The Democratization of Opportunity: The Effects of the U.S. High School Movement

Alison Doxey[*], Ezra Karger[†], and Peter Nencka[‡]

July 12, 2022

Abstract

The construction of public high schools across the United States in the late 1800s and early 1900s transformed the economic opportunities of residents, particularly women, who outnumbered men as high school graduates until the 1940s. We estimate the effect of the "high school movement" on short- and long-run outcomes for both men and women using a novel, complete panel of high schools in towns and cities across the United States. This panel generates a place-level measure of high school access in the late 1800s and early 1900s. We link this panel to school enrollment information from complete-count censuses and the exact birthplaces of notable men and women using structured biographies of federal judges, congresspeople, scientists, businesspeople, and artists. We compare school attendance rates and the long-run outcomes of children who did and did not have access to a public high school, exploiting variation across otherwise similar towns and cities that expanded high schools at different times.

1 Introduction

The rise of public high schools in the United States in the early 20th century led to one of the fastest and most comprehensive increases in human capital in modern history. In 1910, only 9 percent of 18-year-olds held high school degrees. By 1940, 50 percent did (Goldin and Katz, 2010). This rapid increase was fueled by a surge in the number of local high schools, from fewer than 3,000 in the late 1800s to more than 20,000 by the early 1950s.[1] Accessible, publicly-funded, and free high

[*]University of Chicago

[†]Federal Reserve Bank of Chicago

[‡]Department of Economics, Miami University

[1]The population of the United States also grew, but at a much lower rate. The number of school-age children in the United States increased by only 65% from 1890 to 1950.

schools catapulted the United States education system ahead of comparable European countries, creating new opportunities for residents. Increased schooling access particularly benefited women, who graduated from high school at higher rates than men until the 1940s (Goldin, 2008; 2021).

Despite widespread recognition that what has now been dubbed "the high school movement" dramatically changed U.S society, it is difficult to study its impact on individuals. Data is limited; to date, researchers have analyzed the expansion of public high schools across the U.S. using state-level summary reports of high schools or county-level statistics for select states (Goldin, 1998, 1999). However, access to schooling varied dramatically *within* states and counties over time, so it is difficult to map this aggregate data to the exact opportunities available to local residents. The potential effects of schooling are also challenging to measure. The 1940 census was the first national census to ask respondents to self-report educational attainment, and this outcome is measured with significant error since respondents greatly overstated their own level of schooling (Goldin, 1998). Tracking outcomes for women is particularly difficult. Because of name changes after marriage, women cannot easily be linked to their childhood records to understand their educational opportunities.

In this paper, we provide the first estimates of the effect of high school access on the long-run outcomes of both men and women. We construct a novel, complete panel of all public and private high schools in the United States through 1951. The panel contains information about each high school's founding date and location, as well as counts of enrollment and teachers by school, when available. We link this panel to a dataset of notable men and women born between 1870 and 1950, drawn from Federal Judiciary biographies, Congressional biographies, and crowd-sourced structured biographical sketches from Wikidata. Importantly, these biographical sources include exact dates and birthplaces for both men and women. In this paper summary, we focus on the panel of people in Wikidata.

We estimate the effect of high school access on enrollment and adult outcomes, exploiting variation within otherwise similar towns and cities that expanded high school access at different times. Our approach builds on a large literature that links the construction of schools and school spending to short- and long-run local outcomes (e.g., Duflo, 2001; Duflo, 2004; Aaronson and Mazumder, 2011; Schmick and Shertzer, 2020; Navarro-Sola, 2021; Andrews, 2021). We find that the introduction of high schools led to a sharp increase in the rate of children attending school, and that this increase grew over time. We see little-to-no increase in enrollment for children aged 13 and 14, consistent with the idea that many of these students would have been attending middle school or part-time schooling and would therefore have been relatively unaffected by high school entry. For 14–18-year-olds, we see positive enrollment effects that start in the decades immediately after high schools open. These enrollment effects are large: For example, after 2 decades we see an increase in enrollment of about 20 percent for 17-year-olds.

Linking high schools to our panel of notable people, we estimate whether there is a correlation between the timing of the birth of children in a city who would grow up to be eminent and the entry of local high schools. Our outcomes include various indicators for notability: including whether a child grew up to be a federal judge, congressperson, eminent scientist, recognized businessperson, or artist.[2] These are important right-tail human capital outcomes, and we can measure them consistently throughout our entire sample period for both men and women. By focusing on these outcomes, we contribute to a literature examining the predictors of right-tail success, an outcome that plays a key role in driving innovation, economic growth, and inequality (e.g., Bell et al. 2019, Hsieh et al., 2019; Aghion et al., 2019).

We find suggestive evidence that after high schools were built, cities produced more eminent people, and that this increase was largest on a percentage basis for women. However, while we do not observe any pre-trends in our data for these outcomes in the years leading up to high school entry, we caution that post-entry differences are likely a combination of the direct causal effects of high schools, the impact of high schools on local trends (like population growth), and contem-

[2]In future work, we also plan to examine other measures of eminence, including specific awards won (eg. Nobel prizes, National Book Awards, Pulitzer Prizes).

poraneous changes in cities, such as the development of other institutions that could be correlated with long-run human capital outcomes. Our results provide the most geographically detailed information to date on the spread and correlates of the high school movement, and in future work, we anticipate teasing apart these correlates in more detail from the causal effects of high schools on local communities.

2 Data

2.1 High Schools

We construct our panel of high schools in the U.S. from the 1800s to the mid-1900s by combining information from four sources. First, we compile information from censuses of all public and private high schools, collected by the Bureau of Education every 1–2 years from 1873 to 1905 and in 1912. After 1912, the next census of high schools was published in 1951. These censuses contain information about the number of students, the number of teachers, the length of study, and sometimes the founding year of each high school.[3] We use founding years to retrospectively construct a consistent panel of high schools across the United States, even for years when high schools were not directly surveyed.[4]

For later years, we use lists of accredited high schools published by the Bureau of Education every 2–6 years from 1911 through 1944. Accreditation standards varied by state, and public universities used published bulletins to offer admission to local students with a diploma from a high school that met a set of criteria. These criteria often included a requirement (1) that the high school offers four years of study, and (2) that the high school offer at least a minimum number of math, English, and history credits. These lists of accredited high schools contain no information

[3]The founding year was only collected in a few of the censuses, including in 1903.

[4]A limitation of this approach is that we will not observe high schools that closed using retrospective data. In our empirical analysis, we do not focus on individual high schools, but on the existence of any school within a city. It is extremely rare that a city loses access to a high school once one is built, so we do not believe this is a limitation for our analytical purposes.

on founding dates or enrollment, but they do confirm the existence of public and private secondary schools in particular cities and towns.

To complement the founding dates from the earlier high school censuses, we use *Patterson's American Educational Directories*. These provide names of high schools in each county for seven years between 1906 and 1924. The Patterson's directories contain establishment dates for reporting high schools, but these directories were less comprehensive than the contemporaneous censuses of high schools produced by the Bureau of Education [5]

To supplement our data on private high schools, we collect information from the more recent Private School Universe Surveys produced by the National Center for Education Statistics. These surveys include the founding dates of reporting schools. We augment our historical panel with founding dates of high schools reported in the 1989 or 1995 academic year surveys.

We digitize each of these sources, with the exception of the more recent Private School Survey, which is already machine-readable. We combine these four data sources into one panel of high schools containing 360,000 school-by- year observations. The median observation in our panel comes from 1922, but the key data for our analysis comes from the 1889 through 1905 high school censuses, which are available annually.

To prepare this panel for analysis we:

1. Identify transcription errors in the names of towns and cities and correct those errors by looking for places that exist uniquely in the panel, are one letter away from other place names that appear in the panel, and do not conflict on available years.

2. Collapse data to place-year level, calculating enrollment of local schools and the earliest reported founding year of public and private schools.

3. Link each high school's town or city name to GNIS place data, Geocodio, Texas A&M's geocoding platform, and the HERE Geocoder API to identify a longitude-latitude corre-

[5] We use the Patterson's directories for the 1906, 1908, 1912, 1913, 1914, 1920, and 1924 academic years.

sponding to the location of each school's town or city. This step involves standardizing place names and relying on county-level and state-level location identifiers in the high school data when it is available. We use multiple geocoding platforms so that we can look for agreement and disagreement across sources.[6]

4. We take all of the potential geocodes for each place and address discrepancies by privileging geocodes that appear to be within five miles of each other independently in multiple sources. We also require that the implied geocode is within the United States.

At the end of this process, we have a panel of 26,877 places with information about the longitude and latitude of the place where we observe that high school. In Figure 1 we map the spread of high schools over time. Towns and cities with high schools by the 1800s were concentrated in New England and parts of the Midwest. High schools expanded later in the South, and towns and cities in the West also lagged New England in schooling investment.

In Figure 2, we show that despite these general patterns (which have been documented in past work), aggregate state-level high school access measures hide substantial heterogeneity within states. For example, comparing Panels A and D, we see that the average city in Massachusetts received a high school much earlier than the average city in Tennessee. But *within* each state, there is substantial variation in the timing of high school access. This within-state and county variation is particularly evident in the Midwest, as shown in Panels B and C for Ohio and Wisconsin.

In Figure 3 we plot the exponential increase in the number of towns and cities in the United States with at least one high school between 1800 and 1950. We show the cumulative number of cities with a public high school and 'any' high school, which includes private high schools.[7] While

[6]In this step, we can run into difficulties with non-unique matches if there are two cities or towns with the same name in a given county or state. We flag these observations for exclusion in future robustness checks.

[7]Note that we plot the earliest year when a high school could have existed in our town according to our panel. So for example, because we have lower-quality data between 1905 and 1910, we assume that each town that reported a high school in 1910 had access to that high school as of 1906–the first year when our panel implies that a high school could have existed in those towns. This creates minor discontinuities in the growth rate in Figure 3. Our results are robust to excluding high schools assigned to these discontinuous years.

some towns and cities built high schools in the early- to mid-1800s, we see the largest increase in the number of high schools between 1880 and 1910, when the number of towns with a high school more than tripled. After 1910, we no longer have access to the higher-quality high school census data. The flat growth rate of the number of towns with a high school between 1910 and 1915 is driven by a lack of data, and the subsequent upswing in the number of towns with a high school in our panel is caused by the large number of accredited high schools reporting to the Bureau of Education for the first time in the 1920s and 1930s.

In Figure 4 we plot the total enrollment of students in high schools in years of our data with valid enrollment information. Consistent with prior work by Goldin (1999), who reports Bureau of Education aggregate graduation and enrollment information from 1890 to 1970, we find a relatively linear increase in enrollment rates from 1890 to 1910.[8] In years where we have enrollment information aggregated from the high school panel and Goldin (1999) reports aggregate enrollment rates, our reported enrollment rates are quite similar.[9]

2.2 Census

To track school enrollment rates over time, we use the historical decennial U.S. censuses. Each census from 1850 through 1940 asked about school attendance, though the format and reporting patterns of those questions changed over time. We append the 1880-1940 censuses together and focus on enrollment for 13–18-year-olds. We create a local measure of enrollment using a new crosswalk of historical places that standardizes restricted access census location strings into consistently geocoded cities, towns, and unincorporated places (Berkes, Karger, and Nencka, 2022). These places are then linked to the geocoded high school panel, allowing us to measure when a

[8]Goldin (1999) finds a roughly linear increase in graduation rates from 1890 through 1900 and a linear increase in attendance rates from 1900 through 1910, when the Bureau of Education begins to report this information.

[9]For example, in 1910, Goldin (1999) shows a high school enrollment rate of 14–17-year-olds of around 15%, and there were roughly 7 million 14–17-year-olds alive in 1910, which implies an enrollment count of 1.05 million students. Aggregating our high school panel, we estimate that 1.06 million students were enrolled in high schools in 1910. Note that in Figure 4, we report this total in 1912, which was the year of the Bureau of Education Report, but which refers to the 1910 school year.

place had its first nearby high school. In our baseline analysis, we find the first high school within two miles of a given place.

2.3 Notable People

Our data describing notable people come from three sources:

1. Federal Judicial biographies produced by the Federal Judicial Center, an organization within the U.S. Federal Judiciary focused on education and research. This dataset contains structured biographical information about all 4,000 Article III district, appellate, and supreme court judges and justices from the 1700s to the present. Data include birthplaces, education, and pre-nomination occupations.

2. Congressional biographies from the Biographical Directory of the United States Congress–a structured database of information about the more than 10,000 congresspeople to hold office in the U.S. from the 1800s to the present. Data include birthplaces, educational attainment, pre-election or pre-appointment occupations, and election history.

3. Wikidata–a crowdsourced database of eminent people, including more than 100,000 U.S.-born eminent scientists, politicians, businesspeople, and artists. This data describes the exact birthplaces and lifetime achievements of eminent men and women across the U.S.

In this excerpt of our paper, we focus on the third source: Wikidata, a database maintained by the Wikimedia Foundation. Wikidata is an attempt to produce accurate structured knowledge about the world, and it is crowdsourced and maintained by a global network of volunteers. We rely in our analysis on a small number of characteristics (birthplace and sex) which are accurately recorded for eminent people.[10] We clean the data from Wikidata and create a birth-year-by-city panel that includes information on the number of eminent people by gender. We expand the panel

[10] We plan to validate this data by comparing a random sample of entries to official biographical information in the future.

and fill in zero-eminent-birth observations to create a balanced sample of birth years from 1810 forward. After geocoding the cities in Wikidata, we link this panel to the location of high schools and calculate high school entry years for each city in the panel.

In Figure 5, we group our Wikidata panel by decade of birth and separately plot the number of men and women in our sample, showing that the number of eminent men and women who we can track increased throughout the late 1800s and early 1900s. Because of their lower base rates in the early 1800s, women tracked in Wikidata grow at a faster percentage rate than men, though the absolute gap between men and women grows over time.

3 Preliminary results

In this section, we describe our preliminary results. As discussed above, these results are calculated using our new high school panel linked to two sources: high school enrollment from the historical censuses and counts of eminent people from Wikidata.

3.1 School enrollment results

To estimate the effects of high school entry on enrollment, we specify a standard event study model of the form:

$$Enrollment_{ct} = \sum_{e=-4}^{4} \beta_e 1(HS)_{cte} + \delta_c + \gamma_{s,t} + \varepsilon_{ct} \quad (1)$$

where $Enrollment_{ct}$ is the share of children in city c and decennial census year t who report being enrolled in school. The indicator variable $1(HS)_{cte}$ tracks the decades surrounding the first high school entry in place c, and δ_c and $\gamma_{s,t}$ are city and state-year fixed effects (respectively). The coefficients of interest are the vector β_e, which track enrollment effects both before and after first high school entry. Throughout, standard errors are clustered by city.

As is now well known in the literature, this type of event study framework can be biased if there are heterogeneous treatment effects across treated cohorts and time. Since event studies like Equation 1 use early-treated units as controls for later-treated units, treatment effect dynamics can introduce bias.[11] Both the possible existence and magnitude of this bias are difficult to predict *a priori*: it depends on the degree of the treatment effect heterogeneity and the number of never-treated units one has in the estimation sample. To address this concern, we also estimate a version of Equation 1 that follows the two-stage difference-in-difference imputation procedure outlined in Gardner (2021). In the first stage of this process, we identify city, decade, and state-year fixed effect coefficients using a sample of untreated and not-yet-treated observations. We then residualize our enrollment measure ($Enrollment_{ct}$), subtracting off the predicted enrollment from this first stage equation, and estimate Equation 1 using this residualized variable as our outcome in a simple event-study regression without covariates.[12]

We show the results of this estimation in Figure 6 using 14–18-year-old enrollment rates as our outcome variable. Panel A shows the results of the standard two-way fixed effects estimator; Panel B shows the results using the Gardner imputation estimator. In both cases, we see similar patterns: Before a high school enters, 14–18 year enrollment rates are relatively flat, particularly after adjusting for treatment effect heterogeneity using the Gardner approach. After the first high school is built, enrollment increases in the following decades. While precisely estimated, these effect sizes are modest: we see an increase of approximately 2.5 percentage points, or approximately 7 percent relative to the 1880 mean across these ages.[13]

This overall effect masks differences by age. We show this in Figure 7, where we plot the results from the Gardner estimation separately for each age from 13 to 18. We see no evidence for

[11] For a discussion of this bias and various remedies, see Baker, Larcker, and Wang (2022).

[12] Relative to other proposed methods to overcome issues in staggered difference-in-difference settings, the Gardner (2021) approach is extremely flexible, and is easily extendable to our event study setting. In future versions of the paper, we anticipate showing the results of the other common staggered timing estimators as appropriate.

[13] Because of space limitations, we omit robustness exercises for these results. We observe very similar results when we exclude state-year fixed effects or when we instead include county-by-year fixed effects.

an effect for those ages 13 or 14, and only limited evidence of an effect on the school attendance rates of 15-year-olds. This suggests that *overall* school enrollment was not changing at the same time that high schools entered — our effects are concentrated among high schoolers. Effect sizes grow rapidly both in absolute and percentage terms for high school-aged students. For example, after two decades we see an increase in enrollment of about 6 percentage points for 17-year-olds, or roughly 20 percent relative to 1880 enrollment rates. These effect sizes grow to over 10 percentage points by 40 years after high school entry.

The long-lasting and growing effects that we observe suggest that the immediate availability of a high school option did not increase high school enrollment dramatically overnight. This finding is consistent with past work on the high school movement, which showed the largest enrollment effects in later years. As Figure 3 shows, our median city or town had a high school around 1905. However, Goldin (1998) shows that the high school enrollment movement expanded most dramatically starting in 1910 before initially peaking around 1940. This is consistent with the timing of our effects and suggests that places that initially built high schools were also places that later invested more in the schools and increased enrollment relative to places that did not have a high school nearby.

3.2 Wikidata results

We next turn to the Wikidata results. The Wikidata panel is organized as a city-by-birth-year panel. After linking this panel to high schools, for each cohort we calculate the age when a high school first opened in their city. We use this age as our treatment variable and estimate versions of the following model:

$$WikiOutcome_{cb} = \sum_{a=-50}^{50} \beta_a 1(HS)_{cba} + \delta_c + \gamma_{b,s} + \varepsilon_{cb} \tag{2}$$

where $WikiOutcome_{cb}$ is an outcome from the Wikidata for people born in city c and year b. The indicator variable $1(HS)_{cba}$ tracks the ages surrounding the first high school entry in place c, and δ_c and $\gamma_{b,s}$ are place and state-birth-year fixed effects (respectively).

The coefficients of interest are the vector β_a, which track effects across age cohorts. For example, β_{30} is the effect of a high school opening in your birth town at age 30, while β_{12} is the effect of a high school opening in your birth town at age 12. Observing the path of coefficients β_{18} through β_{50} is analogous to looking at pre-trend coefficients in a standard event study. We would not expect there to be any differential effects of living in a city that built their first high school *after* one would have received a high school degree. The coefficients β_{15} through β_{17} represent partially treated cohorts. A nearby high school entering during those ages can still offer benefits, though students could not have a traditional, 4-year high school experience before age 18. Coefficients for ages younger than 14 reflect fully treated cohorts: they had a local high school available to them before traditional high schooling ages. As with the enrollment results, we include a significant "post-period" of children with full access to high schools to account for the possibility of effects that grow over time. For example, β_{-50} indicates children who were born 50 years *after* the first high school existed in their birth city. Throughout, standard errors are clustered by city.

Figure 8 shows the results of estimating Equation 2 on both the extensive and intensive margins of eminent people production. The omitted age is 18. Panel (a) shows the results of a Poisson maximum likelihood equation with the count of the number of eminent people as the outcome variable.[14] Panel (b) shows results from a standard ordinary least squares estimation with a zero-one indicator for whether a given cohort-city produced an eminent person tracked in Wikidata as the outcome variable. In both figures, we see generally flat "pre-trends" in outcomes for people who would have turned 18 or old before high schools opened in their cities. The first-dashed line from the right indicates partly treated children who would have turned 17 years old or younger when their high school opened, and the second dashed line indicates children who were "fully-

[14]There are a significant number of 0 observations in our data (places with no eminent people for a specific birth cohort), so our preferred count specification is the Poisson.

treated" as their local high school opened before age 13. In both panels, we see that relative to children who were aged 18 at the time that high schools opened, later cohorts were more likely to be recorded in Wikidata.

The implied effect sizes in percentage points are large changes from low pre-construction baseline levels. For example, our Poisson estimates can be interpreted as an approximate 50 percent increase in the production of eminent people for children who were fully treated, relative to a mean of 0.014 eminent people per birth cohort. The extensive margin effects over time are less stable and grow for later cohorts, reaching a peak of a 2 percentage point increase in the probability of observing an eminent person, relative to a sample mean of 1.4 percentage points.

In Figure 9 we show results separately for men and women. The pattern of results is similar for both men and women, though the percentage effects are larger for women. For example, the extensive margin results for women imply a 250 percent increase in the probability that a town would have an eminent person born in that town (180% for men), in part owing to the much lower base rate for women. This is also reflected in the Poisson results, though the coefficients for women are less precisely estimated than the male effects.

In Figures 10 and 11 we show a series of robustness checks. In particular, our results are robust to the inclusion of county-year fixed effects and to implementing the Gardner (2021) imputation estimator.

4 Limitations, conclusions, and future work

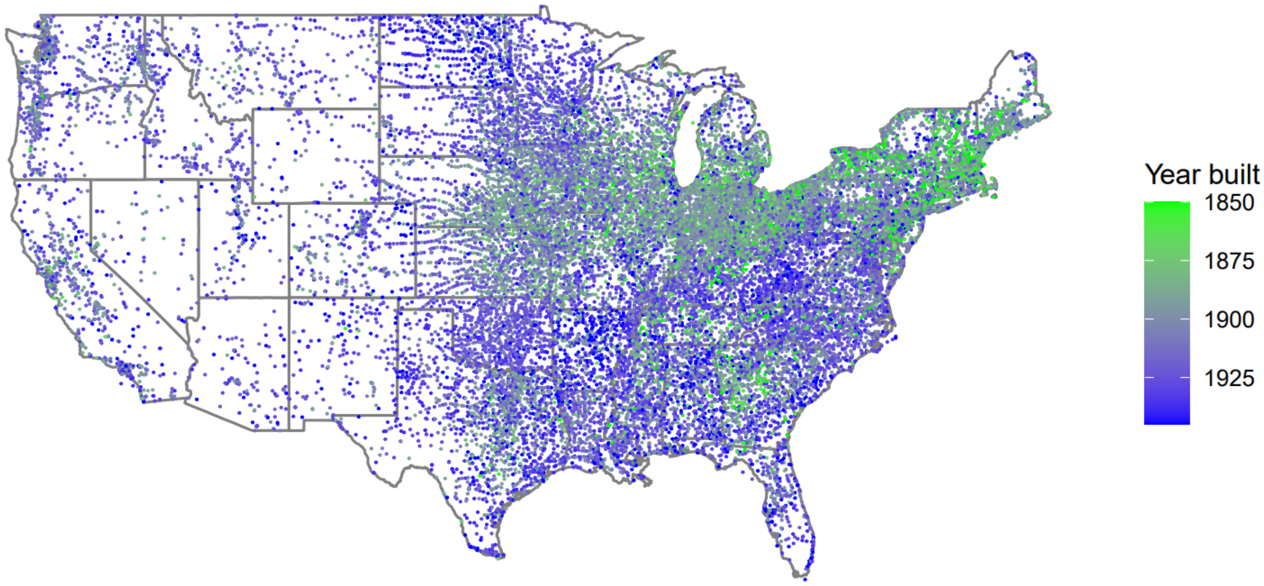
In this paper, we described our new panel of high schools and presented preliminary evidence on the relationship between the timing of high school entry, student enrollment, and the production of eminent men and women.

We anticipate refining and adjusting both our panel of high schools and eminent people in the future. In particular, a concern about Wikidata is that since it is community-reported, information in it could be inaccurate or could exhibit reporting patterns that are systematically correlated with our treatment. As discussed in prior sections, we are combining more "objective" sources of bibliography data with Wikidata to both validate the information in Wikidata and address some of these potential reporting limitations.

From an identification perspective, our event studies suggest that the timing of high school entry is correlated with an increase in both high school enrollment and the production of eminent men and women. Parallel pre-trends suggest that places that received a high school were not differentially changing relative to non-or-later treated places in the years leading up to high school entry. However, if other features of towns or cities are changing in the years after high schools enter, our results combine both the causal effect of high schools and the impact of other time-varying characteristics on our outcome variables. Of course, both types of effects are interesting: the latter tells us important information about the types of cities that received high schools even within narrowly-defined geographic regions, dramatically improving our understanding of the correlates and causes of the high school movement. In addition, some of the effects of high schools on local child outcomes may be indirect, but still causal. For example, this could occur if high schools affected migration decisions of lower-SES or higher-SES families.

To demonstrate these multiple dimensions of possible effects, in Figure 12 we plot results after conditioning on time-varying log city population for a subsample of our cities with valid population data. Panel A shows the Poisson count model and Panel B shows the linear probability model for the probability of observing an eminent person. Results are less precisely estimated than our main results (due to reduced sample sizes) and — particularly for the intensive margin results — show attenuated effects. This is consistent with the fact that places that received high schools grew faster than places that did not receive high schools. Future work will explore these issues more completely and construct alternative control groups: including matched treatment-control pairs and approaches that use only later-treated cities as controls for earlier treated cities. In addition, we plan to use city financial statistics and census data to document additional characteristics of

Figure 1: Map of cities, by year of first high school constructed

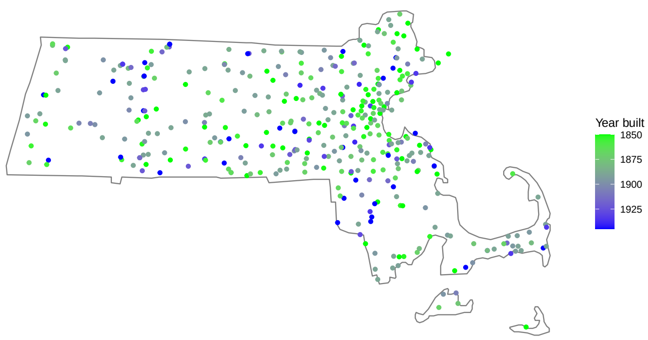


Notes: This map shows the founding dates of the first high school built in each of 26,877 places that we map to a valid longitude-latitude.

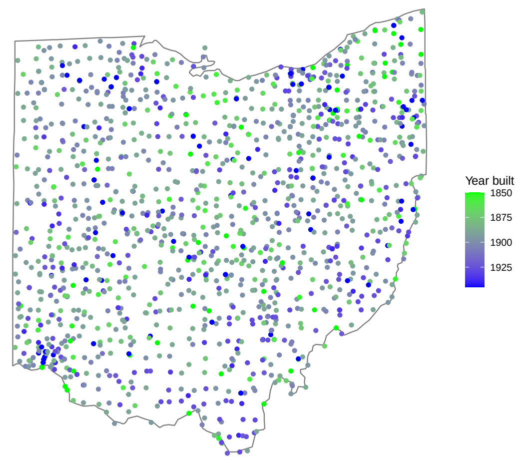
cities that predict high school entry.

Figure 2: Map of cities, by year of first high school constructed (selected states)

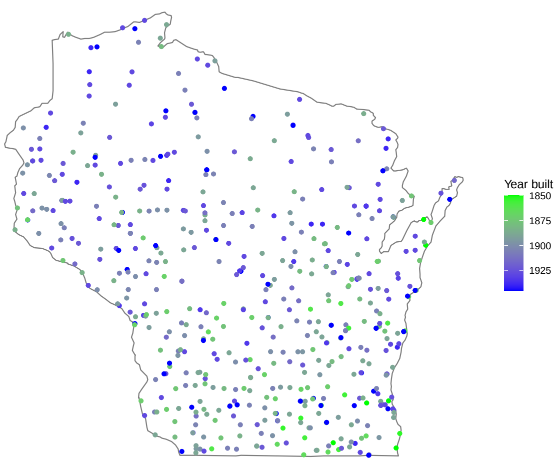
(a) Massachusetts



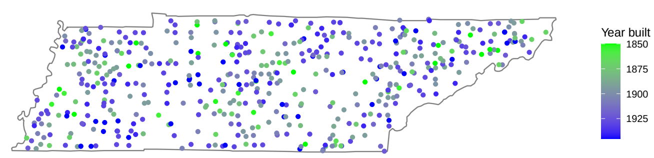
(b) Ohio



(c) Wisconsin

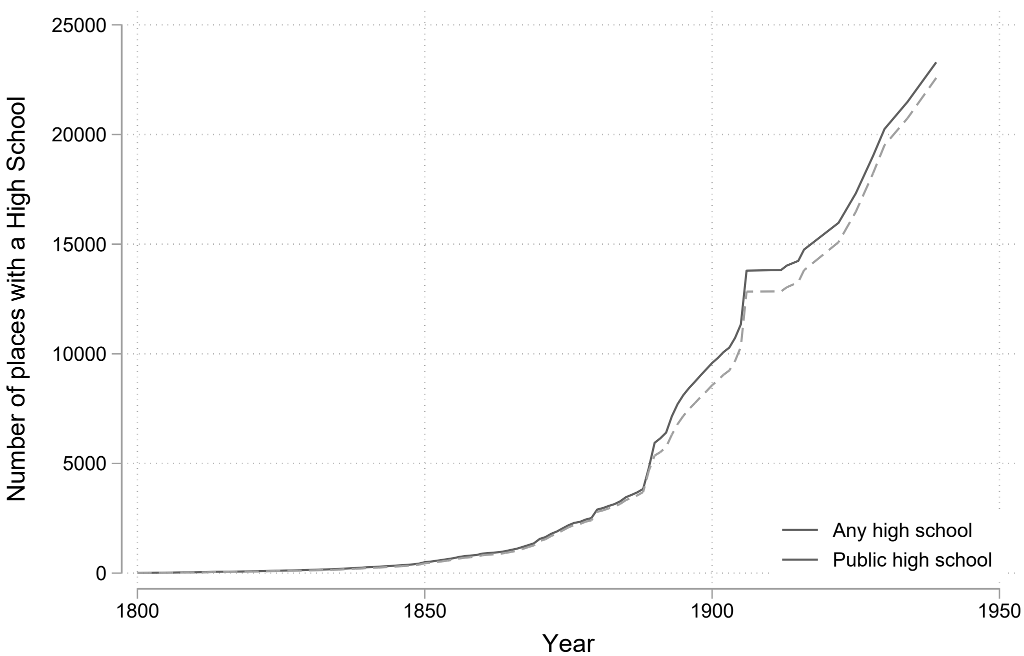


(d) Tennessee



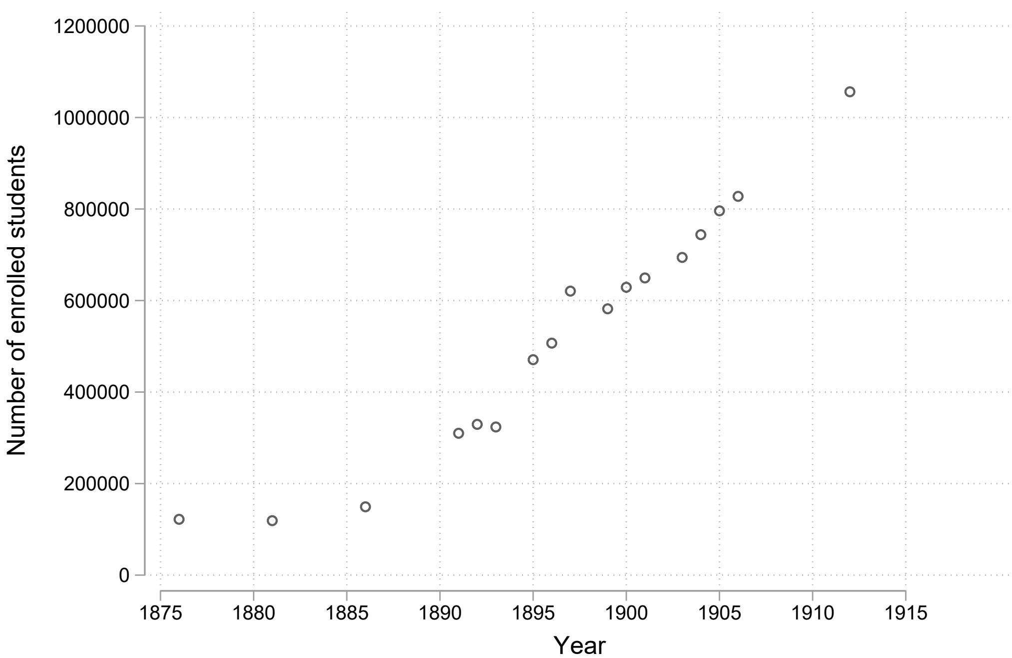
Notes: This map shows the founding dates of the first high school built in each town we can geolocate in Massachusetts (a), Ohio (b), Wisconsin (c), and Tennessee (d).

Figure 3: Number of towns in the U.S. with high schools, by year



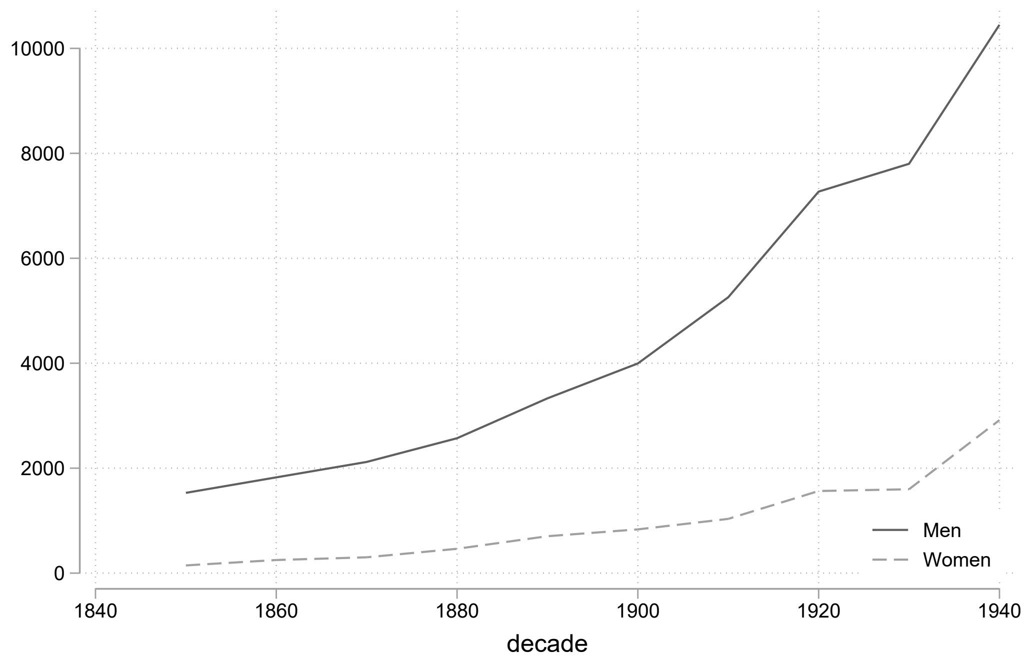
Notes: This plot shows the number of towns and cities with any high school (the solid line) and with any public high school (the dashed line) from 1800 to 1950. The founding dates are drawn from the panel described in the Data section above.

Figure 4: Number of students enrolled in high school, by year



Notes: This plot shows the number of students enrolled in high schools in years for which we have valid enrollment data.

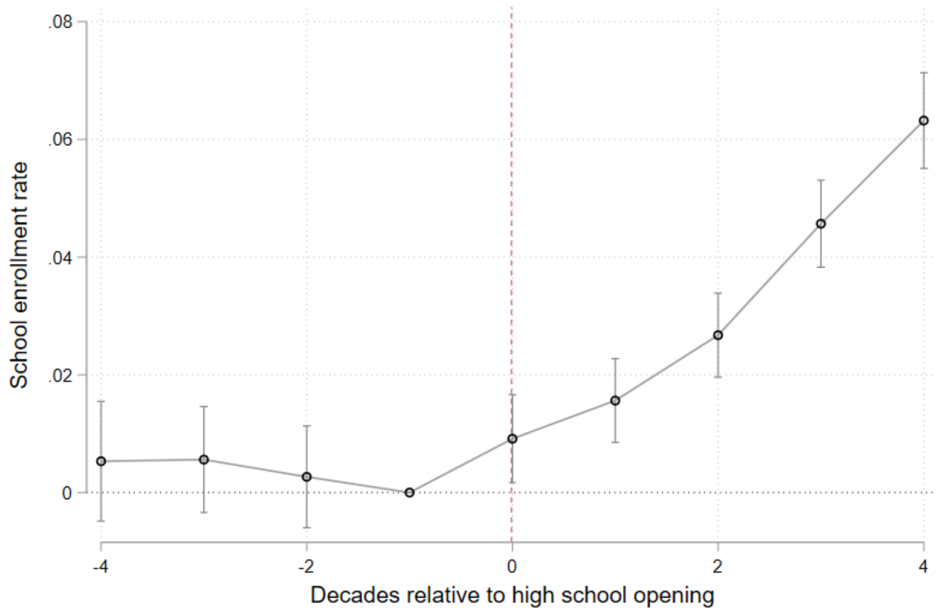
Figure 5: Count of Wikidata eminent people, by birth year and sex



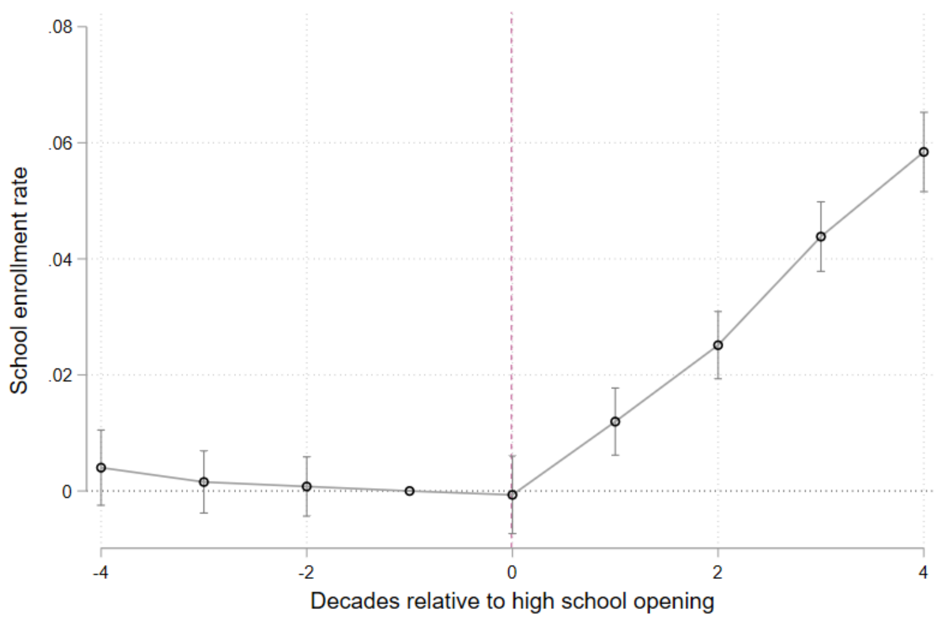
Notes: This plot shows the number of eminent men (solid line) and women (dashed line) in our panel of eminent people from Wikidata. The x-axis represents the birth year of the eminent people and the Y-axis is the count of people in our panel.

Figure 6: Effect of high school entry on 14–18-year-old school enrollment

(a) Baseline, two-way fixed effect



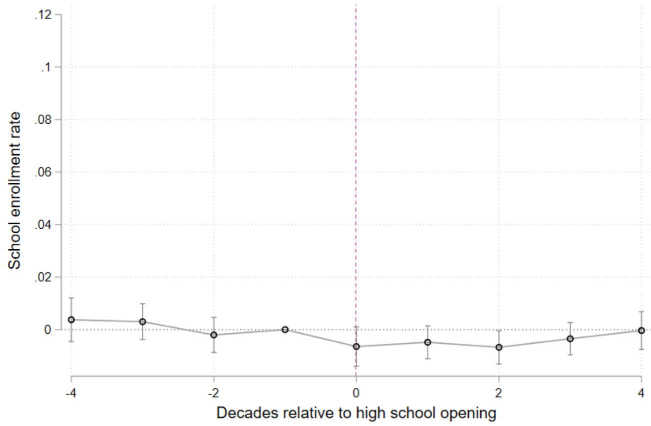
(b) Gardner (2021) imputation method



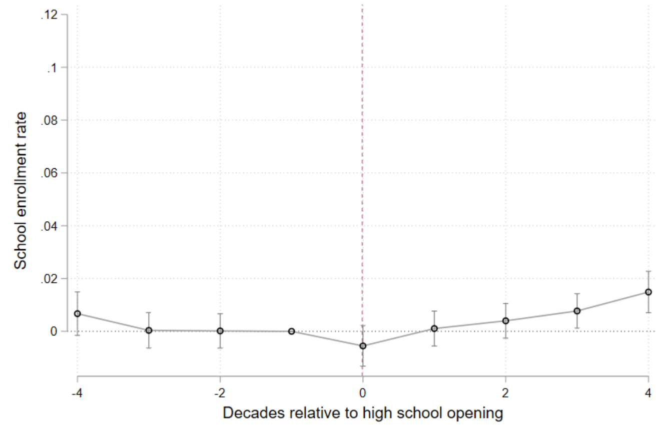
Notes: This shows our baseline event study results on school enrollment estimated from Equation 1. Standard errors are clustered by city and each model includes state-year fixed effects. Panel A shows results from a standard two-way fixed effect event study. Panel B shows results using the Gardner (2021) imputation method.

Figure 7: Effect of high school entry on 13–18-year-old school enrollment, by age

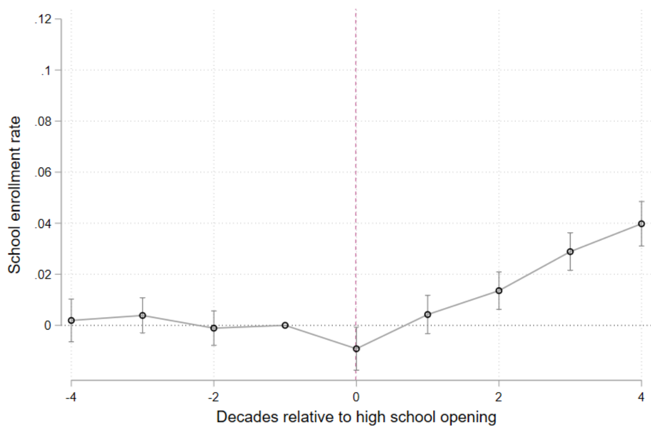
(a) Age 13



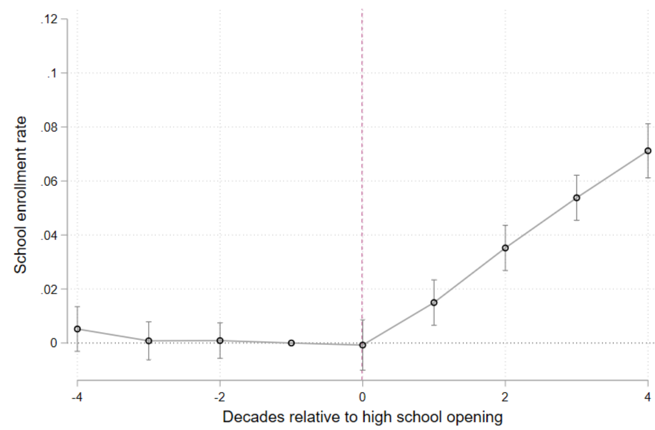
(b) Age 14



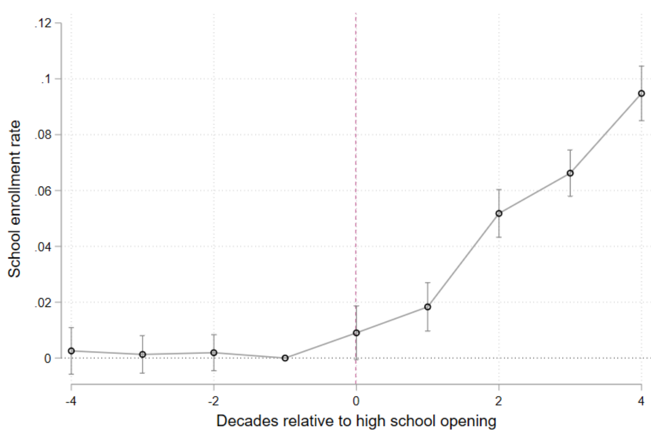
(c) Age 15



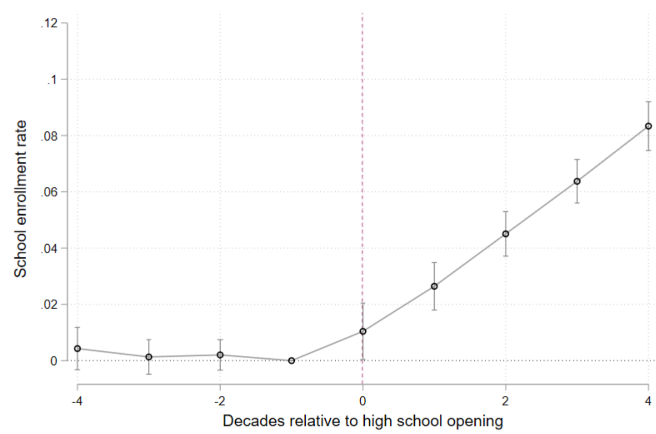
(d) Age 16



(e) Age 17



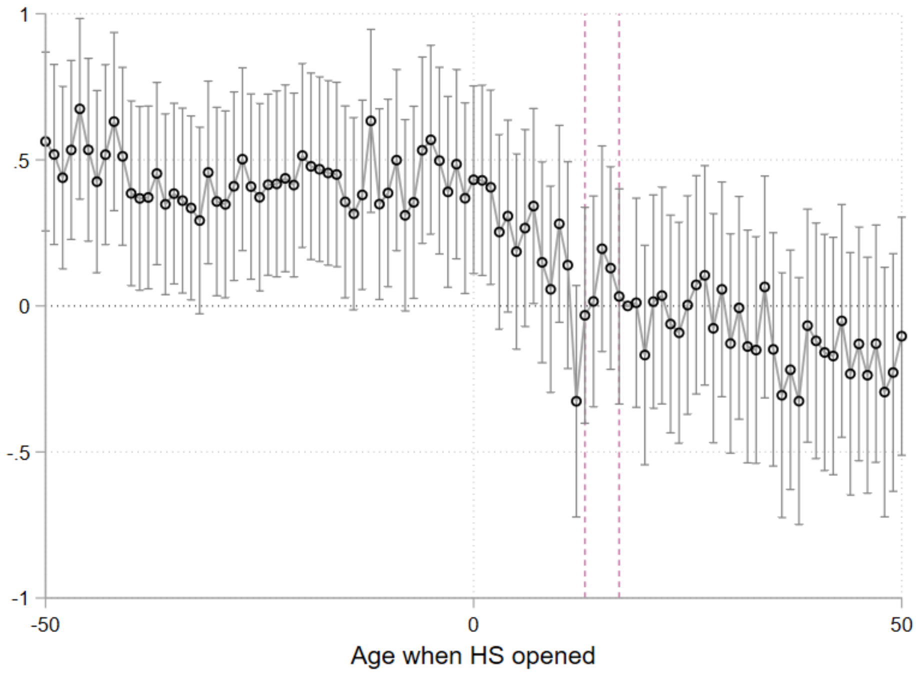
(f) Age 18



Notes: This shows our baseline event study results on school enrollment estimated from Equation 1 separately by age. Standard errors are clustered by city and each model includes state-year fixed effects. We show results using the Gardner (2021) imputation method.

Figure 8: Effect of high school entry on eminent people production

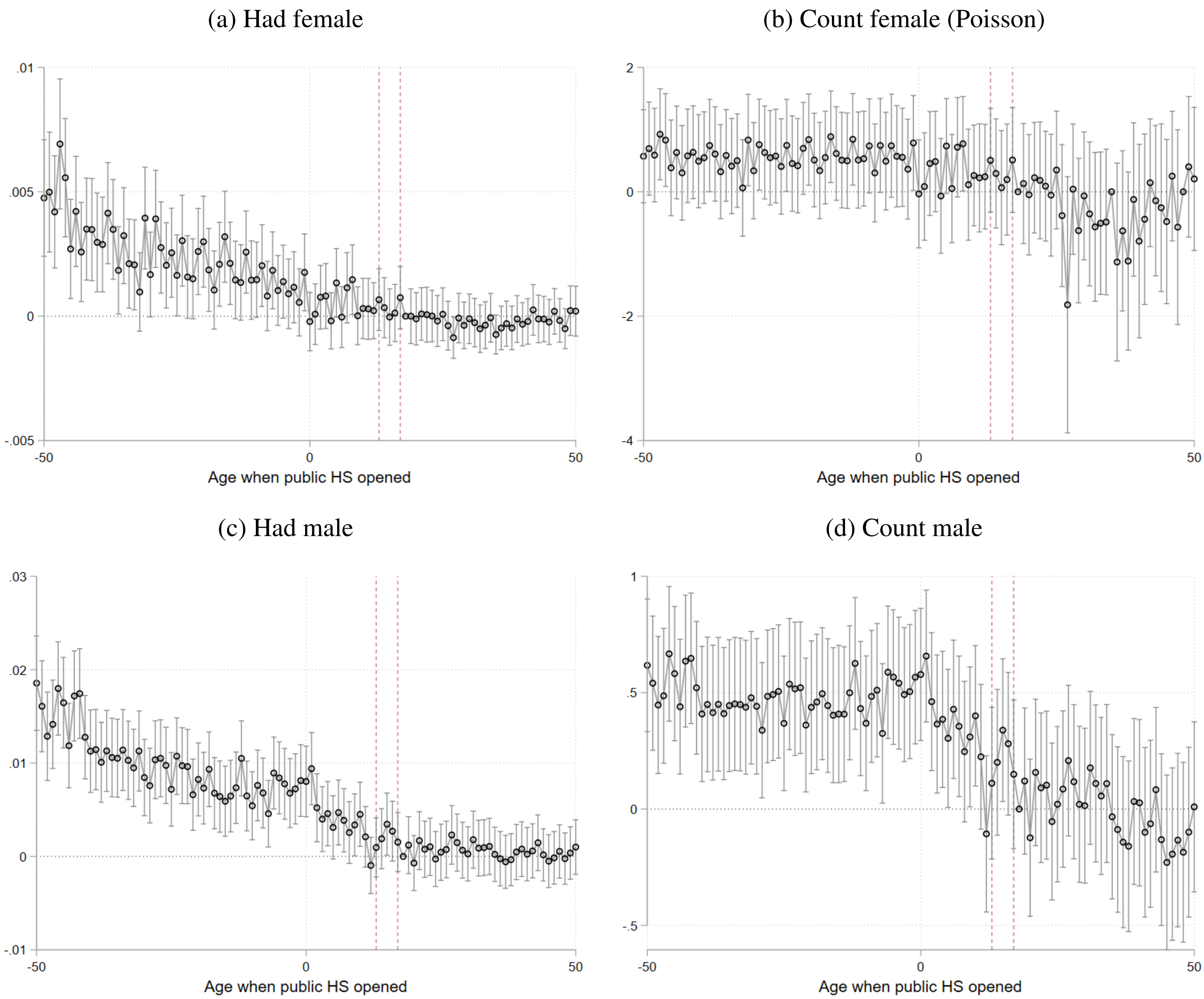
(a) Poisson count (intensive and extensive margin)



(b) Probability of observing an eminent person (extensive margin)

Notes: This shows our baseline event study results for eminent people using Equation 2. Standard errors are clustered by city and each model includes state-year fixed effects.

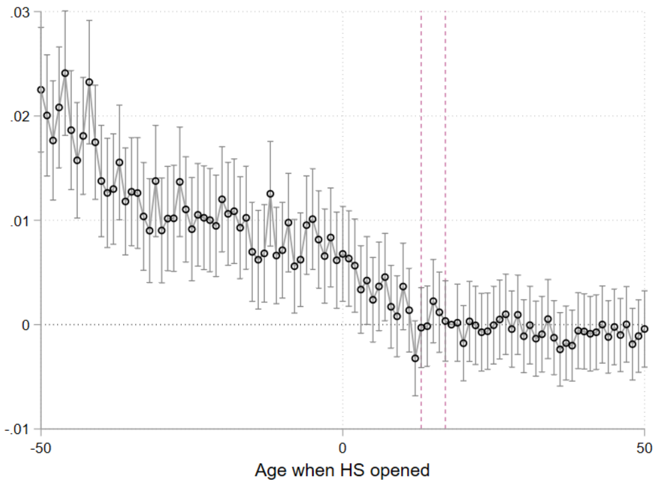
Figure 9: Effect of high school entry on eminent people production, by gender



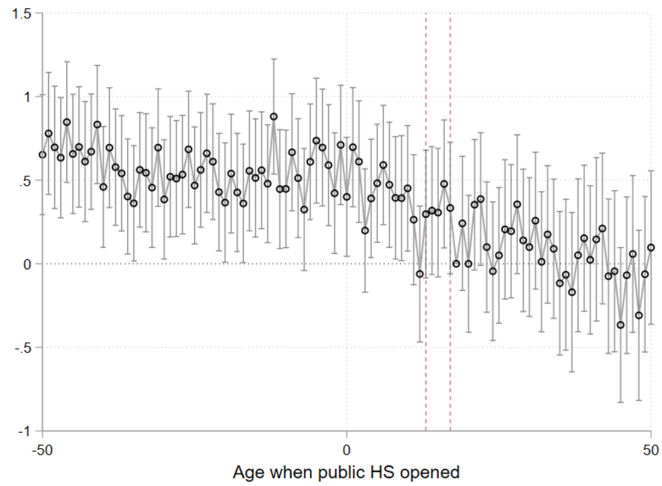
Notes: This shows our baseline event study results for eminent people using Equation 2 separately by gender. Standard errors are clustered by city and each model includes state-year fixed effects.

Figure 10: Effect of high school entry on eminent people production, robustness

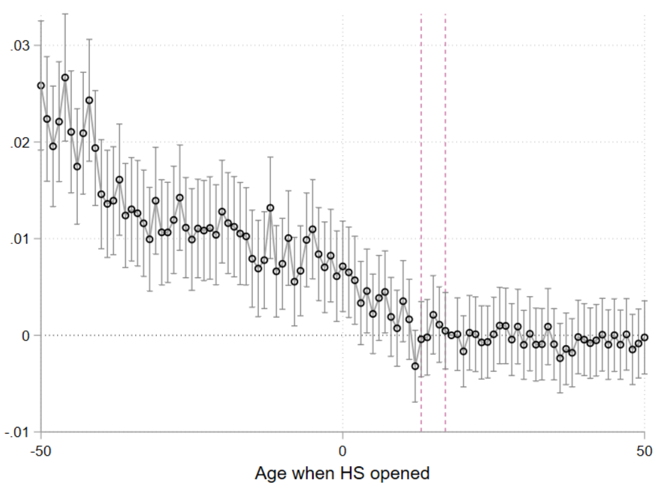
(a) Had eminent person (county-year fixed effects)



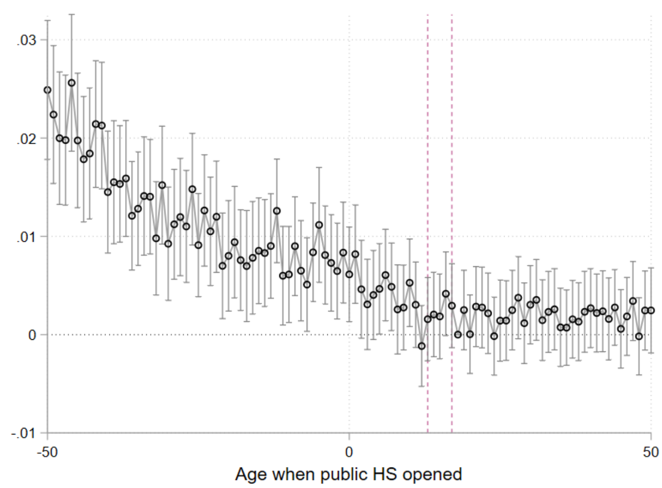
(b) Count eminent people (Poisson, county-year fixed effects)



(c) Count eminent people (counts, state-year fixed effects)



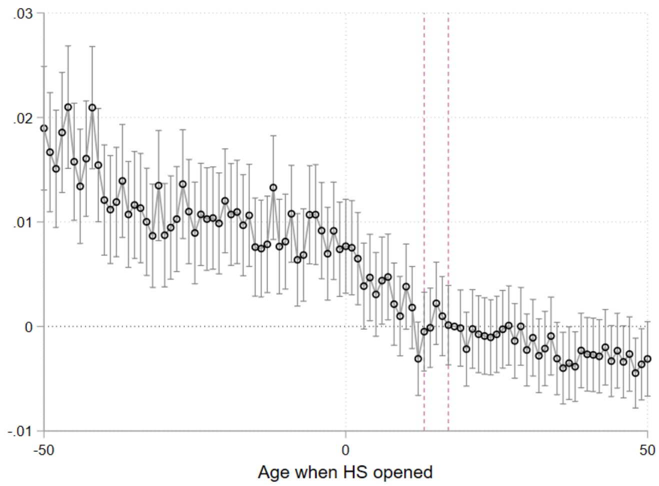
(d) Count eminent people (counts, county-year fixed effects)



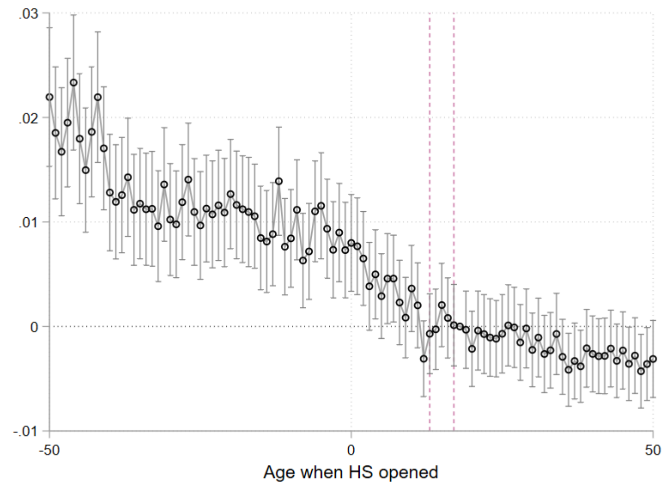
Notes: This shows our baseline event study results for eminent people using Equation 2 for alternative specifications and measures of eminence. Standard errors are clustered by city and each model includes the indicated fixed effects.

Figure 11: Effect of high school entry on eminent people production, Gardner (2021) robustness

(a) Had eminent person, TWFE

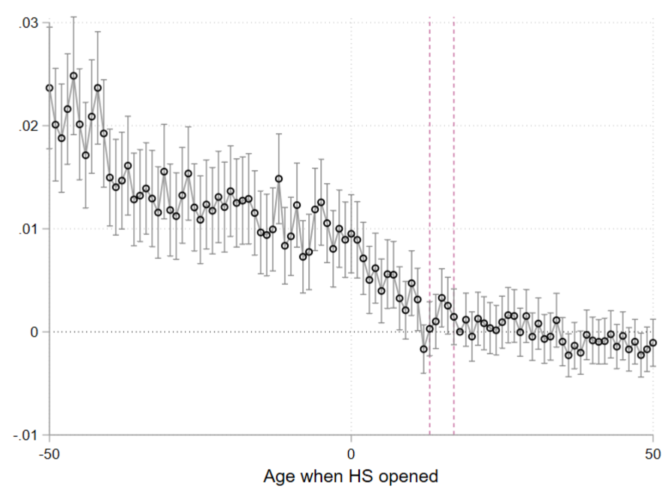


(b) Count eminent people (counts), TWFE



(c) Had eminent person, Gardner (2021)

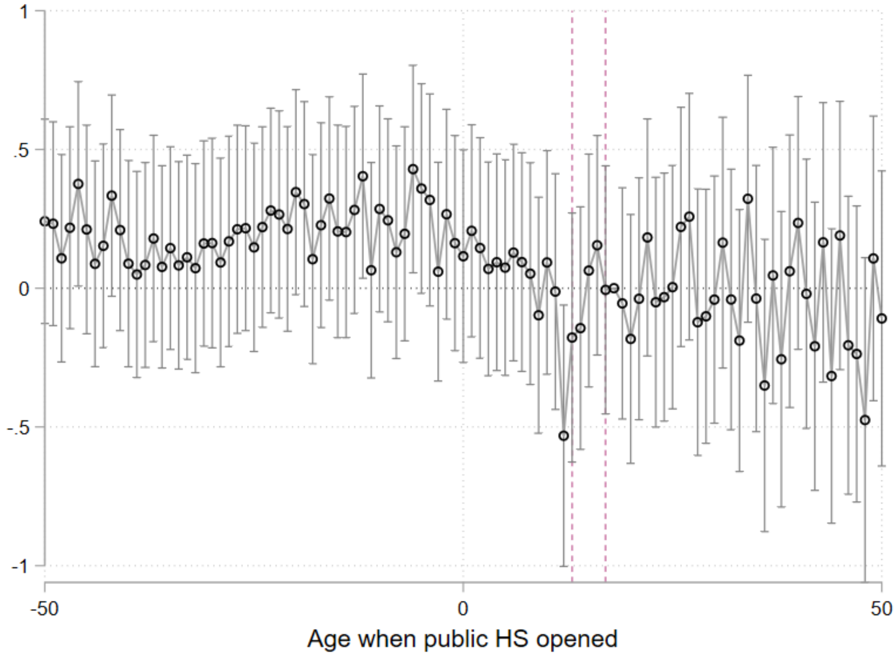
(d) Count eminent people, (counts, Gardner), TWFE



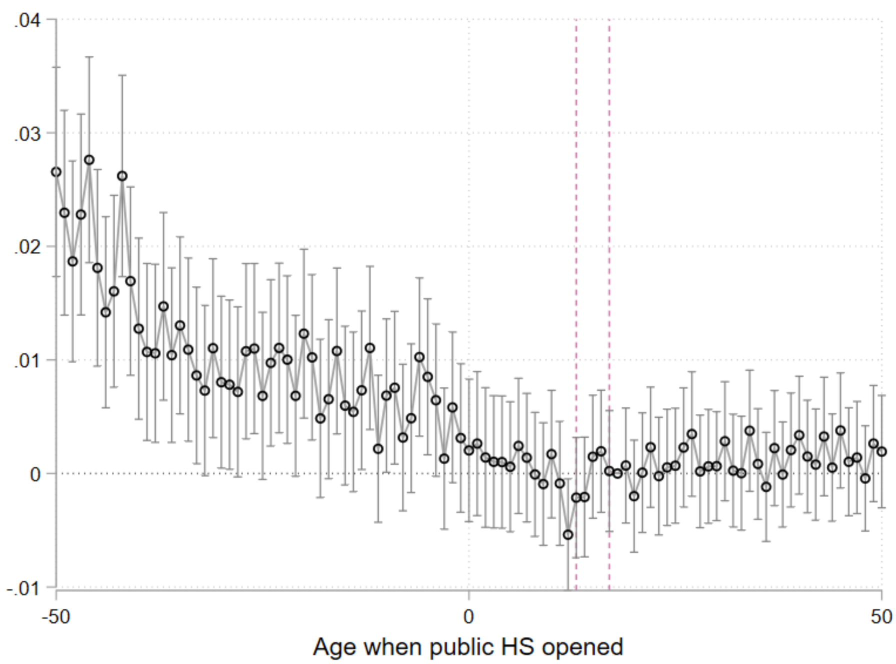
Notes: This shows our baseline event study results for eminent people using Equation 2 for alternative specifications and measures of eminence. Standard errors are clustered by city and models are estimated using the Gardner (2021) method.

Figure 12: Effect of high school entry on eminent people production, conditional on log time-varying population

(a) Poisson count (intensive and extensive margin)



(b) Probability of observing an eminent person (extensive margin)



Notes: This shows our baseline event study results for eminent people using Equation 2 for alternative specifications and measures of eminence. Standard errors are clustered by city. Models include state-year fixed effects and log, time-varying population controls.

References

[1] Aaronson, Daniel and Bhashkar Mazumder. "The Impact of Rosenwald Schools on Black Achievement." *Journal of Political Economy* Vol. 119, No. 5, pp. 821-888, 2011.

[2] Aghion, Philippe, Ufuk Akcigit, Antonin Bergeaud, Richard Blundell, and David Hémous. "Innovation and top income inequality." *Econometrica* Vol. 86, No. 1, pp. 1-45, 2019.

[3] Andrews, Michael. "How Do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of U.S. Colleges Mike Andrews." *American Economic Journal: Economic Policy (conditionally accepted).*

[4] Baker, Andrew C., David F. Larcker, and Charles CY Wang. "How much should we trust staggered difference-in-differences estimates?." *Journal of Financial Economics* 144, no. 2 (2022): 370-395.

[5] Bell, Alex and Chetty, Raj and Jaravel, Xavier and Petkova, Neviana and Van Reenen, John. "Who becomes an inventor in America? The importance of exposure to innovation." *The Quarterly Journal of Economics* Vol. 134, No. 2, pp. 647-713, 2019.

[6] Berkes, Enrico and Karger, Ezra and Nencka, Peter. "The Census Place Project: A Method for Geolocating Unstructured Place Names", 2022

[7] Duflo, Esther. "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment." *American Economic Review* Vol. 91, No. 4, pp. 795-813, 2001.

[8] Duflo, Esther. "The medium run effects of educational expansion: evidence from a large construction program in Indonesia." *Journal of Development Economics* Vol. 74, pp. 163-197, 2004.

[9] Gardner, John "Two-stage difference-in-differences." *Working paper*, 2021.

[10] Goldin, Claudia. "America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century." *Journal of Economic History* Vol. 58, No. 2, pp. 345-374, 1998.

[11] Goldin, Claudia. *Career and Family: Women's Century-Long Journey toward Equity*. Princeton University Press, 2021.

[12] Goldin, Claudia. "Egalitarianism and the Returns to Education during the Great Transformation of American Education." *Journal of Political Economy* Vol. 107, No. S6, pp. S65-S94, 1999.

[13] Goldin, Claudia and Lawrence F Katz, *The Race Between Education and Technology*, Harvard University Press, 2010

[14] Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. "The allocation of talent and US economic growth." *Econometrica* Vol. 87, No. 5, pp. 1439-1474, 2019.

[15] Navarro-Sola, Laia "Secondary Schools with Televised Lessons: The Labor Market Returns of the Mexican Telesecundaria." *Working paper*, 2021.

[16] Schmick, Ethan J., and Allison Shertzer. "The Impact of Early Investments in Urban School Systems in the United States" *Working paper*, 2020.