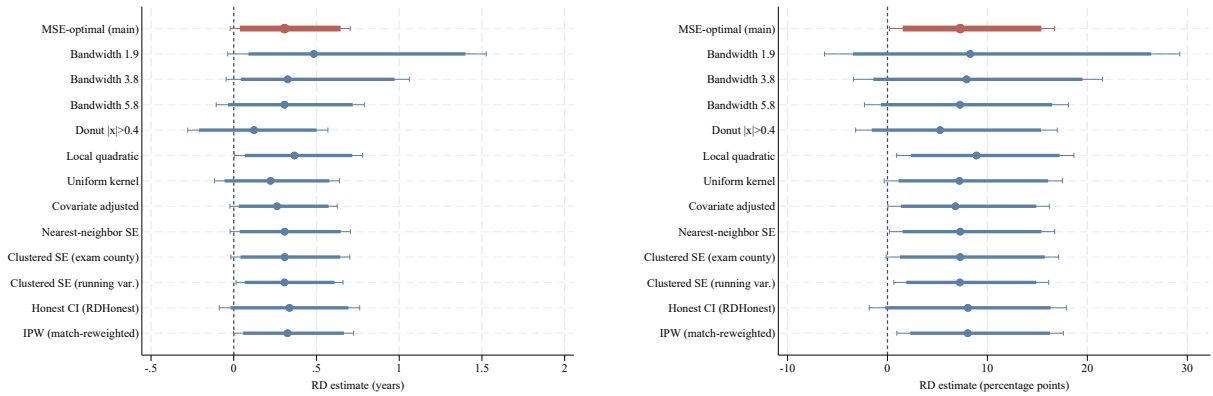
(c) Labor force participation in 1930



(d) Labor force participation in 1940

Figure B20: Robustness: Labor Force Participation

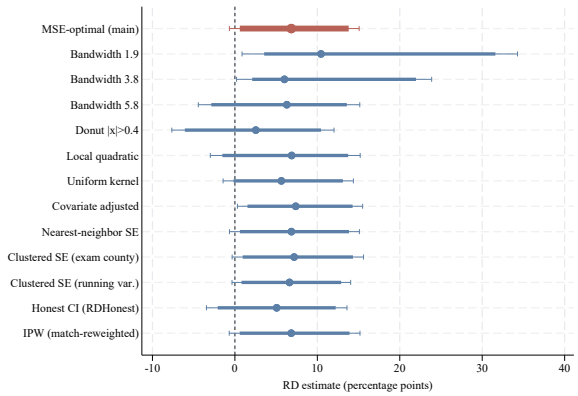
Notes: The red marker is the main estimate, using the MSE-optimal bandwidth. Blue markers denote the specifications for the twelve alternative specifications. These are bandwidths of 1.9, 3.8, and 5.8 percentage points; a donut that drops observations within 0.4 percentage points of the cutoff; a local-quadratic fit; a uniform kernel; controls for age at exam, education at exam, and birth year; nearest-neighbor standard errors; standard errors clustered by exam county; standard errors clustered by the value of the running variable; honest confidence intervals (RDHonest, triangular kernel, MSE-optimal bandwidth); and an estimate that reweights the linked sample to match characteristics of the sample who match to a baseline census. Thick bars are 90 percent and thin capped lines 95 percent confidence intervals, bias-corrected for the local-polynomial specifications and honest for the RDHonest estimate.



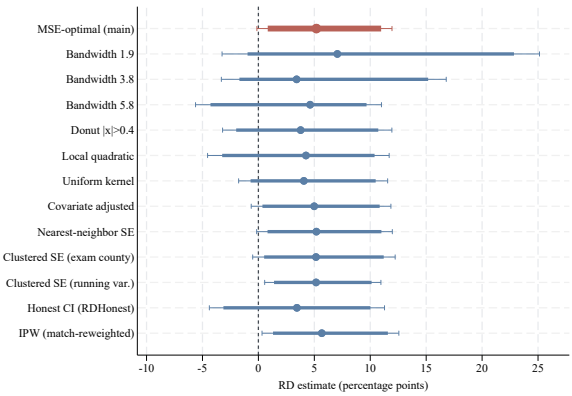
(a) Years of education in 1940 **(b)** 14+ years of schooling, 1940

Figure B21: Robustness: Education

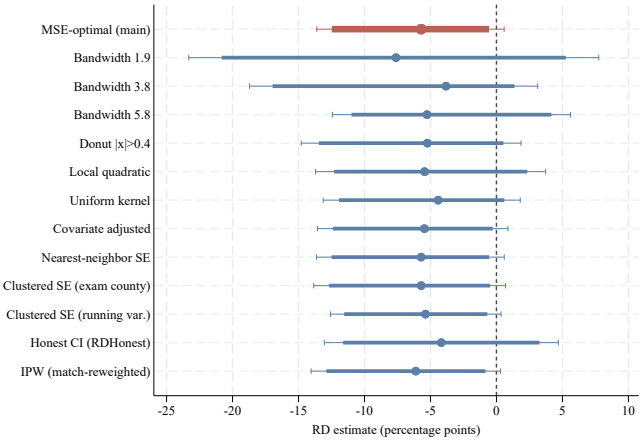
Notes: The red marker is the main estimate, using the MSE-optimal bandwidth. Blue markers denote the specifications for the twelve alternative specifications. These are bandwidths of 1.9, 3.8, and 5.8 percentage points; a donut that drops observations within 0.4 percentage points of the cutoff; a local-quadratic fit; a uniform kernel; controls for age at exam, education at exam, and birth year; nearest-neighbor standard errors; standard errors clustered by exam county; standard errors clustered by the value of the running variable; honest confidence intervals (RDHonest, triangular kernel, MSE-optimal bandwidth); and an estimate that reweights the linked sample to match characteristics of the sample who match to a baseline census. Thick bars are 90 percent and thin capped lines 95 percent confidence intervals, bias-corrected for the local-polynomial specifications and honest for the RDHonest estimate.



(a) Never married in 1930



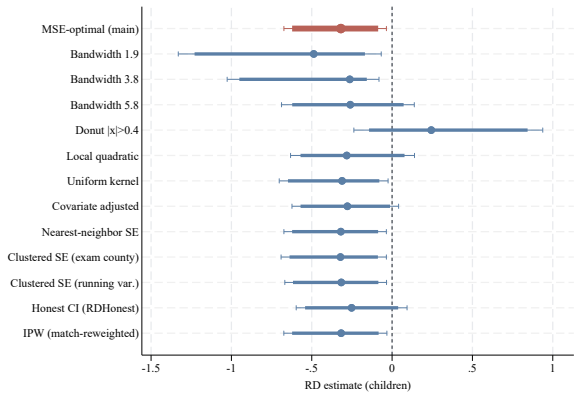
(b) Never married in 1940



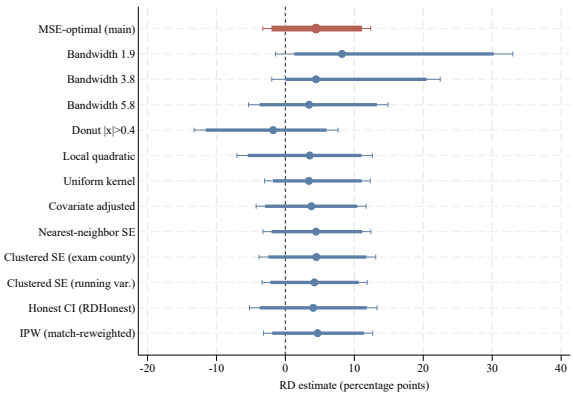
(c) Married, spouse present, 1940

Figure B22: Robustness: Marriage

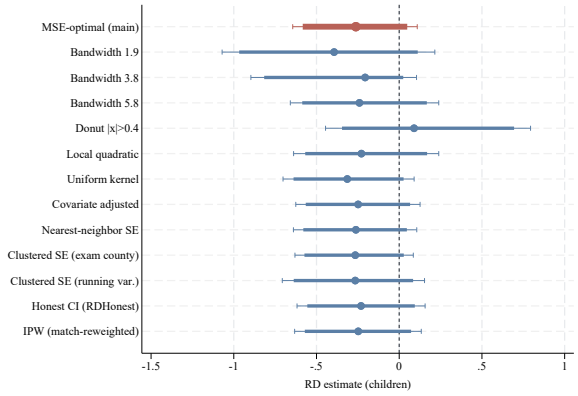
Notes: The red marker is the main estimate, using the MSE-optimal bandwidth. Blue markers denote the specifications for the twelve alternative specifications. These are bandwidths of 1.9, 3.8, and 5.8 percentage points; a donut that drops observations within 0.4 percentage points of the cutoff; a local-quadratic fit; a uniform kernel; controls for age at exam, education at exam, and birth year; nearest-neighbor standard errors; standard errors clustered by exam county; standard errors clustered by the value of the running variable; honest confidence intervals (RDHonest, triangular kernel, MSE-optimal bandwidth); and an estimate that reweights the linked sample to match characteristics of the sample who match to a baseline census. Thick bars are 90 percent and thin capped lines 95 percent confidence intervals, bias-corrected for the local-polynomial specifications and honest for the RDHonest estimate.



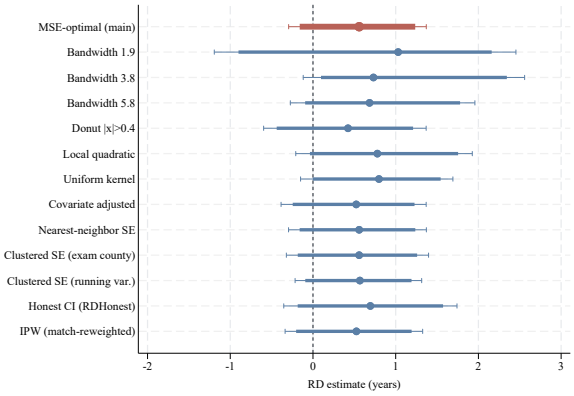
(a) Estimated fertility



(b) No children by 1940



(c) Fertility conditional on positive



(d) Age at first birth

Figure B23: Robustness: Fertility

Notes: The red marker is the main estimate, using the MSE-optimal bandwidth. Blue markers denote the specifications for the twelve alternative specifications. These are bandwidths of 1.9, 3.8, and 5.8 percentage points; a donut that drops observations within 0.4 percentage points of the cutoff; a local-quadratic fit; a uniform kernel; controls for age at exam, education at exam, and birth year; nearest-neighbor standard errors; standard errors clustered by exam county; standard errors clustered by the value of the running variable; honest confidence intervals (RDHonest, triangular kernel, MSE-optimal bandwidth); and an estimate that reweights the linked sample to match characteristics of the sample who match to a baseline census. Thick bars are 90 percent and thin capped lines 95 percent confidence intervals, bias-corrected for the local-polynomial specifications and honest for the RDHonest estimate.

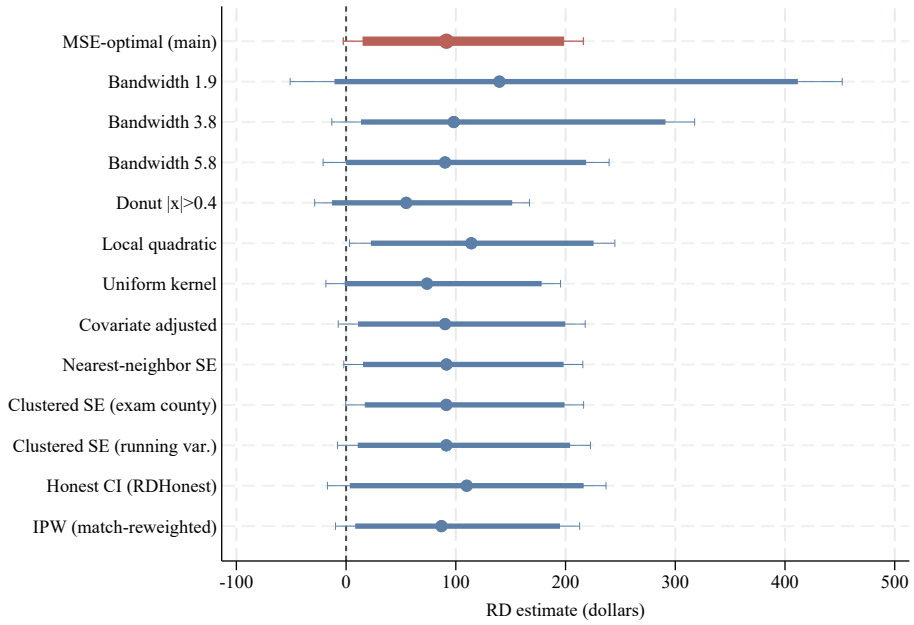


Figure B24: Robustness: Own Wage Income in 1940

Notes: The red marker is the main estimate, using the MSE-optimal bandwidth. Blue markers denote the specifications for the twelve alternative specifications. These are bandwidths of 1.9, 3.8, and 5.8 percentage points; a donut that drops observations within 0.4 percentage points of the cutoff; a local-quadratic fit; a uniform kernel; controls for age at exam, education at exam, and birth year; nearest-neighbor standard errors; standard errors clustered by exam county; standard errors clustered by the value of the running variable; honest confidence intervals (RDHonest, triangular kernel, MSE-optimal bandwidth); and an estimate that reweights the linked sample to match characteristics of the sample who match to a baseline census. Thick bars are 90 percent and thin capped lines 95 percent confidence intervals, bias-corrected for the local-polynomial specifications and honest for the RDHonest estimate.

E Placebo Cutoff Distributions

I reestimate the RD across a range of false cutoffs and plot the true estimate against the distribution of placebo estimates. For each outcome I place a false cutoff at every raw point of the cumulative score from 150 points below to 75 points above the true passing threshold (11.5 percentage points of the average below to 5.8 above), excluding a narrow window within 5 raw points of the true cutoff, and estimate the RD. Each placebo estimate uses observations on the same side of the true cutoff as the placebo cutoff to avoid including the true discontinuity in the estimation (Imbens and Lemieux, 2008). This produces 215 placebo estimates per outcome.

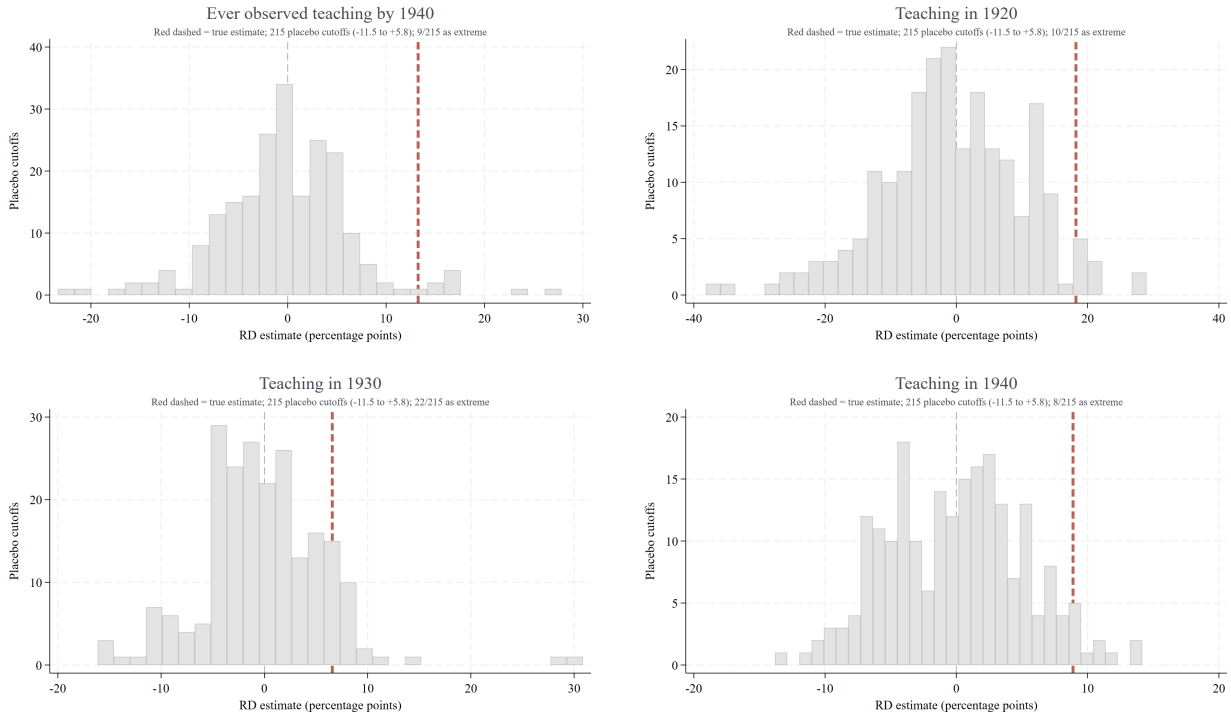


Figure B25: Placebo-Cutoff Distributions: Teaching

Notes: Gray bars are the distribution of RD point estimates at 215 false cutoffs, placed at every raw point of the cumulative score from 150 points below to 75 points above the true threshold (excluding a window within 5 raw points of it). The red dashed line is the true estimate at the certification threshold. Each RD estimate using the placebo cutoff uses observations on the same side of the threshold to which it was shifted. This is to exclude the true discontinuity from the placebo estimates. Estimates use the MSE-optimal bandwidth (HC3 standard errors). The count in each panel is the number of placebo estimates at least as large as the true estimate in the same direction. Binary outcomes are in percentage points and continuous outcomes in their natural units.

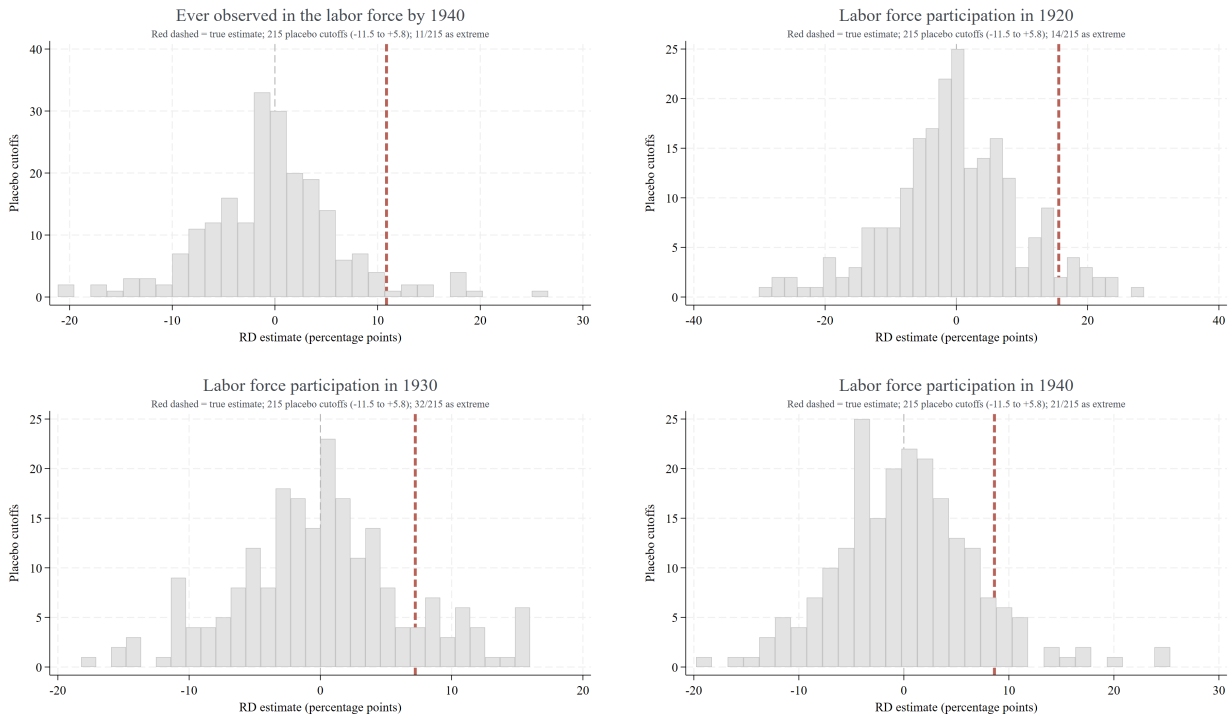


Figure B26: Placebo-Cutoff Distributions: Labor Force Participation

Notes: Gray bars are the distribution of RD point estimates at 215 false cutoffs, placed at every raw point of the cumulative score from 150 points below to 75 points above the true threshold (excluding a window within 5 raw points of it). The red dashed line is the true estimate at the certification threshold. Each RD estimate using the placebo cutoff uses observations on the same side of the threshold to which it was shifted. This is to exclude the true discontinuity from the placebo estimates. Estimates use the MSE-optimal bandwidth (HC3 standard errors). The count in each panel is the number of placebo estimates at least as large as the true estimate in the same direction. Binary outcomes are in percentage points and continuous outcomes in their natural units.

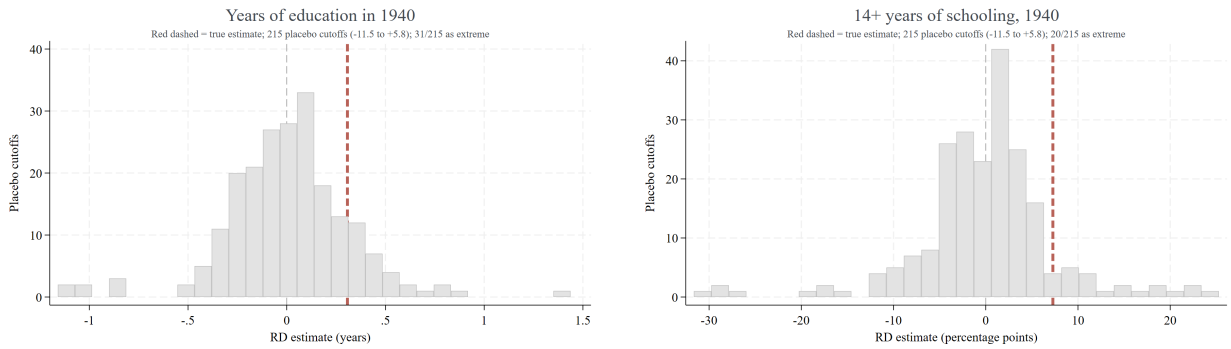


Figure B27: Placebo-Cutoff Distributions: Education

Notes: Gray bars are the distribution of RD point estimates at 215 false cutoffs, placed at every raw point of the cumulative score from 150 points below to 75 points above the true threshold (excluding a window within 5 raw points of it). The red dashed line is the true estimate at the certification threshold. Each RD estimate using the placebo cutoff uses observations on the same side of the threshold to which it was shifted. This is to exclude the true discontinuity from the placebo estimates. Estimates use the MSE-optimal bandwidth (HC3 standard errors). The count in each panel is the number of placebo estimates at least as large as the true estimate in the same direction. Binary outcomes are in percentage points and continuous outcomes in their natural units.

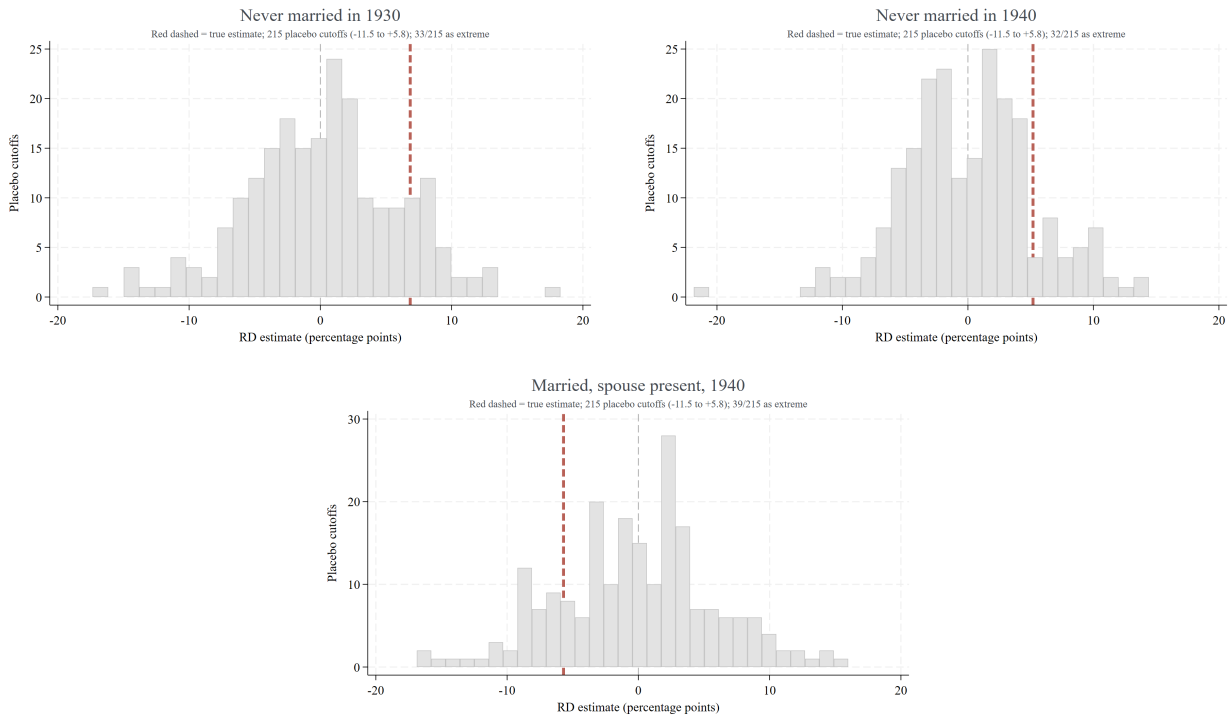


Figure B28: Placebo-Cutoff Distributions: Marriage

Notes: Gray bars are the distribution of RD point estimates at 215 false cutoffs, placed at every raw point of the cumulative score from 150 points below to 75 points above the true threshold (excluding a window within 5 raw points of it). The red dashed line is the true estimate at the certification threshold. Each RD estimate using the placebo cutoff uses observations on the same side of the threshold to which it was shifted. This is to exclude the true discontinuity from the placebo estimates. Estimates use the MSE-optimal bandwidth (HC3 standard errors). The count in each panel is the number of placebo estimates at least as large as the true estimate in the same direction. Binary outcomes are in percentage points and continuous outcomes in their natural units.

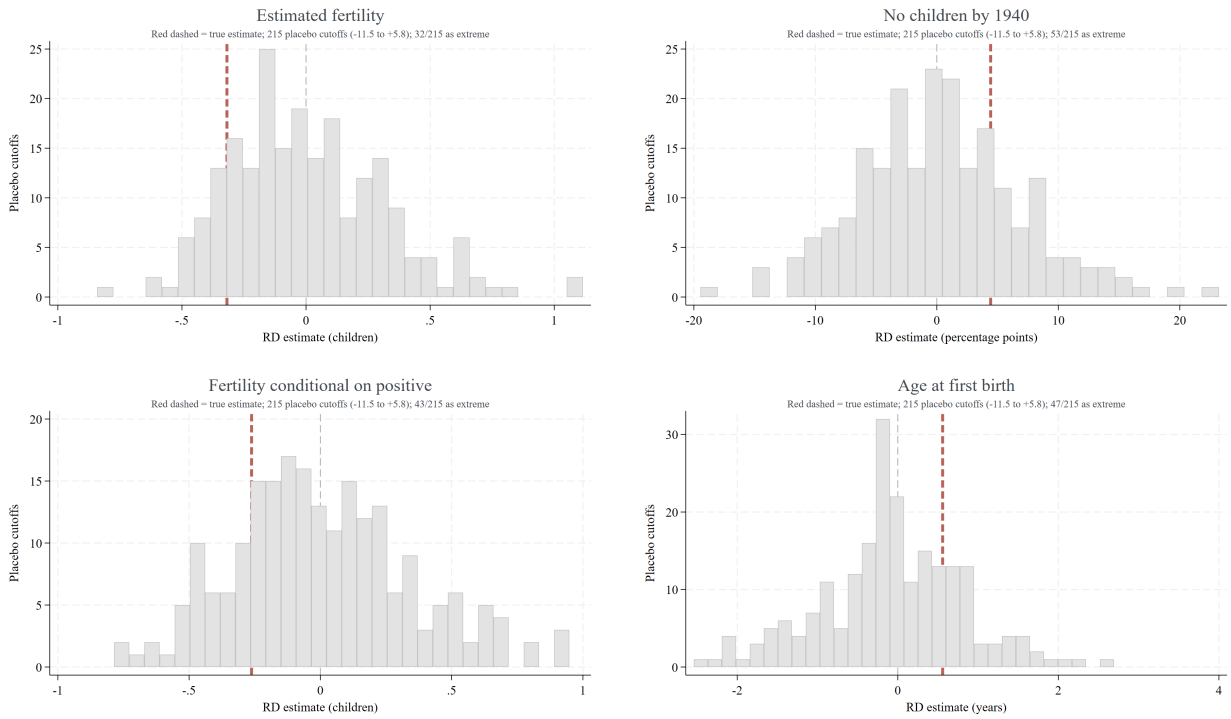


Figure B29: Placebo-Cutoff Distributions: Fertility

Notes: Gray bars are the distribution of RD point estimates at 215 false cutoffs, placed at every raw point of the cumulative score from 150 points below to 75 points above the true threshold (excluding a window within 5 raw points of it). The red dashed line is the true estimate at the certification threshold. Each RD estimate using the placebo cutoff uses observations on the same side of the threshold to which it was shifted. This is to exclude the true discontinuity from the placebo estimates. Estimates use the MSE-optimal bandwidth (HC3 standard errors). The count in each panel is the number of placebo estimates at least as large as the true estimate in the same direction. Binary outcomes are in percentage points and continuous outcomes in their natural units.

Figure B30: Placebo-Cutoff Distributions: Own Wage Income in 1940

Notes: Gray bars are the distribution of RD point estimates at 215 false cutoffs, placed at every raw point of the cumulative score from 150 points below to 75 points above the true threshold (excluding a window within 5 raw points of it). The red dashed line is the true estimate at the certification threshold. Each RD estimate using the placebo cutoff uses observations on the same side of the threshold to which it was shifted. This is to exclude the true discontinuity from the placebo estimates. Estimates use the MSE-optimal bandwidth (HC3 standard errors). The count in each panel is the number of placebo estimates at least as large as the true estimate in the same direction. Binary outcomes are in percentage points and continuous outcomes in their natural units.