Rural Education, Nation-building, and Human Capital in Post-Revolutionary Mexico

Ariel Gómez
Northeastern University
School of Public Policy

9/14/2023
Please click here for the most recent version of this paper

Abstract

Can low-quality investments in schooling displace the creation of higher-quality schools in the long run? I investigate whether Mexico's rapid expansion of rural education during the 1920s successfully promoted human capital acquisition. By combining newly digitized schools data with census samples from 1930 and 1960, I find that Mexico's first major expansion in public education increased literacy rates and language homogenization, but did not induce primary school completion. Through a difference-in-differences analysis, I first show that rural schools achieved short-run education goals central to Mexico's nation-building efforts: by 1930, treated cohorts were up to 10% more likely to be literate and, among indigenous communities, up to 20% more likely to speak Spanish and 4% less likely to speak an indigenous language. I then show that while literacy gains persist into 1960, other long-run education measures suggest a lack of broader human capital growth. Affected cohorts were more likely to have attended but less likely to have completed primary school. These findings highlight the potential drawbacks of rapid, low-quality investments in human capital: the rural school program increased literacy in the short run, but without additional public investments in education, affected cohorts may have had fewer opportunities to complete primary education.

1 Introduction

Are rapid expansions in access to education an effective policy effective for a broad and equitable acquisition of human capital? Can low-quality investments in schooling displace the creation of higher-quality schools in the long run? To address these questions, I investigate the expansion of the federal education system in post-Revolutionary Mexico, which led to the opening of 2,600 rural schools[1] between 1922 and 1926. This program, one of the first state-led development projects in Mexico after the Revolution (1910-1920), anticipated future state intervention in economic policy. Government officials

[1]Tabulations indicate that more schools were established but some were abandoned. Further archival evidence is needed to identify schools that either failed or were transferred to local governments.

at the time believed that a robust public education system was a critical first step in modernizing Mexico's economy and society, through the spread of literacy and a Mexican national identity. To evaluate whether this program achieved these policy goals, I assess whether access to rural schools led to human capital acquisition and cultural assimilation in the short and long run.

Because the allocation of schools across space was not random, a simple comparison between treated and untreated locations cannot cleanly identify the effect of schools on future outcomes. To evaluate the impacts of this program, I have collected and digitized information on the locations of each school in 1925 and 1926. I combine these data with a microsample from the 1930 census, which reports whether an individual is literate, speaks Spanish, and speaks any indigenous language. For my long-run analysis, I use the IPUMS microsample of the 1960 census to assess long run impacts. With these data, I can account for selection in the allocation of rural schools across space with a difference-in-differences design. Using a two-way fixed effects estimator, I interact geographic variation in the locations of federal rural schools with cross-cohort variation in each census to estimate the impacts of exposure to rural schools on my outcomes. In the short run, I show that children exposed to these schools were more likely to be literate in Spanish by 1930. In predominantly indigenous areas, exposed children additionally experienced language homogenization: they were more likely to be fluent in Spanish and less likely to speak any indigenous language, an indication of cultural assimilation.

Using data from the 1960 census, however, I find little evidence of additional educational gains in the long run. Through a difference-in-differences design, I again find that exposed cohorts report a higher literacy rate. For other educational outcomes, I find mixed results. On one hand, younger cohorts in areas with more schools per capita were less likely to have completed primary school and show marginally significant losses in years of schooling. On the other hand, these same individuals also reported a higher likelihood of having attended primary school at all.

How can these results be reconciled? Through a combination of descriptive data and historical evidence, I suggest that initial, low-quality investments in education may have limited future investments as Mexico urbanized and as improved its fiscal position. Historical evidence shows that many rural schools were provided with few resources from the federal government. Moreover, they were run by teachers with just enough training to teach literacy and Spanish. In later years, as the government continued to expand educational access, it may have established better quality schools only in areas untouched by the 1920s school expansion. Moreover, as road access grew and populations urbanized, children in areas not treated by the rural school program may have gained greater access to urban schools with better resources and formally trained teachers.

Heterogeneity in the share of indigenous language speakers in the 1930s further supports this explanation. In contrast to the overall effects, exposed cohorts in majority-indigenous areas experienced a significant increase in the rate of primary school

completion.An explanation consistent with this finding is that while the 1920s and 1930s may have been a period of increased public investment in predominantly indigenous areas, by the 1940s, political support within the government for further investment may have waned. Evidence gathered by historians, along with contemporary primary sources I have reviewed, support this interpretation. From this perspective, rural schools may have been an educational trap for non-indigenous rural populations as well as one of the few significant public investments in indigenous communities in the post-Revolutionary period.

I mainly contribute to work on school expansions in developing countries and their role in education, nation-building and language homogenization (Alesina et al., 2020, Almargo and Andrés-Cerezo, 2019, Anderson, 1984, Assouad, 2021, Bazzi et al., 2021, Duflo, 2001, Weber, 1978). Unlike Duflo (2001), I find that the school expansion in Mexico did not lead to increases in years of education, suggesting that the quality of initial educational investments may hinder future improvements. My work is also related to Blanc and Kubo (2021), who show that state-sponsored education in France led populations to adopt the modern French dialect. I instead examine language homogenization among indigenous populations that managed to retain their native languages after hundreds of years of colonization.

I also contribute to the historical and empirical literature on the period of reconstruction and state-building in Mexico after the Revolution. (Krauze et al., 1981, Meyer, 1976, Sánchez-Talanquer, 2018, Vaughan, 1982) Elizalde et al. (2021), for example, find that the Mexican highway system did not expand into areas where politically hierarchical indigenous groups were dominant. While some ethnic groups opposed infrastructure they viewed as a threat to their way of life, I show that rural education successfully penetrated indigenous communities in Mexico. Resistance to roads and modernization, combined with changes in government priorities, may have inhibited further educational investments in indigenous areas.

I also add to a growing literature on the history of indigenous populations in Latin America (Angeles and Elizalde, 2017, Diaz-Cayeros, 2011, Diaz-Cayeros and Jha, 2022, Elizalde, 2020, Valencia Caicedo, 2019). In particular, Diaz-Cayeros (2011), Diaz-Cayeros et al. (2021), and Diaz-Cayeros and Jha (2022) study how variation in the rapacity of the Spanish crown during the colonial era impacted the survival of indigenous groups and patterned their process of cultural assimilation. Unlike the colonial policies of earlier periods, the rural education program in the 1920s had the explicit goal of molding indigenous people into Spanish-speaking *mestizos*. My results suggest that in the absence of directed state action, indigenous peoples would have experienced lower rates of assimilation, at least in the short run.[2]

[2]The Mexican government was not the only one that successfully assimilated native populations through schooling. In a series of papers, Feir (2016a,b), Feir and Auld (2021) and Gregg (2018) study forced assimilation policies in North America, where native children were taken from their homes and sent to boarding schools. While Mexico's rural schools appears to have been less coercive, some communities did see them as a forced and foreign imposition. Contemporary reports from school inspectors note that the local clergy was often suspicious of secular schools (Secretaría de Educación Pública, 1927), while

The rest of this paper is organized as follows. I first discuss the historical context of the rural schooling program. I then describe my various data sources, explaining how I link my newly digitized information on rural schools to existing databases. Next, I outline my empirical strategy and present my short-run results. I follow with my long-run results along with empirical tests. I conclude by suggesting future directions for research on education policy in post-Revolutionary Mexico.

2 Historical Context

2.1 Education Policy in Post-Revolutionary Mexico

The rural education program in 1920s Mexico was intimately linked to the Mexican Revolution, a multi-sided civil war that began in 1910. In spite of deep ideological differences, most revolutionary leaders believed education to be central to the future development of Mexico, as well as a way of gaining popular support. While fighting continued until 1920, the liberal faction of the war established a constitution in 1917. Among other things, the new constitution mandated primary schooling, prohibited clerical primary education, and devolved authority over public education to the local level (Vaughan, 1982). Instability continued in the following years, including a wave of school closures due to the lack of support from the federal government. In 1920, generals Álvaro Obregón and Plutarco Elias Calles took advantage of frustrations with the new government and staged a successful coup. The subsequent election of Obregón as president in 1920 marked the end of the Revolution.

During this time, José Vasconcelos, a revolutionary intellectual and former rector of the National University, believed that schooling and literacy were necessary for the modernization of rural Mexico. Under the new regime, he advocated for a constitutional amendment in 1921 which would return authority over public education to the federal government. This amendment established a new education department, the *Secretaría de Educación Publica* (SEP).[3]

As the director of the SEP from 1921 to 1923, Vasconcelos led a wave of educational and cultural initiatives[4] to promote his vision of Mexican nationalism. As a vanguard of the *indigenista* movement, he rejected theories of European racial superiority and advocated for *mestizaje*, the cultural and racial mixing of indigenous and Spanish heritage. Vasconcelos believed this would be achieved through education, which would teach indigenous people Spanish while imparting them with Western values and a Mexican, *mestizo* identity(Vaughan, 1982).

historians have suggested that the growth in public secular education contributed to religious conflict between 1926 to 1929 (Meyer, 1976, Vaughan, 1982).

[3]The structure of the SEP was based on the previous regime's education department. Although pre-war federal education policy was limited to Mexico City and unincorporated territories, the head of the department at the time pushed for an expansion in rural schooling, though he did not have the time or resources to implement this before the Revolution (Vaughan, 1982).

[4]This included the financing of Mexican muralists like Diego Rivera, an avowed communist who emphasized the role of indigenous people in his renditions of Mexican history and the Revolution.

Table 1: Growth in Federal Rural Schools by State, 1922–1926

State	1922	1923	1924	1925	1926
Aguascalientes	5	0	6	25	24
Baja California	0	0	0	0	0
Campeche	0	2	0	17	25
Chiapas	10	32	82	77	121
Chihuahua	15	15	46	65	71
Coahuila	3	9	9	26	51
Colima	0	8	21	37	39
Distrito Federal	5	7	4	102	1
Durango	6	2	7	31	47
Guanajuato	8	11	22	107	145
Guerrero	22	30	22	79	131
Hidalgo	27	50	88	87	132
Jalisco	11	22	50	135	184
Michoacán	13	30	80	129	187
Morelos	13	13	24	34	35
México	24	54	91	200	216
Nayarit	14	12	51	47	80
Nuevo León	5	10	19	83	105
Oaxaca	38	72	84	101	167
Puebla	35	63	109	141	165
Querétaro	5	15	25	30	58
Quintana Roo	0	0	0	9	0
San Luis Potosí	8	21	48	87	148
Sinaloa	6	13	18	51	48
Sonora	0	16	34	33	76
Tabasco	5	3	0	0	18
Tamaulipas	7	14	12	6	40
Tlaxcala	0	4	6	35	51
Veracruz	11	37	55	52	109
Yucatán	0	0	0	0	2
Zacatecas	13	13	31	100	124
Total	**309**	**578**	**1,044**	**1,926**	**2,600**

Note: The list of rural schools in the capital, Mexico City, or Distrito Federal, includes a number of "cultural missions" (another SEP program) and schools which were eventually incorporated into the urban school system in Mexico City. All of my analyses exclude the capital.

Sources: *Boletín de la Secretaria de Educación Pública* 1, no. 3 & 4 (1923); and Secretaría de Educación Pública (1927)

2.2 The Rural School Program

Between 1922 and 1926, the SEP established 2,600 rural schools throughout Mexico (Vaughan, 1982).[5] This began through the work of cultural missionaries, teachers tasked with locating rural and indigenous villages in need of schools. Missionaries also located and informally trained resident teachers, who would focus on teaching literacy and basic arithmetic. By 1923, the rural schooling program began to formalize under the title *Casa del Pueblo* ("Home of the Village").

In addition to creating a literate populace, these schools served as part of a broader nation-building effort meant to modernize rural and indigenous populations. The role of schools as a tool for socialization on the one hand, and as a community center on the other, were further clarified under the new SEP leadership during the Calles administration. In 1927, President Calles wrote,

> "[The SEP's work] will consist not only of combating illiteracy, but also in attaining a harmonic development of the *spirit* of the peasant and indigenous populations...so [they] can be completely incorporated into civilization." (Secretaría de Educación Pública, 1927)[6] [emphasis added]

For context, Table 14 reports state-level data on federal rural school enrollment in September 1926,[7] showing that a total of 143,661 children[8] were enrolled in rural schools. This can be benchmarked against either the rural population,[9] more than 11 million or about 66.5% of the total population, or all children ages five to fourteen, just under 4 million in 1930.[10] Assuming that 66.5% of the 4 million children lived in rural areas in 1930, and using this as a proxy for rural children in 1926, about 5.4% of rural children would have been enrolled in a SEP rural school.

3 Data

For this paper, I collect several Mexican data sources from the first half of the 20th century. The central component is a list of rural schools funded by the SEP in 1925 and 1926, which I obtain from contemporary statistical reports (Secretaría de Educación Pública, 1926b, 1927). For each year, I have digitized the names of the locality, municipality, and state associated with each school. Although I do not know the exact year in which schools were established, the rural schooling program began in 1922. Therefore, all these schools were founded over a five-year period.

[5] See Table 1 for state-level data on the growth of the rural schools program between 1922 and 1926. This growth is for federal rural schools only. For context, in 1925, federal schools made up approximately 48% of all rural schools in Mexico (Secretaría de Educación Pública, 1926a).

[6] This statement comes from a 1927 government report published in the midst of a violent religious uprising known as the *Cristro* War (1926-1929). This revolt was a backlash against Calles' anti-clericalism, including the perceived imposition of public secular education (Meyer, 1976).

[7] See **??** for state-level attendance data.

[8] These schools also offered night classes for older teenagers and adults, with 40,200 enrolled.

[9] Localities (which can be thought of as towns or villages) with no more than 2,500 inhabitants.

[10] Existing 1930 census tabulations do not show the number of school aged children in rural localities.

Table 2: Federal Rural School Enrollment, September 1926

State	Enrollment				Adult to Child Enrollment	
	Boys	Girls	Men	Women	Male	Female
Aguascalientes	546	519	230	22	0.42	0.04
Campeche	721	423	279	31	0.39	0.07
Coahuila	1,376	1,326	823	652	0.60	0.49
Colima	1,159	1,098	570	60	0.49	0.05
Chihuahua	1,770	1,342	313	64	0.18	0.05
Chiapas	4,528	1,517	910	50	0.20	0.03
Distrito Federal	74	71	100	42	1.35	0.59
Durango	1,628	1,316	777	87	0.48	0.07
Guanajuato	4,816	3,616	2,324	504	0.48	0.14
Guerrero	4,112	2,343	1,181	107	0.29	0.05
Hidalgo	6,011	3,090	2,363	190	0.39	0.06
Jalisco	6,181	6,188	2,906	1,047	0.47	0.17
México	8,253	5,527	4,367	266	0.53	0.05
Michoacán	5,574	5,094	2,241	483	0.40	0.09
Morelos	1,106	958	542	33	0.49	0.03
Najarit	1,827	1,917	461	139	0.25	0.07
Nuevo León	3,034	2,474	1,141	338	0.38	0.14
Oaxaca	6,768	2,128	1,824	114	0.27	0.05
Puebla	6,615	3,538	2,587	233	0.39	0.07
Querétaro	1,945	1,060	929	50	0.48	0.05
San Luis Potosí	5,948	4,309	3,346	188	0.56	0.04
Sinaloa	1,159	1,207	520	59	0.45	0.05
Sonora	1,211	1,083	324	95	0.27	0.09
Tabasco	578	193	182	17	0.31	0.09
Tamaulipas	765	529	591	31	0.77	0.06
Tlaxcala	1,576	1,471	766	201	0.49	0.14
Veracruz	3,687	937	1,027	50	0.28	0.05
Yucatán	26	30	0	0	0.00	0.00
Zacatecas	2,601	2,779	1,331	89	0.51	0.03
Total	**85,578**	**58,083**	**34,958**	**5,242**	**0.41**	**0.09**

Source: Secretaría de Educación Pública (1927)

My main outcomes are literacy, a measure of educational attainment, and fluency in Spanish and indigenous languages, the latter of which proxy for cultural assimilation. The most basic goal of rural schooling was to increase literacy rates, which officials viewed as a prerequisite for the transformation of the rural sector. The other central goal was to "civilize" indigenous populations by inculcating them with a sense of modernity and a Mexican identity. Although this change was to encompass many cultural attributes, language was understood to be the most consequential dimension of ethnic assimilation. Increases in Spanish fluency would suggest that schools were successful in integrating children into Spanish-speaking society, whereas reductions in indigenous language fluency would indicate that schooling limited the transmission of indigenous culture.

To observe these outcomes, I use a 10% micro-sample of surviving documents from the 1930 Mexican census,[11] which has been digitized by a group of researchers at Universidad Autónoma Chapingo (Zamudio Sánchez et al., 2015, 2018). These data provide basic demographic information for each individual sampled, including sex, age,[12] disability status, and state of birth, as well as information on literacy, contemporaneous school attendance,[13] language spoken, and occupation. The data also includes information on the state, municipality, and locality[14] of residence in 1930.

The language data are separated into a question on fluency in Spanish, followed by a section in which enumerators wrote in any other language the individual spoke. The digitized data groups responses together, allowing me to create separate indicators for speaking Spanish and speaking any indigenous language. Literacy and language were collected only for individuals above the age of five.[15]

For my long-run estimates, I combine data on the number of rural schools by municipalities with the IPUMS 10% microsample of the 1960 census (Minnesota Population Center). These data allow me to observe each individual's literacy, years of schooling, and primary and secondary school completion.

4 Short-Run Analysis

In this section, I examine the short-run effects of Mexico's rural school program. Using a microsample from the 1930 census, I assess how literacy and language homogenization

[11] Some of the underlying census documents have been lost, but many of them are not relevant here. In particular, all of the original census forms for Mexico City, the national capital, and large portions of some small states have not survived (Zamudio Sánchez et al., 2015). Although the loss of Mexico City may be problematic for national estimates, the population of interest for this paper is rural, and thus does not include the capital.

[12] The census officially began in May 15, 1930, so the approximate year of birth can be obtained through a simple difference between 1930 and age (Instituto Nacional de Estadística y Geografía).

[13] I cannot observe educational attainment or years of schooling as they were not collected in 1930.

[14] These are the three main administrative divisions in Mexico, in descending order. Localities can be thought of as villages, cities, or other population centers.

[15] For a small proportion of the data, younger people, including some newborns, are reported to be literate and speaking Spanish or another language. This may be due to errors in reported age.

changed in response to increased access to schooling in both indigenous and non-indigenous communities.

4.1 Endogeneity in Rural School Allocation

It is useful to first consider why a simple OLS regressions may not identify the causal effects of rural schools on literacy and language. The primary concern is the possibility of selective allocation of schools across locations. Even if the SEP did allocate these schools to rural areas only, treated localities may have been closer to cities or the transportation network. Children in these areas may have already had access to urban schools, such that educational outcomes would have been higher compared to untreated localities even without the presence of rural schools. Moreover, indigenous communities in such a locality would have been more likely to interact with an urban, Spanish-speaking population. Such a proximity to urban centers may have incentivized indigenous communities to adopt Spanish to take advantage of economic exchange with Spanish speakers. If, instead, the government targeted more remote areas, where access to education was low and indigenous communities were more likely to retain their culture, a simple OLS estimate of the effects of schools on literacy and assimilation would be downward-biased. It is also possible that communities selected themselves into treatment by petitioning the government for schools, potentially biasing estimates toward increases in education and assimilation.[16]

Archival evidence, however, suggests that negative selection in the allocation of rural schools was more likely. For example, reports by SEP officials tasked with visiting rural communities provided information on each village, including a recommendation for whether to establish a local rural school. These reports were relatively standardized, often including: the total population of the locality; the school-age population, whether it was an indigenous community, and if so, with which ethnic group it was populated; whether there were pre-existing school; and whether the community appeared to be willing to contribute, both materially and socially, to the functioning of the school. Smaller localities lacking schools and with sizable indigenous communities were more likely to be recommended for the rural school program. Currently available data allows me to restrict my sample to rural localities with small populations and indigenous populations. Even combining these restrictions with fixed-effects and controls, however, a simple OLS regression may not account for other selection criteria or other sources of selection bias.

4.2 Census Sample

I make four main restrictions on the 1930 census sample. First, I include only individuals born between 1890 and 1924. Following Duflo (2001), I define a set of "young," fully

[16]For instance, according to one anecdote from a 1927 government report, in one community, the people were so well organized and civically engaged that they successfully petitioned the government for both the redistribution of land and schools. I also found archival evidence that some communities wrote letters to the president asking for a school to be built nearby.

exposed cohorts that were born between 1920 and 1924, or ages two to six in 1926 and ages six to ten in 1930. Based on my observations on school attendance in 1930, these cohorts were exposed to the rural education program at the ages at which they were most likely to have attended school. Children born after 1924 were age five or younger in 1930 and unlikely to be attending school during my sample period , so they are excluded from the analysis. I further define a set of "old," fully unexposed cohorts, born between 1890 and 1908; these cohorts would have been at least 14 years old in 1922, the first year of the SEP rural schools program. Because school attendance fell rapidly after age 13 in 1930, these cohorts were unlikely to have attended these schools. I consider the remaining cohorts, born between 1909 and 1919, to have been partially exposed.

I use locality of residence in 1930 as a proxy for the locality of birth and restrict my sample to individuals living in their state of birth in 1930.[17] Locality of residence may incorrectly classify an individual's treatment status if they moved at any time after age six, inducing attenuation bias. To limit this, I exclude individuals not residing in their state of birth in 1930, though they make only up about 4% of my original sample. I further restrict the sample to localities in municipalities that had at least one locality with at least one federal rural school. I make this restriction because localities in treated municipalities were likely to be different from those in untreated municipalities.

To focus on the program's intended population, I also restrict to rural localities, that is those with populations no greater than 3,000 in 1930[18] and which were not classified as municipal capitals or cities in 1930 or earlier.[19] Given changes over time in the official definition of a rural locality,[20] I use 3,000 as the cutoff for rurality because the largest locality in my list of rural schools had a population of just over 2,700 in 1930.

In my main specification, I do not put additional restrictions on the sample. To assess the effects of rural schools on language homogenization in indigenous communities, a second specification is restricted to predominantly indigenous-speaking localities. Given that fluency in indigenous languages is one of my outcomes, I use individuals born before 1890 to estimate the proportion of indigenous speakers. For localities with at least one indigenous language-speaker born before 1890, 75% of localities have approximately 55% of their population speaking an indigenous language. I use this as a cutoff for defining a locality as primarily indigenous or non-indigenous. Alternative cutoffs give qualitatively similar results. In the Appendix, Figure 11 plots the cumulative distribution of this proportion. The vast majority of localities have zero observations of older indigenous

[17] It is possible that individuals from older cohorts moved to a different locality at higher rates than younger cohorts. This would induce more attenuation bias in older cohorts relative to younger cohorts, creating an upward bias in the magnitudes of my estimates.

[18] I obtain locality population from the historical time-series of census tabulations maintained by the Mexican statistical agency, INEGI.

[19] Exposure to the transportation network may confound my estimates. Although I do not have data on roads or railways, some localities are classified as train stations in either 1930 or earlier. I also drop these observations from my main sample.

[20] The official definition of a rural locality changed across different censuses, with population cutoffs set at 4,000 in 1910, 2,000 in 1921, and 2,500 in 1930 (Secretaría de Economía, 1930). Moreover, a statistical report from the SEP for 1925 notes that rural schools were then defined as schools in localities with no more than 1,000 people.

people in my sample, indicating these are locations with small indigenous populations. The distribution of indigenous-speaking populations suggests considerable segregation: most localities with at least one indigenous language speaker had a substantial majority of indigenous language speakers.

4.3 Descriptive Statistics

Table 3: Descriptive Statistics for Census Sample

VARIABLES	(1) N	(2) Mean	(3) SD
Literate	239,506	0.199	0.399
Attendance (6-15)	57,212	0.272	0.445
Speaks Spanish	239,506	0.742	0.437
Speaks Indig. Lang.	239,506	0.174	0.379
Non-Migrant	239,506	0.959	0.198
Age	239,452	22.38	18.14
Male	239,471	0.503	0.500
Treat	239,506	0.293	0.455
Field Worker (M 15+)	120,493	0.436	0.496
Homemaker (F 15+)	71,029	0.955	0.208

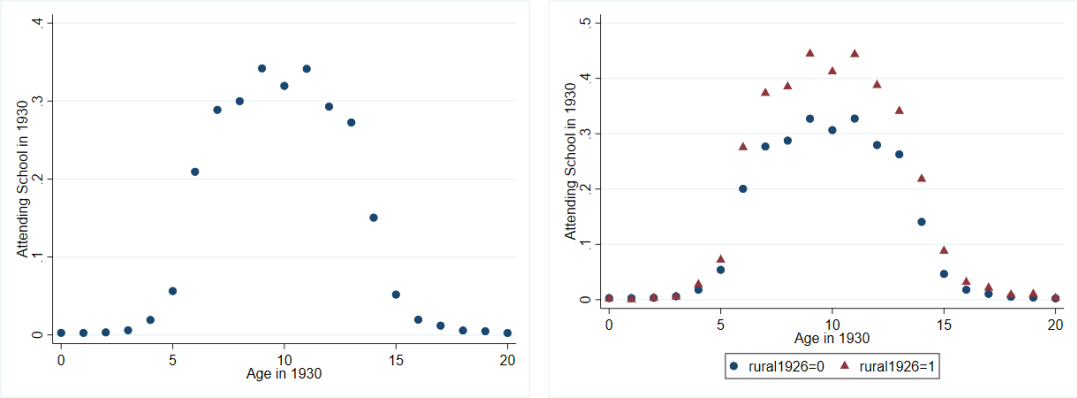
Data is restricted to localities that were not classified as municipal capitals, cities, or railroad stations in 1930 or earlier. I also exclude localities that are part of municipalities which did not have any federal rural schools in either 1925 or 1926, as well as localities with populations greater than 3000 in 1930.

Table 3 shows descriptive statistics for individuals in the 1930 census sample. The proportion of the population that speaks Spanish is lower than the reported national value,[21] and although literacy rates and school attendance are low relative to modern standards, they are far from negligible during this period. Very few individuals live outside of their state of birth, reducing concerns about the exclusion of migrants. Finally, I note that about 30% of the sample resided in a treated locality in 1930.

Although official school-going ages were from six to fourteen, it is useful to get a sense of what actual attendance, and therefore exposure to schools, looked like in 1930. In Figure 1a, I plot a binned-scatter plot for school attendance by age for people twenty years old and younger in 1930. For those below five and above seventeen, attendance rates are virtually zero. The primary age range for school attendance appears to be from six to thirteen, peaking at nearly 40% for ages nine and eleven. Assuming that the distribution of school attendance by age was similar during the 1920s (and not accounting for adults attending night school), the earliest cohort plausibly exposed to the rural school program would have been approximately thirteen years old in 1922 (born in 1909). In addition, the first cohort I know to have been fully exposed based

[21] Based on census tabulations, approximately 91.4% of the Mexican population spoke Spanish in 1930.

on the list of 1926 schools was six in 1926, thus born in 1920. This gives a range of birth years from 1909 to 1919 of plausibly partially exposed cohorts versus fully exposed cohorts born between 1920 and 1924, excluding children younger than six in 1930.



(a) Ages 0 to 20 **(b)** By Treatment Status

Figure 1: School Attendance Rates in 1930

In the second panel, Figure 1b shows the same plot split by whether the individual resided in a locality with a rural school in 1925 or 1926. Attendance rates in 1930 are about 10% higher in treated locations versus untreated locations for ages six to fourteen, suggesting that rural schools may have indeed increased school attendance.

4.4 Empirical Specification

To account for these endogeneity concerns, my empirical strategy exploits spatial and cohort variation in the exposure to federal rural schools. My difference-in-differences specification first compares outcomes for individuals in cohorts older than age thirteen in 1922, likely too old to benefit from new schools, to cohorts born between 1920 and 1924, young enough to have benefited but old enough to be in school by 1930. The second difference is a cross-sectional comparison between the outcomes of individuals in localities with and without federal rural schools in either 1925 or 1926. Following Duflo (2001) and Fabregas (2021), my main individual-level empirical specification is

$$Y_{ilt} = \alpha + \beta \text{Young}_t \times \text{School}_l + \psi X_i + \gamma_t + \delta_l + \phi_{mt} + \epsilon_{ilt} \tag{1}$$

where a literacy or language indicator Y_{itl} is defined for individual i born in year t residing in locality l in 1930. $Young_t$ is equal to one for individuals born after the threshold birth year (1920–1924) and zero otherwise. The main specification excludes partially treated cohorts (born between 1909 and 1919), such that the comparison cohorts were born between 1890 and 1908. $School_l$ is an indicator for localities that had a federal rural school in either 1925 or 1926. The main effects for $Young_t$ and $School_l$ are absorbed by the cohort and locality fixed effects, as explained below. X_i is a vector of individual characteristics, sex/gender, and disability status.[22]

[22]The census asks whether individuals suffer from a variety of ailments that may affect their vision, communication, physical movement, or learning rates. This control is an indicator equal to one if an individual is reported to have any disability.

I also include a set of fixed-effects to account for unobserved individual and location characteristics. γ_t is a cohort fixed-effect, which controls for differences between cohorts that do not vary with locality. If, for instance, older cohorts have more years of schooling and higher literacy rates than younger cohorts, cohort fixed-effects will account for these differences.

As discussed above, there may be either positive or negative selection in the allocation of rural schools. To the extent that selection is based on time-invariant characteristics of a location, δ_l, a set of locality fixed-effects, will account for them. Because exposure to treatment is defined by locality-cohort pairs, locality-by-cohort fixed-effects would absorb my identifying variation. However, I can interact fixed effects for municipality[23] m with cohort fixed-effects to control for local time-varying characteristics, ϕ_{mt}. For example, imagine a municipal capital was connected to the transportation network in the middle of my study period. Then these fixed-effects would control for the impact of this change on outcomes in the average locality within the affected municipality. Finally, to account for correlated shocks across individuals in the same location, and since my treatment variable is defined at the locality level, I cluster standard errors by locality.

The coefficient of interest in this regression is β, which will give the effect on the outcome Y_{itl} of being exposed to a rural school by age six if my identification assumptions hold.

Alternative specifications also include an interaction between treatment and an indicator for partial exposure, Partial_t:

$$\begin{aligned} Y_{ilt} = \alpha + \pi_1 \text{Young}_t \times \text{School}_l + \pi_2 \text{Partial}_t \times \text{School}_l \\ + \psi X_i + \gamma_t + \delta_l + \phi_{mt} + \epsilon_{ilt} \end{aligned} \tag{2}$$

All other terms are defined as in Equation 1. Once again, the main effects for $Young_t$, Partial_t, and $School_l$ are exclude because absorbed by the cohort and locality fixed effects. To check graphically for differential effects between cohorts and potential pre-trends, I additionally estimate a dynamic difference-in-difference specification interacting treatment with cohort indicators:

$$Y_{ilt} = \alpha + \sum_t \beta_t \text{School}_{lt} + \psi X_i + \gamma_t + \delta_l + \phi_{mt} + \epsilon_{ilt} \tag{3}$$

Because of the fuzziness of the treatment cutoff, in my dynamic plots below, I use the oldest cohort as the baseline and expect significant changes only for cohorts born after 1908. I run specifications grouping cohorts into bins of two, three, or five years. Therefore, Treatment_{lt} is an treatment indicator equal to one for an individual born in year t residing in a treated locality l in 1930. Given that base rates for literacy and language vary by gender, I run regressions separately for males and females. Finally, I

[23] Recall that municipalities are the administrative level just above localities. Then many localities may be located within the same municipality.

Table 4: Results: Female Literacy

	(1)	(2)	(3)	(4)
	Respondent Can Read and Write			
VARIABLES	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) × Rural School	0.0742**	0.0713**	0.0708***	0.0704***
	(0.0308)	(0.0313)	(0.0213)	(0.0212)
	[2.411]	[2.277]	[3.324]	[3.324]
Born 1909-1919 (Partial) × Rural School		0.0449***		0.0721***
		(0.0145)		(0.0119)
		[3.086]		[6.047]
Outcome Mean	0.0955	0.0991	0.230	0.254
Observations	10,610	15,456	36,774	55,948
R-squared	0.544	0.525	0.457	0.453
Controls	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Locality FE	Y	Y	Y	Y
Muni-Cohort FE	Y	Y	Y	Y
Clusters	Localities	Localities	Localities	Localities
Localities	512	518	2087	2103
F-stat	5.813	5.289	11.05	18.46
R2 w/in adj.	0.00337	0.00219	0.00138	0.00143

Standard errors clustered by localities in parentheses, t-stats in brackets. All regressions restrict to individuals born between 1890 and 1924 residing in their state of birth in 1930. The left out group in all columns is the 1890-1908 cohort. Localities are restricted to those in municipalities with at least one rural school, excluding municipal capitals, cities, and train stations, keeping localities with populations no greater than 3000 in 1930. All regressions control for gender/sex and disability status, as well as locality, cohort, and municipality-by-cohort. Odd numbered columns show estimates for fully exposed cohorts only, while even columns include an additional interaction for partially exposed cohorts. Columns 1 and 2 restrict to indigenous localities, while Columns 3 and 4 restrict to non-indigenous localities.

run additional regressions restricting to localities estimated to have significant indigenous populations.

Conditional on municipality-by-cohort fixed-effects, the above two-way fixed-effects estimators should identify the causal effects of the rural schooling program under the assumptions of parallel trends, no anticipation effects, and homogeneity in treatment effects across localities and between cohorts (Roth et al., 2022). If my treated cohorts are exposed to schools during the same time period, then my dynamic specification will account for heterogeneous effects. In the results below, I show that pre-trends are not a concern for most of my results.

4.5 Results

My first set of results demonstrates that, in predominantly indigenous localities, the rural education program was successful in increasing literacy rates among both girls and boys, whereas only girls benefited in non-indigenous localities. Although I do not use

Table 5: Results: Male Literacy

	(1)	(2)	(3)	(4)
	Respondent Can Read and Write			
VARIABLES	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) ×	0.0813**	0.0890***	0.0188	0.0179
Rural School	(0.0336)	(0.0338)	(0.0217)	(0.0214)
	[2.421]	[2.636]	[0.870]	[0.834]
Born 1909-1919 (Partial) ×		0.0690***		0.0271**
Rural School		(0.0222)		(0.0138)
		[3.113]		[1.969]
Outcome Mean	0.196	0.210	0.314	0.322
Observations	10,619	15,111	35,919	55,341
R-squared	0.494	0.485	0.433	0.424
Controls	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Locality FE	Y	Y	Y	Y
Muni-Cohort FE	Y	Y	Y	Y
Clusters	Localities	Localities	Localities	Localities
Localities	512	516	2088	2102
F-stat	5.860	5.856	0.757	1.972
R2 w/in adj.	0.00198	0.00191	4.55e-05	9.19e-05

Standard errors clustered by localities in parentheses, t-stats in brackets. All regressions restrict to individuals born between 1890 and 1924 residing in their state of birth in 1930. The left out group in all columns is the 1890-1908 cohort. Localities are restricted to those in municipalities with at least one rural school, excluding municipal capitals, cities, and train stations, keeping localities with populations no greater than 3000 in 1930. All regressions control for gender/sex and disability status, as well as locality, cohort, and municipality-by-cohort. Odd numbered columns show estimates for fully exposed cohorts only, while even columns include an additional interaction for partially exposed cohorts. Columns 1 and 2 restrict to indigenous localities, while Columns 3 and 4 restrict to non-indigenous localities.

the changes in literacy as a first stage, the results suggest that education is the likeliest mechanism through which language may have changed in indigenous communities. I then present results on fluency in Spanish and indigenous languages. My findings show that schools increased Spanish fluency among children in indigenous communities while decreasing their likelihood of speaking an indigenous language.

The tables in this section all follow the same format. In Columns 1 and 3, I use my main specification (see Equation 1), which excludes partially treated cohorts and compares only young, fully exposed cohorts to older, fully unexposed cohorts. As in Equation 2, I add a second interaction term for all partially treated cohorts in Columns 2 and 4. In Columns 1 and 2, I restrict my sample to indigenous localities, whereas in Columns 3 and 4, I restrict to non-indigenous localities. Columns 3 and 4 also serve as a placebo check on my language homogenization results: although rural schools may increase literacy rates in non-indigenous communities, we should not see a increase in Spanish fluency or loss in indigenous languages.

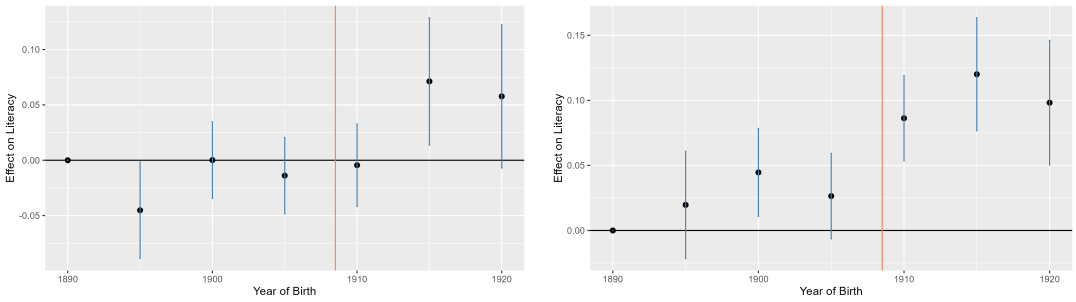
Table 6: Placebo Check: Literacy

	(1)	(2)	(3)
	Respondent Can Read and Write		
VARIABLES	Pooled	Female	Male
Born After 1900 ×	-0.00428	0.0186	-0.0169
Rural School	(0.0118)	(0.0130)	(0.0233)
	[-0.363]	[1.430]	[-0.724]
Outcome Mean	0.0911	0.0392	0.142
Observations	14,866	7,192	6,944
R-squared	0.330	0.398	0.439
Controls	Y	Y	Y
Cohort FE	Y	Y	Y
Locality FE	Y	Y	Y
Muni-Cohort FE	Y	Y	Y
Clusters	Localities	Localities	Localities
Localities	519	506	505
F-stat	0.132	2.045	0.524
R2 w/in adj.	-7.06e-05	0.000174	-9.03e-05

Standard errors clustered by localities in parentheses, t-stats in brackets. All regressions restrict to individuals born between 1890 and 1924 residing in their state of birth in 1930. The left out group in all columns is the 1890-1908 cohort. Localities are restricted to those in municipalities with at least one rural school, excluding municipal capitals, cities, and train stations, keeping localities with populations no greater than 3000 in 1930. All regressions control for gender/sex and disability status, as well as locality, cohort, and municipality-by-cohort. Odd numbered columns show estimates for fully exposed cohorts only, while even columns include an additional interaction for partially exposed cohorts. Columns 1 and 2 restrict to indigenous localities, while Columns 3 and 4 restrict to non-indigenous localities.

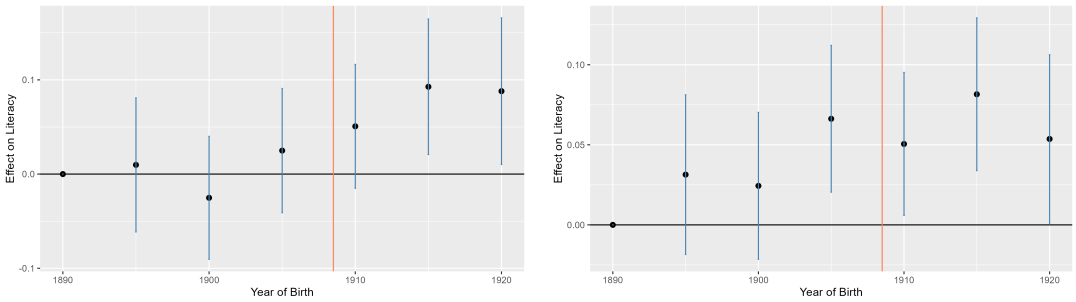
Literacy. Columns 1 and 2 in Table 4 and Table 5 show that in indigenous communities, both girls and boys exposed to federal rural schools were more likely to be able read and write than older, unexposed cohorts. Columns 1 and 2 in Table 4 and Table 5 show that effects for males and females are similar in magnitude, about 7-9%. However, since the female literacy rate (approximately 10%) is about half that of males in indigenous localities, these results suggest a doubling of literacy rates for females and an increase in 50% for males. In Column 2 of Table 4 and Table 5, the interaction for the partially exposed cohorts shows an increase of about 5% for girls and 7% for boys. These estimates are all statistically significant at the 1% level. Compared to mean literacy rates in the indigenous sample, these effects are smaller in both absolute and relative magnitude, but still suggest sizable increases in literacy. Given that school attendance rates are about 10% in 1930 for localities in rural schools compared to those without, a back of the envelope calculation suggests that about 70-80% of students who attended rural schools become literate.

In non-indigenous localities, the estimated effects are qualitatively different. First, Columns 3 and 4 in Table 5 show that fully exposed boys did not experience a significant change in literacy, with partially treated cohorts experiencing and increase of about 10%



(a) Indigenous Localities **(b)** Non-Indigenous Localities

Figure 2: Dynamic DiD: Effect of Rural Schools on Literacy, Females



(a) Indigenous Localities **(b)** Non-Indigenous Localities

Figure 3: Dynamic DiD: Effect of Rural Schools on Literacy, Males

relative to the sample mean. For girls, Columns 3 and 4 in Table 4 show a significant increase in literacy of around 7%, or about a 30% increase. These differences between indigenous and non-indigenous localities may be due in part to the fact that these populations already have relatively higher literacy rates, just under 30%.[24]

Taken together, these results suggest that the federal rural schooling program increased literacy, an important component of nation-building, in both indigenous and non-indigenous localities. Because the positive effects in non-indigenous communities are unlikely to lead to changes in language, these localities will serve as a placebo check for my results in the next set of outcomes.

As a check on the plausibility of the parallel trends assumption, in Table 6, I formally test for pre-trends in the indigenous sample by comparing cohorts born between 1890 and 1899 to those born between 1900 and 1908, both of which I expect to have been unexposed to rural schools. In Column 1, I test for pre-trends using the pooled sample, in Column 2 I restrict the sample to females, and in Column 3, I restrict my sample to males. In all three columns, the effects are insignificant and in Columns 1 and 3 they are small relative to the sample mean. While the size of the coefficient for females is about half the mean literacy rate, overall literacy is very low for older females. These results suggest that there are no statistically significant pre-trends in literacy.

As an additional tests for pre-tends, Figure 2 and Figure 3 plot the coefficients from my dynamic specification, allowing for a visual check for changes before the treatment

[24]If non-indigenous communities where more likely to have pre-existing schools, it is possible that boys already had access to education in some locations. If so, the SEP's literacy campaign may have additionally fostered gender equity in education.

period. In all dynamic plots, coefficients are benchmarked against the difference in literacy for the oldest cohorts, corresponding to the years between 1890 and 1894. The light red line marks the year 1909, which I use as the threshold for partial exposure in my main specification.

All four plots show that there is little evidence for significant changes in literacy for cohorts born before 1909,

In Figure 2a and Figure 3a, I plot coefficients from the dynamic specifications for the indigenous female and male samples, respectively. In both cases, pre-trends are relatively flat, whereas there appears to be an increase in literacy for the youngest cohorts. Figure 2b shows a similar pattern for non-indigenous females, but for males, Figure 3a shows noisy pre-trends without a clear break in literacy trends for later cohorts. These plots are consistent with the results in my regression tables. Moreover, along with my formal placebo checks in Table 6, the evidence suggests the plausibility of the parallel trends assumption, supporting a causal interpretation of my findings.

Overall, these results are evidence that, as part of its nation-building strategy, the federal government succeeded in increasing literacy through rural schools. Given the implied increases in literacy in non-indigenous localities, this sample will serve as an additional placebo test, as they should not have experienced changes in either Spanish or indigenous-language fluency. I next turn to my results on Spanish fluency.

Spanish fluency. Table 7 shows the estimated effects of exposure to rural schools on the propensity for girls to speak Spanish in 1930. In indigenous localities, Columns 1 and 2 show that exposure to schools increases the probability that a girl spoke Spanish by about 9%. Compared to a mean Spanish fluency rate of about 45%, this amounts to an increase of about 20%. Effects on partially treated cohorts are smaller but significant. For boys, the estimated effects Columns 1 and 2 of Table 8 are about 8%, which is a relative increase of about 15% compared to the sample mean of around 55%. In both tables tables, Columns 3 and 4 serve as placebo checks on non-indigenous localities. These results show that there is no evidence of any change in Spanish fluency among boys or girls in non-indigenous areas.

To test for pre-trends, Table 9 presents results from a formal placebo check using older, unexposed cohorts. A positive and significant estimate in any of the columns would suggest an upward trend in Spanish fluency before the schools were established, rendering the parallel trends assumption implausible. If anything, there appears to be a negative and significant downward trend for males in indigenous localities. This may suggest that my estimates effects are actually a lower bound of the true effect, limiting concerns over a potential violation of the parallel trends assumption.

Table 7: Results: Female Spanish Fluency

	(1)	(2)	(3)	(4)
	Respondent Speaks Spanish			
VARIABLES	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) × Rural School	0.0910***	0.0905***	-0.00248	-0.00270
	(0.0277)	(0.0279)	(0.00246)	(0.00239)
	[3.289]	[3.248]	[-1.006]	[-1.129]
Born 1909-1919 (Partial) × Rural School		0.0561***		-0.000801
		(0.0186)		(0.00234)
		[3.020]		[-0.342]
Outcome Mean	0.449	0.446	0.989	0.990
Observations	10,610	15,456	36,774	55,948
R-squared	0.706	0.706	0.484	0.456
Controls	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Locality FE	Y	Y	Y	Y
Muni-Cohort FE	Y	Y	Y	Y
Clusters	Localities	Localities	Localities	Localities
Localities	512	518	2087	2103
F-stat	10.82	6.917	1.012	0.640
R2 w/in adj.	0.00272	0.00204	-3.60e-06	-1.98e-05

Standard errors clustered by localities in parentheses, t-stats in brackets. All regressions restrict to individuals born between 1890 and 1924 residing in their state of birth in 1930. The left out group in all columns is the 1890-1908 cohort. Localities are restricted to those in municipalities with at least one rural school, excluding municipal capitals, cities, and train stations, keeping localities with populations no greater than 3000 in 1930. All regressions control for gender/sex and disability status, as well as locality, cohort, and municipality-by-cohort. Odd numbered columns show estimates for fully exposed cohorts only, while even columns include an additional interaction for partially exposed cohorts. Columns 1 and 2 restrict to indigenous localities, while Columns 3 and 4 restrict to non-indigenous localities.

To visually check whether pre-trends may threaten my identification, Figure 4 and Figure 5 plot coefficients from my dynamic specification for the male and female samples, respectively. Figure 5a suggests that the significant decrease in Spanish fluency in Table 9 may be due to noise, as one coefficient jumps up relative to the rest without a clear pattern. The rest of the plots are consistent with my previous findings: no evidence of significant pre-trends and increases in Spanish fluency for younger cohorts in indigenous localities, and no effect for non-indigenous populations.

My findings in this section suggest that rural schools in 1920s Mexico induced language homogenization in indigenous areas by stimulating the adoption of Spanish, one of the key nation-building goals of the post-Revolutionary government. I next assess whether these schools may have simultaneously led to a reduction in the use of indigenous languages.

Indigenous Languages. Table 10 and Table 11 present my estimates for the effect of the federal rural school expansion on the propensity for children to speak any indigenous language in 1930. In Columns 1 and 2 of Table 10, I find that girls exposed

Table 8: Results: Male Spanish Fluency

	(1)	(2)	(3)	(4)
	Respondent Speaks Spanish			
VARIABLES	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) ×	0.0834***	0.0836***	-0.00174	-0.00148
Rural School	(0.0291)	(0.0295)	(0.00210)	(0.00211)
	[2.868]	[2.836]	[-0.829]	[-0.701]
Born 1909-1919 (Partial) ×		0.0266		-0.00445**
Rural School		(0.0173)		(0.00179)
		[1.537]		[-2.487]
Outcome Mean	0.545	0.553	0.991	0.991
Observations	10,619	15,111	35,919	55,341
R-squared	0.677	0.671	0.408	0.405
Controls	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Locality FE	Y	Y	Y	Y
Muni-Cohort FE	Y	Y	Y	Y
Clusters	Localities	Localities	Localities	Localities
Localities	512	516	2088	2102
F-stat	8.227	4.179	0.688	3.098
R2 w/in adj.	0.00208	0.00140	-2.00e-05	4.34e-05

Standard errors clustered by localities in parentheses, t-stats in brackets. All regressions restrict to individuals born between 1890 and 1924 residing in their state of birth in 1930. The left out group in all columns is the 1890-1908 cohort. Localities are restricted to those in municipalities with at least one rural school, excluding municipal capitals, cities, and train stations, keeping localities with populations no greater than 3000 in 1930. All regressions control for gender/sex and disability status, as well as locality, cohort, and municipality-by-cohort. Odd numbered columns show estimates for fully exposed cohorts only, while even columns include an additional interaction for partially exposed cohorts. Columns 1 and 2 restrict to indigenous localities, while Columns 3 and 4 restrict to non-indigenous localities.

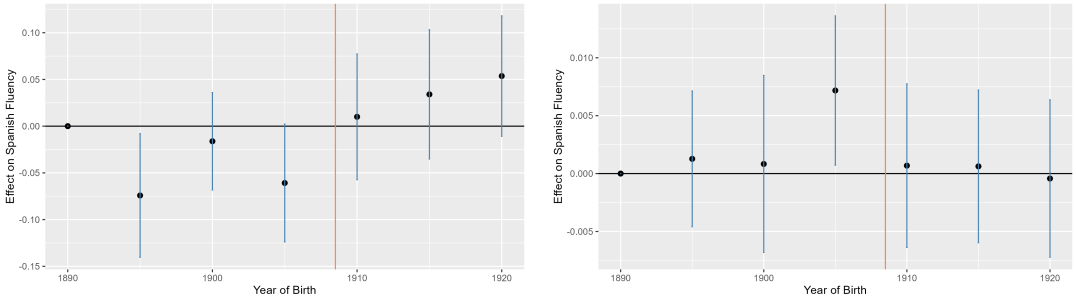
to rural schools in indigenous localities were about 4% less likely to report speaking any indigenous language. Relative to an indigenous-language fluency rate in the sample mean of 90%, this suggests slightly more than a 4% reduction in the propensity of a child to speak an indigenous language. Surprisingly, Table 11 does not shows a significant effect of schools on the likelihood of a boy speaking an indigenous language, despite similar rates of indigenous-language fluency for both males and females. This might be rationalized by considering that the relative effects of school exposure on other outcomes generally appear to be bigger for girls than for boys. In both tables, Columns 3 and 4 show no evidence of any change in indigenous language in non-indigenous localities.

Table 12 formally checks for pre-trends through a placebo check, and does not find evidence of a negative and significant change in any of the three columns. In Column 3, males again show evidence of pre-trends opposite to the direction that would raise identification concerns. Moreover, given that males do not show significant decreases in their propensity to speak indigenous languages, this result does not directly threaten my findings on the loss of indigenous-language fluency.

Table 9: Placebo Check: Spanish

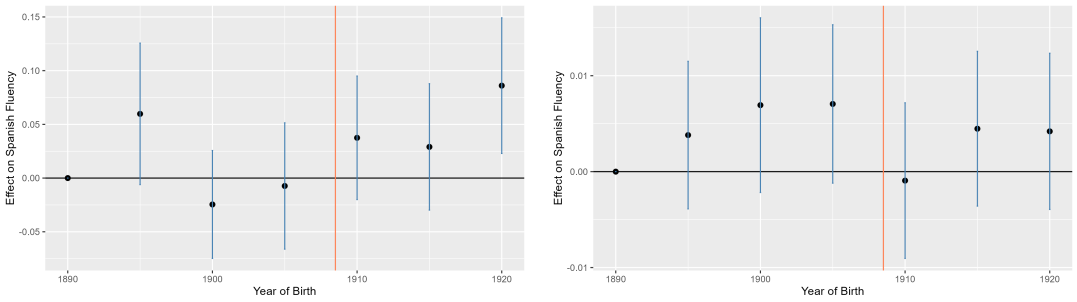
	(1)	(2)	(3)
	Respondent Speaks Spanish		
VARIABLES	Pooled	Female	Male
Born After 1900 ×	-0.0251*	-0.0110	-0.0402**
Rural School	(0.0144)	(0.0203)	(0.0195)
	[-1.748]	[-0.540]	[-2.064]
Outcome Mean	0.488	0.429	0.541
Observations	14,866	7,192	6,944
R-squared	0.678	0.736	0.733
Controls	Y	Y	Y
Cohort FE	Y	Y	Y
Locality FE	Y	Y	Y
Muni-Cohort FE	Y	Y	Y
Clusters	Localities	Localities	Localities
Localities	519	506	505
F-stat	3.055	0.292	4.258
R2 w/in adj.	0.000137	-0.000142	0.000418

Standard errors clustered by localities in parentheses, t-stats in brackets. All regressions restrict to individuals born between 1890 and 1924 residing in their state of birth in 1930. The left out group in all columns is the 1890-1908 cohort. Localities are restricted to those in municipalities with at least one rural school, excluding municipal capitals, cities, and train stations, keeping localities with populations no greater than 3000 in 1930. All regressions control for gender/sex and disability status, as well as locality, cohort, and municipality-by-cohort. Odd numbered columns show estimates for fully exposed cohorts only, while even columns include an additional interaction for partially exposed cohorts. Columns 1 and 2 restrict to indigenous localities, while Columns 3 and 4 restrict to non-indigenous localities.



(a) Indigenous Localities **(b)** Non-Indigenous Localities

Figure 4: Dynamic DiD: Effect of Rural Schools on Spanish, Females



(a) Indigenous Localities **(b)** Non-Indigenous Localities

Figure 5: Dynamic DiD: Effect of Rural Schools on Spanish, Males

Plots for the coefficients in my dynamic specification may shed light on the different results for males and females in indigenous localities. In Figure 7a, coefficients appear to move up then down in a smooth pattern, suggesting that any estimates of the effects of rural schools on males' propensity to speak an indigenous language are due primarily to noise. In other words, this plot confirms that there is no effect on males. For the female sample, Figure 6a is consistent with reductions in the use of indigenous language due to exposure to schools, with no evidence of significant pre-trends. Figure 6b and Figure 7b again show there is no evidence of an effect on non-indigenous populations for language.

While the magnitudes of these effects are much smaller in relative terms compared the effects of rural schools on for literacy and Spanish fluency, local indigenous languages likely continued to be important for most children in majority-indigenous localities. In additional analyses not presented here, I define my outcome as an indicator for speaking both Spanish and an indigenous language. I find significant results, suggesting that the same individuals who stopped speaking native languages also learned Spanish. Even for individuals who learned Spanish, however, it is not obvious that they would either forget their native language or that their parents would not pass it on. The fact that some children did not gain ability to communicate with their broader community may signal the important role of the federal rural schooling program.[25]

Taken together, my findings suggest that the post-Revolutionary Mexican government accomplished the main goals of its nation-building program through the federal rural education. Literacy increased in all treated localities whereas children exposed to rural schools in indigenous-speaking locations were more likely to culturally assimilate through language. Not only did were these children more likely to adopt Spanish, but they were also less likely to know any indigenous language.

[25]Due to data limitations, I cannot check whether this effect holds only for children whose parents speak Spanish or another non-indigenous language. This is because in my data, children and young adults appear to start forming their own households at age fifteen and older. Given that my fully unexposed sample was born no later than 1908, and thus would been twenty-two years of age in 1930, it is uncommon for them to live with other adults old enough to potentially be either their parents or grandparents.

Table 10: Results: Female Indigenous Fluency

VARIABLES	(1)	(2)	(3)	(4)
	Respondent Speaks an Indigenous Language			
	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) × Rural School	-0.0425**	-0.0448**	0.000817	0.00107
	(0.0175)	(0.0175)	(0.00285)	(0.00278)
	[-2.427]	[-2.557]	[0.287]	[0.383]
Born 1909-1919 (Partial) × Rural School		-0.0279*		-0.00132
		(0.0145)		(0.00280)
		[-1.933]		[-0.473]
Outcome Mean	0.908	0.907	0.0197	0.0186
Observations	10,610	15,456	36,774	55,948
R-squared	0.463	0.469	0.529	0.506
Controls	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Locality FE	Y	Y	Y	Y
Muni-Cohort FE	Y	Y	Y	Y
Clusters	Localities	Localities	Localities	Localities
Localities	512	518	2087	2103
F-stat	5.892	3.455	0.0821	0.265
R2 w/in adj.	0.000890	0.000712	-3.22e-05	-3.28e-05

Standard errors clustered by localities in parentheses, t-stats in brackets. All regressions restrict to individuals born between 1890 and 1924 residing in their state of birth in 1930. The left out group in all columns is the 1890-1908 cohort. Localities are restricted to those in municipalities with at least one rural school, excluding municipal capitals, cities, and train stations, keeping localities with populations no greater than 3000 in 1930. All regressions control for gender/sex and disability status, as well as locality, cohort, and municipality-by-cohort. Odd numbered columns show estimates for fully exposed cohorts only, while even columns include an additional interaction for partially exposed cohorts. Columns 1 and 2 restrict to indigenous localities, while Columns 3 and 4 restrict to non-indigenous localities.

5 Long-Run Analysis

The previous section showed that, in the short run, Mexico's rapid expansion in rural schooling during the 1920s was largely successful: it promoted literacy and language homogenization, particularly among indigenous communities. However, these schools were not meant to be equivalent to complete primary schools and likely did not prepare children for secondary education. It is therefore not obvious that these schools would have successfully expanded human capital beyond literacy and language. In this section, I explore whether cohorts exposed to these schools experienced human capital gains beyond literacy. Because limited information on language is available in the 1960 IPUMS census sample, I focus the analysis on measures of educational attainment rather than language homogenization.

Table 11: Results: Male Indigenous Fluency

VARIABLES	(1)	(2)	(3)	(4)
	Respondent Speaks an Indigenous Language			
	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) ×	-0.0226	-0.0243	0.00513	0.00487
Rural School	(0.0175)	(0.0172)	(0.00314)	(0.00304)
	[-1.288]	[-1.414]	[1.637]	[1.600]
Born 1909-1919 (Partial) ×		-0.00346		0.00515
Rural School		(0.0128)		(0.00331)
		[-0.271]		[1.558]
Outcome Mean	0.906	0.903	0.0208	0.0192
Observations	10,619	15,111	35,919	55,341
R-squared	0.495	0.502	0.544	0.516
Controls	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Locality FE	Y	Y	Y	Y
Muni-Cohort FE	Y	Y	Y	Y
Clusters	Localities	Localities	Localities	Localities
Localities	512	516	2088	2102
F-stat	1.658	1.195	2.681	1.503
R2 w/in adj.	0.000179	9.10e-05	4.34e-05	3.58e-05

Standard errors clustered by localities in parentheses, t-stats in brackets. All regressions restrict to individuals born between 1890 and 1924 residing in their state of birth in 1930. The left out group in all columns is the 1890-1908 cohort. Localities are restricted to those in municipalities with at least one rural school, excluding municipal capitals, cities, and train stations, keeping localities with populations no greater than 3000 in 1930. All regressions control for gender/sex and disability status, as well as locality, cohort, and municipality-by-cohort. Odd numbered columns show estimates for fully exposed cohorts only, while even columns include an additional interaction for partially exposed cohorts. Columns 1 and 2 restrict to indigenous localities, while Columns 3 and 4 restrict to non-indigenous localities.

5.1 Empirical Specification

As in the short-run, I use a difference-in-differences analysis to account for potential selection in the allocation of schools across space. I use similar cut-offs for cohort comparisons as before, except I now extend my young cohort to individuals born as late as 1940, as they would be at least 20 years old in 1960. The main difference in this section is the granularity of an individual's geographic location. Whereas the 1930 data contains information on a person's locality of residence, the 1960 data only states a person's municipality of residence. My specification and treatment must therefore be redefined for a municipality-level analysis. To measure cross-sectional exposure to rural schools for each municipality, I take the sum of the number of localities in each municipality with at least one rural school in either 1925 or 1926. I then divide this by the population of the respective municipality in 1930 and multiply by 10,000. The treatment variable can now be interpreted as the approximate number of federal rural schools per 10,000 people in that municipality in 1930. Then Equation 1 can be

Table 12: Placebo Check: Indigenous

	(1)	(2)	(3)
	Speaks an Indigenous Language		
VARIABLES	Pooled	Female	Male
Born After 1900 × Rural School	0.0128	0.000337	0.0339**
	(0.00950)	(0.0157)	(0.0135)
	[1.342]	[0.0214]	[2.515]
Outcome Mean	0.925	0.926	0.927
Observations	14,866	7,192	6,944
R-squared	0.392	0.439	0.477
Controls	Y	Y	Y
Cohort FE	Y	Y	Y
Locality FE	Y	Y	Y
Muni-Cohort FE	Y	Y	Y
Clusters	Localities	Localities	Localities
Localities	519	506	505
F-stat	1.801	0.000459	6.325
R2 w/in adj.	2.64e-05	-0.000185	0.000617

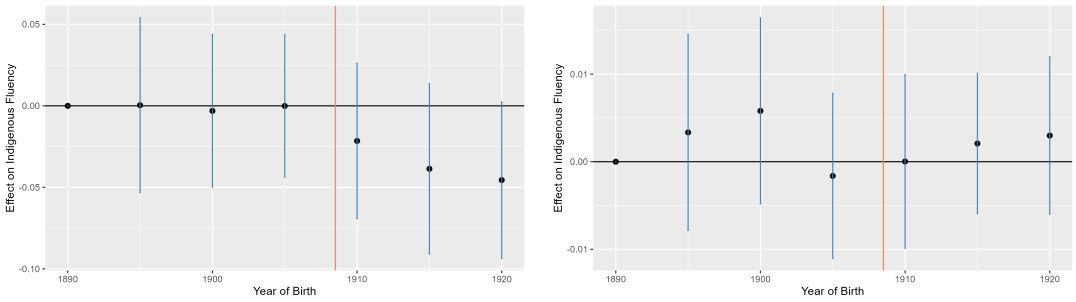
Standard errors clustered by localities in parentheses, t-stats in brackets. All regressions restrict to individuals born between 1890 and 1924 residing in their state of birth in 1930. The left out group in all columns is the 1890-1908 cohort. Localities are restricted to those in municipalities with at least one rural school, excluding municipal capitals, cities, and train stations, keeping localities with populations no greater than 3000 in 1930. All regressions control for gender/sex and disability status, as well as locality, cohort, and municipality-by-cohort. Odd numbered columns show estimates for fully exposed cohorts only, while even columns include an additional interaction for partially exposed cohorts. Columns 1 and 2 restrict to indigenous localities, while Columns 3 and 4 restrict to non-indigenous localities.

redefined as follows:

$$Y_{imt} = \alpha + \beta \text{Young}_t \times \text{School per 10k}_m + \psi X_i + \gamma_t + \delta_m + \phi_{st} + \epsilon_{imt} \tag{4}$$

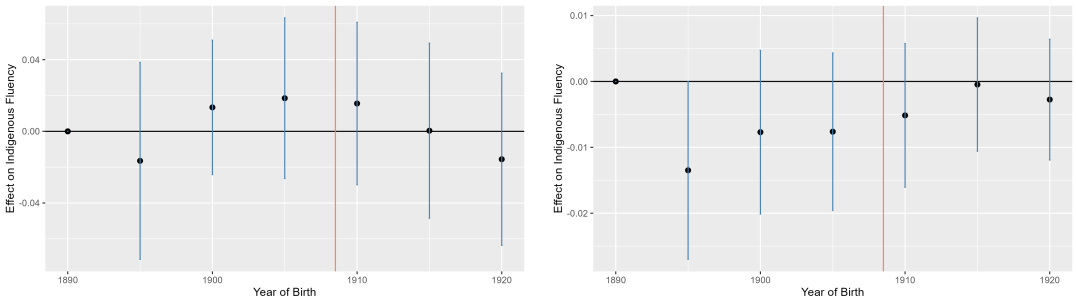
where a literacy or language indicator Y_{itm} is defined for individual i born in year t in municipality m in 1930. $Young_t$ is equal to one for individuals born after the threshold birth year (1920–1940) and zero otherwise. The main specification excludes partially treated cohorts (born between 1909 and 1919), such that the comparison cohorts were born between 1890 and 1908. School per $10k_m$ is the number of per rural schools per 10,000 people in 1930. The main effects are absorbed by the fixed-effects described below. X_i controls for a person's sex/gender.

I also include a set of fixed-effects to account for unobserved individual and location characteristics. γ_t is a cohort fixed-effect, which controls for differences between cohorts that do not vary with locality. If, for instance, older cohorts have more years of schooling and higher literacy rates than younger cohorts, cohort fixed-effects will account for these differences.



(a) Indigenous Localities **(b)** Non-Indigenous Localities

Figure 6: Dynamic DiD: Effect of Rural Schools on Indigenous, Females



(a) Indigenous Localities **(b)** Non-Indigenous Localities

Figure 7: Dynamic DiD: Effect of Rural Schools on Indigenous, Males

As discussed earlier, there may be either positive or negative selection in the allocation of rural schools. To the extent that selection is based on time-invariant characteristics of a location, δ_m, a set of municipality fixed-effects, will account for them. Because exposure to treatment is defined by municipality-cohort pairs, municipality-by-cohort fixed-effects would absorb my identifying variation. However, I can interact fixed effects for state s with cohort fixed-effects to control for state-level time-varying characteristics, ϕ_{st}. Finally, to account for correlated shocks across individuals in the same location, and since my treatment variable is defined at the municipality level, I cluster standard errors by municipality.

The coefficient of interest in this regression is β, which will give the effect on the outcome Y_{itm} of being exposed to a rural school by age six if my identification assumptions hold.

All other terms are defined as in Equation 1. To check graphically for differential effects between cohorts and potential pre-trends, I additionally estimate a dynamic difference-in-difference specification interacting treatment with cohort indicators:

$$Y_{imt} = \alpha + \sum_t \beta_t \text{School per 10k}_m \times \mathbb{1}_t + \psi X_i + \gamma_t + \delta_m + \phi_{mt} + \epsilon_{imt} \tag{5}$$

Because of the fuzziness of the treatment cutoff, in my dynamic plots below, I use the oldest cohort as the baseline and expect significant changes only for cohorts born after 1908. I run specifications grouping cohorts into bins of five years.

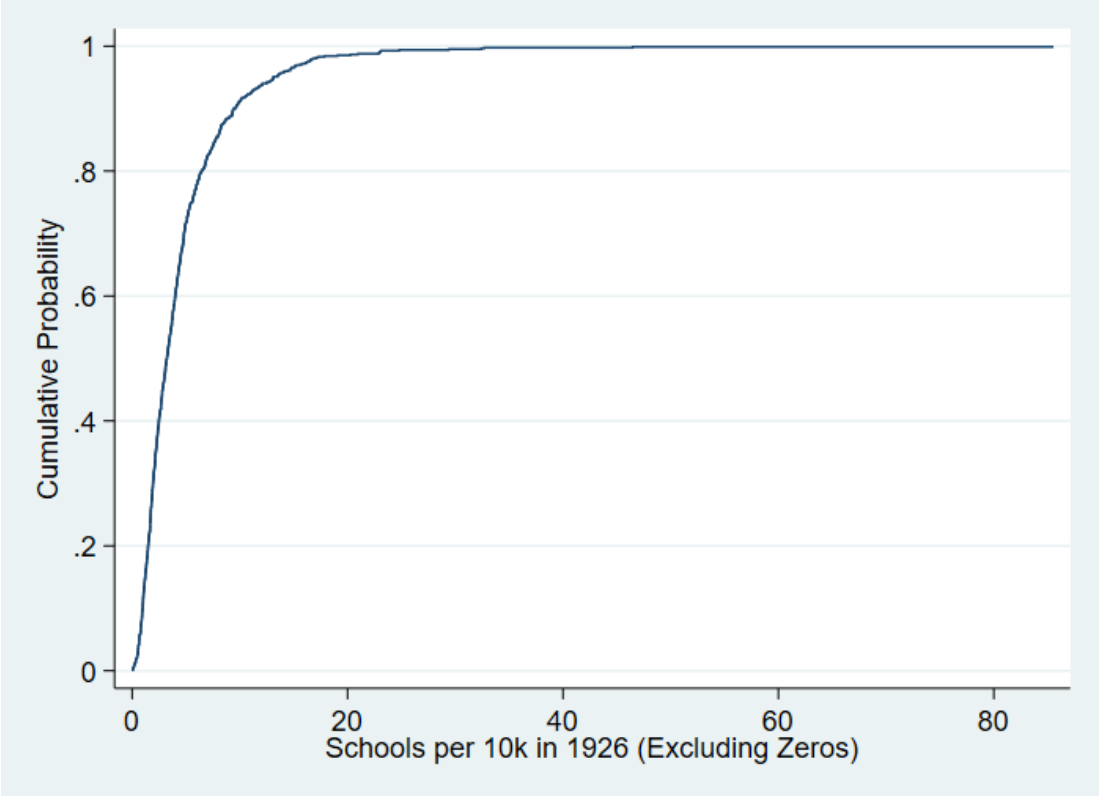


Figure 8: *CDF of Schools per Capita for Munis. with Schools*

In practice, instead of using the estimated number schools per 10,000 individuals in the 1920s, an indicator for a municipality being above or below the median number of schools per capita will offer a more simple interpretation while also increasing statistical power. Because the majority of municipalities had no federal rural schools in this period, the true median would be 0. Instead, I define the median based on the distribution of schools per capita among municipalities with at least one rural school. The median in this distribution, which is plotted as a CDF in Figure 8, is just above 3 schools per 10,000 individuals. I therefore create an indicator for whether a given municipality had at least three schools per 10,000, which I use to refer to high-exposure municipalities or municipalities with schools per capita above the median.

5.2 Migration

As alluded to earlier, the census data have important empirical limitations. First, the census collects data on a person's state of birth, but not their municipality of birth, a reasonable proxy for the municipality in which a person would have been educated Duflo (2001). Excluding individuals born outside their state of residence, the best proxy for municipality of birth would therefore be municipality of residence. This assumption may be plausible in 1930, but given the rural-urban migration associated with increasing growth rates after the 1940s, municipality of residence may no longer be a good proxy for municipality of birth by 1960. If migration was random, for example, differences in outcomes between places with more and fewer schools would be attenuated toward

zero. If migration is selective, such that education stimulated rural-urban out-migration, my estimates would be biased downward. The following results must therefore be interpreted with caution. While I cannot fully account for the full effects of migration, I do exclude individuals who resided outside of their state of birth in 1960.

5.3 Results

In this section, I present long-run estimates of the effect of the 1920s rural school program on educational outcomes. The first subsection presents results using microdata the 1960 and 1970 censuses, in which I can observe a person's municipality of residence in the given year, but not place of birth. In the second subsection, I report additional analyses and descriptive statistics to assess the plausibility of migration and mortality bias in my estimates.

In Table 13, Panel A presents results from my difference-in-differences specification for literacy, partial primary schooling, primary school completion, and years of schooling in 1960. First considering Columns 1 and 2, exposure to the federal rural school program in the 1920s and 1930s appears to have significantly increased both literacy rates and the probability of having only one to five years of schooling, which would mean some but incomplete primary education. The latter suggests that individuals exposed to rural schools were less likely to be completely uneducated, but were also less likely to advance further in their education.

Although a positive effect on literacy is consistent with my previous findings in **?**, both the absolute and relative magnitudes are much lower. Notably, mean literacy rates appear to be higher in 1960 than in 1930. This can be partly attributed to the inclusion of younger cohorts than in **?**, as well as to the fact that I cannot distinguish between localities that would have been rural or urban in 1930. The attenuation of the effect may also be due to a broad increase in the national literacy rate over time. As more schools were established over time, adult literacy classes may have become more accessible in municipalities with no schools in 1926. Moreover, urbanization and exposure to formal markets may have incentivized illiterate adults to learn basic reading and numeracy skills on their own.

In Columns 3 and 4, I instead find a negative effect on educational attainment: exposure to rural schools appears to have reduced primary school completion and the number of years in school, though the later effect is noisy. This may seem counterintuitive for a school expansion program, but the effects are consistent with Column 2. A possible explanation for the simulatneous increase in basic education and reduction overall education may be that a high policy emphasis on literacy may have reduced further investments in education.

Government reports in the 1920s made it clear that increasing literacy among adults was a key education policy goal for the government (Secretaría de Educación Pública, 1926b). There was also some skepticism that the curricula of urban schools were relevant to rural communities. Thus beyond literary, basic numeracy, and civics

Table 13: Education in 1960

VARIABLES	(1) Literacy	(2) Some Primary Only	(3) Primary Complete	(4) Years of Schooling
Panel A: Main Specification				
Above Med. Rural Sch. Per 10k in 1926 ×	0.0235**	0.0235**	-0.0160***	-0.0829*
Young Cohort (YOB 1920-1939)	(0.0103)	(0.0105)	(0.00561)	(0.0481)
	[2.278]	[2.238]	[-2.852]	[-1.726]
Outcome Mean	0.599	0.411	0.118	1.962
Observations	7,306,886	7,509,695	7,493,950	7,499,712
Municipalities	1832	1832	1832	1832
Panel B: Indigenous Muni. Interaction				
Above Med. Rural Sch. Per 10k in 1926 ×	0.0262**	0.0309***	-0.0208***	-0.100*
Young Cohort (YOB 1920-1939)	(0.0113)	(0.0115)	(0.00614)	(0.0518)
	[2.326]	[2.680]	[-3.383]	[-1.937]
Above Med. Sch × Young ×	-0.0191	-0.0566**	0.0414***	0.177
Majority Indigenous 1930	(0.0289)	(0.0274)	(0.0133)	(0.128)
	[-0.660]	[-2.067]	[3.122]	[1.387]
Outcome Mean	0.599	0.411	0.118	1.962
Observations	7,306,886	7,509,695	7,493,950	7,499,712
Municipalities	1832	1832	1832	1832

Standard errors clustered by municipality are in parentheses, t-statistics are in brackets. All regressions include birth cohort fixed effects, state-by-cohort fixed effects, municipality fixed effects, and control for sex/gender. Samples are restricted to individuals living in their state of birth in 1960, born between 1890 and 1939, excluding those born between 1909 and 1919. Observations from the states of Baja California, Baja California Sur, Quintana Roo, Sonora, Yucatan, and Mexico City are excluded. The outcome in Column 1 is an indicator for whether a person can read and write; Column 2 is an indicator for having only between 1 and 5 years of schooling; Column 3 is an indicator for having completed primary schooling or a higher grade level; and Column 4 uses the reported number of years spent in school. In both panels, outcomes are regressed on an interaction between an indicator equal to one if a municipality had at least 3 rural schools per 10,000 individuals in 1926 and an indicator for individuals born between 1920 and 1939. In Panel B, regressions also include a triple interaction of schools per capita, being part of the young cohort, and an indicator for whether the municipality was majority indigenous by language in 1930.

education, the rural school curricula focused most on hygiene, agriculture, and the manual manufacture of locally useful goods. Moreover, the typical rural school was not meant to offer complete primary education, usually offering between two to four years of material. Consequently, the limitations of these schools may not have prepared students to transition to full primary schools, much less advance to higher grades. Additionally, anecdotal reports of the time complained of the low supply of qualified teachers, who in some cases only received brief training before being assigned to a school. Teachers had their own complaints, as some frequently went without pay from the federal government and often lacked teaching materials. Another possibility may be that

as the government continued to aggressively expand the education system, communities who initially established lower quality schools may not have been prioritized for further development of local schools, though more qualitative evidence is needed to support this last explanation.

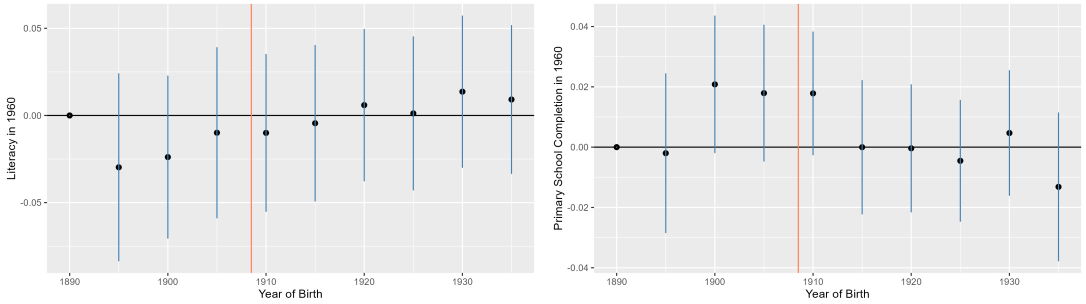
In Panel B of Table 13, I run a similar specification, now including triple interaction for municipalities with a majority of indigenous language speakers in 1930. The main effects in all four columns are quantitatively similar to the main effects in Panel A. However, the triple interactions for Columns 2 and 3, for incomplete primary and complete primary education, appear to flip the sign of the overall effect in indigenous municipalities. Specifically, exposed cohorts in these municipalities are significantly more likely to have completed primary schooling and less likely to have an incomplete primary education. The point estimates for years of schooling follows a similar pattern, but are noisy and not significant at the 5% level.

A possible explanation for this reversal may be that indigenous communities had such low access to education to begin with, that even small, low-quality schools may have allowed exposed individuals to complete primary schooling. It may also be that to counter the perceived educational deprivation of indigenous communities, the government may have made more significant investments in schools in these areas. More qualitative evidence is needed to substantiate these possibilities, however.

I also plot the coefficients from a dynamic specification, in which birth cohorts are grouped into five year ranges, with individuals born between 1890 and 1894 as the baseline cohort. Points before the salmon-colored line indicate birth cohorts born on or before 1908, the last cohort to reach 14 years of age before 1922. Data from 1930 shows that school attendance rates dropped dramatically after 14 (**?**), whereas 1922 is the first year of the rural school program. Then cohorts before the line can be considered to be the older, unexposed group, while cohorts born after 1908 could have been exposed.

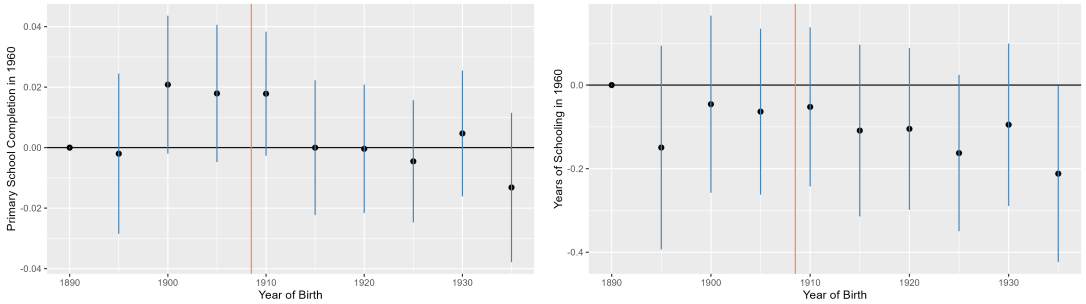
These coefficients are noisy and almost every one of them is statistically insignificant. Figure 9a is visually consistent with a positive effect on literacy in 1960 when comparing oler and younger cohorts. However, there appears to be a slight upward trend among the older cohorts, and the coefficients for the latter cohorts are of about the same size as the baseline. The remaining plots are consistent with either no effect or a negative effect on incomplete primary, primary completion, and years of schooling. This includes **??**, which in my regression tables showed a positive effect. The lack of precision in my dynamic specifications, however, limits what can be said about temporal heterogeneity in the effects and does not allow for a reliable visual check of the parallel trends assumption.

Overall, these results suggest an ambiguous and contradictoy effect of rural eduation in the long-run. For the most part, the government was able to achieve its goal of increasing literacy in the short and long-run through the rapid expansion of rural schools in the 1920s. However, possibly due to the low quality of the schools, it appears that the beneficiaries of these schools were not able to translate access into further educational gains. Nonetheless, when interpreting these results, it is important to keep in mind what the implied counterfactual is. If in a counterfactual world, treated communities would



(a) *Literacy* **(b)** *Incomplete Primary School*

Figure 9: Dynamic Plots in 1960



(a) *Primary School Completion* **(b)** *Years of Schooling*

Figure 10: Dynamic Plots of Effects on Primary School and Years of Schooling

not have had access to schools for many years, then even if the rural school program did not promote higher levels of education, literacy and a couple years in school may have still been beneficial relative to no education at all in an increasingly urbanized Mexico. On the other hand, it may be that in the absence of a rural school, treated locations would have had to wait only a bit longer to receive a better school, possibly with a full primary education curriculum. In this case, a small, initial investment may have precluded later, larger investments, negatively affecting education overall. With the current data, however, it is difficult to assess which of these scenarios represents the most plausible counterfactual. Qualitative evidence as well as later archival evidence on the expansion of education may offer information on whether once treated, rural communities were thereafter ignored in later education expansions.

6 Conclusion

In this paper, I show that Mexico's first major effort to expand public education increased literacy rates and induced language homogenization, but did not stimulate additional human capital acquisition in the long run, at least compared to locations with fewer rural schools. Through a difference-in-differences specification, I find that, in the short run, rural schools achieved education goals central to the Mexican government's nation-building policy in the 1920s: spreading literacy and Spanish fluency, while reducing the use of indigenous languages. In my long-run analysis, I find that, while literacy

gains associated with exposure to rural schools in the 1920 persists, other measures of educational attainment suggest a broader failure to stimulate growth in human capital. Affected cohorts are more likely to have attended primary school, but have fewer years of schooling and are less likely to have completed primary school. These findings may highlight the potential drawbacks of rapid, low-quality investments in human capital: those exposed to the rural school program were more likely to become literate in the short run, but without additional public investments in education, they appear to have been stuck with incomplete primary education. Although the interpretation of my findings depends on the counterfactual assumed, it appears that at best, the rural school expansion, while increasing literacy, did not on its own set the stage for sustained increases in educational attainment.

References

Alberto Alesina, Bryony Reich, and Alessandro Riboni. *Nation-building, nationalism, and wars*, volume 25. Springer US, 2020. ISBN 0123456789. doi: 10.1007/s10887-020-09182-7. URL https://doi.org/10.1007/s10887-020-09182-7.

Milena Almargo and David Andrés-Cerezo. The Construction of National Identities. 2019.

Benedict Anderson. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*, volume 1984. 1984. doi: 10.3817/0684060227.

Luis Angeles and Aldo Elizalde. Pre-colonial institutions and socioeconomic development: The case of Latin America. *Journal of Development Economics*, 124:22–40, 2017. ISSN 03043878. doi: 10.1016/j.jdeveco.2016.08.006. URL http://dx.doi.org/10.1016/j.jdeveco.2016.08.006.

Lydia Assouad. Charismatic Leaders and Nation-Building. 2021. URL https://www.dropbox.com/s/93e05vg2euksrgk/JMP{_}Assouad{_}PSE.pdf?dl=0.

Samuel Bazzi, Masyhur Hilmy, and Benjamin Marx. Islam and the State: Religious Education in the Age of Mass Schooling. 2021.

Guillaume Blanc and Masahiro Kubo. Schools, Language, and Nations: Evidence From a Natural Experiment in France. 2021.

Alberto Diaz-Cayeros. Indian Identity, Poverty and Colonial Development in Mexico. 2011.

Alberto Diaz-Cayeros and Saumitra Jha. Conquered but not Vanquished: Complementarities and Indigenous Entrepreneurs in the Shadow of Violence. 2022.

Alberto Diaz-Cayeros, Juan Espinosa Balbuena, and Saumitra Jha. Pandemic Spikes and Broken Spears: Indigenous Resilience after the Conquest of Mexico. 2021.

Esther Duflo. Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4):795–813, 2001. ISSN 00028282. doi: 10.1257/aer.91.4.795.

Aldo Elizalde. On the economic effects of Indigenous institutions: Evidence from Mexico. *Journal of Development Economics*, 147(August):102530, 2020. ISSN 03043878. doi: 10.1016/j.jdeveco.2020.102530. URL https://doi.org/10.1016/j.jdeveco.2020.102530.

Aldo Elizalde, Eduardo Hidalgo, and Nayeli Salgado. Public good or public bad? Indigenous institutions and the demand for public goods. 2021.

Raissa Fabregas. Broadcasting Human Capital? The Long-Run Effects of Mexico's Telesecundarias. 2021.

Donna L. Feir. The long-term effects of forcible assimilation policy: The case of Indian boarding schools. *Canadian Journal of Economics*, 49(2):433–480, 2016a. ISSN 15405982. doi: 10.1111/caje.12203.

Donna L. Feir. The intergenerational effects of residential schools on children's educational experiences in Ontario and Canada's Western Provinces. *International Indigenous Policy Journal*, 7(3), 2016b. ISSN 19165781. doi: 10.18584/iipj.2016.7.3.5.

Donna L. Feir and M. Christopher Auld. Indian residential schools: Height and body mass post-1930. *Canadian Journal of Economics*, 54(1):126–163, 2021. ISSN 15405982. doi: 10.1111/caje.12495.

Matthew T. Gregg. The long-term effects of American Indian boarding schools. *Journal of Development Economics*, 130(September 2017):17–32, 2018. ISSN 03043878. doi: 10.1016/j.jdeveco.2017.09.003.

Instituto Nacional de Estadística y Geografía. Quinto Censo de Población 1930. URL `https://www.inegi.org.mx/programas/ccpv/1930/`.

Enrique Krauze, Jean A. Meyer, and Cayetano Reyes García. *La reconstrucción económica*. Colegio de México, México, 1981.

Jean A. Meyer. *The Cristero Rebellion: The Mexican People between Church and State, 1926-1929*. Cambridge University Press, Cambridge, 1976. ISBN 0521210313. doi: 10.2307/1856581.

Minnesota Population Center. Integrated public use microdata series, international: Version 7.3 [dataset]. *IPUMS, 2020*. URL `https://doi.org/10.18128/D020.V7.3`.

Jonathan Roth, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. 2022. URL `http://arxiv.org/abs/2201.01194`.

Mariano Sánchez-Talanquer. Legacies of Revolution: Popular Militias and the Rule of Law. 2018.

Dirección General de Estadística Secretaría de Economía. *Resumen General del Quinto Censo de Población*. México, 1930.

Secretaría de Educación Pública. *Noticia estadística sobre la educación publica en México correspondiente al año de 1925*. México, 1926a.

Secretaría de Educación Pública. *La educación pública en Mexico a traves de los mensajes presidenciales desde la consumacion de la independencia hasta nuestros dias*. México, 1926b.

Secretaría de Educación Pública. *El sistema de escuelas rurales en México*. 1927.

Secretaría de Educación Pública. *El esfuerzo educativo en México: La obra del gobierno federal en el ramo de educación pública durante al administración del presidente Plutarco Elías Calles (1924-1928)*. México, 1928.

Felipe Valencia Caicedo. The Mission: Economic Persistence, Human Capital Transmission and Culture in South America. *The Quarterly Journal ofEconomics*, 134(1): 507–556, 2019. doi: 10.1093/qje/qjy024.Advance.

Mary Kay Vaughan. *The State, Education, and Social Class in Mexico, 1880-1928*. Northern Illinois University Press, DeKalb, Illinois, 1982. ISBN 0-87580-079-3.

Eugen Weber. *Peasants into Frenchmen: The Modernization of Rural France, 1870-1914*, volume 25. Stanford University Press, Stanford, 1978. doi: 10.2307/481200.

Francisco José Zamudio Sánchez, Roxana Ivette Arana Ovalle, Waldenia Cosmes Martínez, Javier Santibáñez Cortés, and Margoth Laredo Rojas. Análisis de los microdatos del censo de 1930: a 80 años del México posrevolucionario. *Realidad, datos y espacio*, 6(3):16–35, 2015. ISSN 2007-2961. URL `www.inegi.org.mx`.

Francisco José Zamudio Sánchez, Waldenia Cosmes Martínez, Roxana Ivette Arana Ovalle, Marco Antonio Andrade Barrera, Geovanni Becerril González, and Luz Judith Rodríguez Esparza. Una estratificación socioeconómica para comparar dos momentos del desarrollo en México: 1930-2010. *Economía Sociedad y Territorio*, xviii:259–289, 2018. ISSN 1405-8421. doi: 10.22136/est0982.

A Appendix

A.1 Appendix Tables

Table 14: Federal Rural School Enrollment, September 1926

State	Enrollment				Adult to Child Enrollment	
	Boys	Girls	Men	Women	Male	Female
Aguascalientes	546	519	230	22	0.42	0.04
Campeche	721	423	279	31	0.39	0.07
Coahuila	1,376	1,326	823	652	0.60	0.49
Colima	1,159	1,098	570	60	0.49	0.05
Chihuahua	1,770	1,342	313	64	0.18	0.05
Chiapas	4,528	1,517	910	50	0.20	0.03
Distrito Federal	74	71	100	42	1.35	0.59
Durango	1,628	1,316	777	87	0.48	0.07
Guanajuato	4,816	3,616	2,324	504	0.48	0.14
Guerrero	4,112	2,343	1,181	107	0.29	0.05
Hidalgo	6,011	3,090	2,363	190	0.39	0.06
Jalisco	6,181	6,188	2,906	1,047	0.47	0.17
México	8,253	5,527	4,367	266	0.53	0.05
Michoacán	5,574	5,094	2,241	483	0.40	0.09
Morelos	1,106	958	542	33	0.49	0.03
Najarit	1,827	1,917	461	139	0.25	0.07
Nuevo León	3,034	2,474	1,141	338	0.38	0.14
Oaxaca	6,768	2,128	1,824	114	0.27	0.05
Puebla	6,615	3,538	2,587	233	0.39	0.07
Querétaro	1,945	1,060	929	50	0.48	0.05
San Luis Potosí	5,948	4,309	3,346	188	0.56	0.04
Sinaloa	1,159	1,207	520	59	0.45	0.05
Sonora	1,211	1,083	324	95	0.27	0.09
Tabasco	578	193	182	17	0.31	0.09
Tamaulipas	765	529	591	31	0.77	0.06
Tlaxcala	1,576	1,471	766	201	0.49	0.14
Veracruz	3,687	937	1,027	50	0.28	0.05
Yucatán	26	30	0	0	0.00	0.00
Zacatecas	2,601	2,779	1,331	89	0.51	0.03
Total	**85,578**	**58,083**	**34,958**	**5,242**	**0.41**	**0.09**

Source: Secretaría de Educación Pública (1927)

Table 15: National Budget 1925-1928, Categories with at least 10 Million Pesos

	1925	1926	1927	1928
Defense	$ 84,778,689.20	$ 84,093,700.80	$ 84,562,996.64	$ 84,523,505.44
Public Debt	$ 84,529,000.00	$ 63,701,690.00	$ 59,927,258.65	$ 32,500,000.00
Treasury/Tax Collection	$ 85,511,744.23	$ 47,717,076.35	$ 30,914,725.58	$ 26,235,958.74
Communications/Transport	$ 31,166,362.30	$ 39,873,584.42	$ 36,685,444.30	$ 36,297,226.60
Public Education	$ 21,970,813.22	$ 26,707,729.47	$ 25,808,764.17	$ 27,014,697.80
Agriculture	$ 13,002,027.20	$ 29,722,471.44	$ 30,394,245.03	$ 31,548,629.70
Manufacturing	$ 13,928,892.30	$ 17,296,252.69	$ 15,353,645.00	$ 13,097,254.57

All values in current Mexican pesos.
Source: Secretaría de Educación Pública (1928)

A.2 Appendix Figures

A.2.1 Distribution of Indigenous Language Speakers

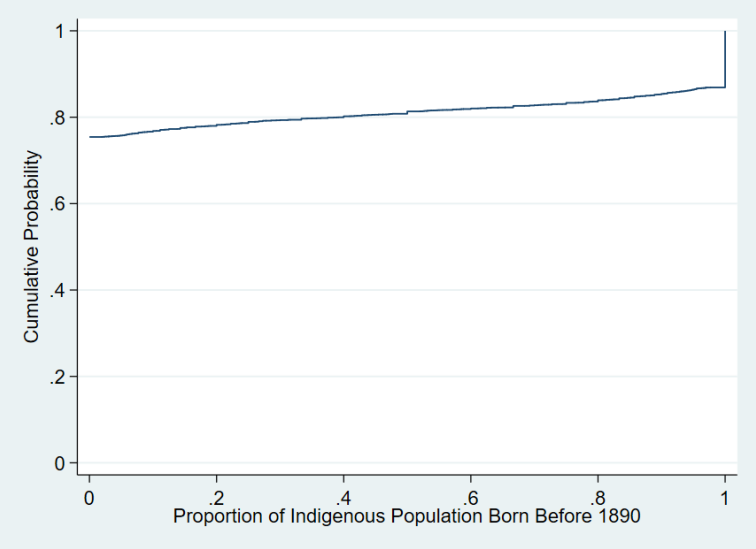


Figure 11: Cumulative Distribution of Older Indigenous Population

A.2.2 Dynamic DiD Plots, 3-Year Bins

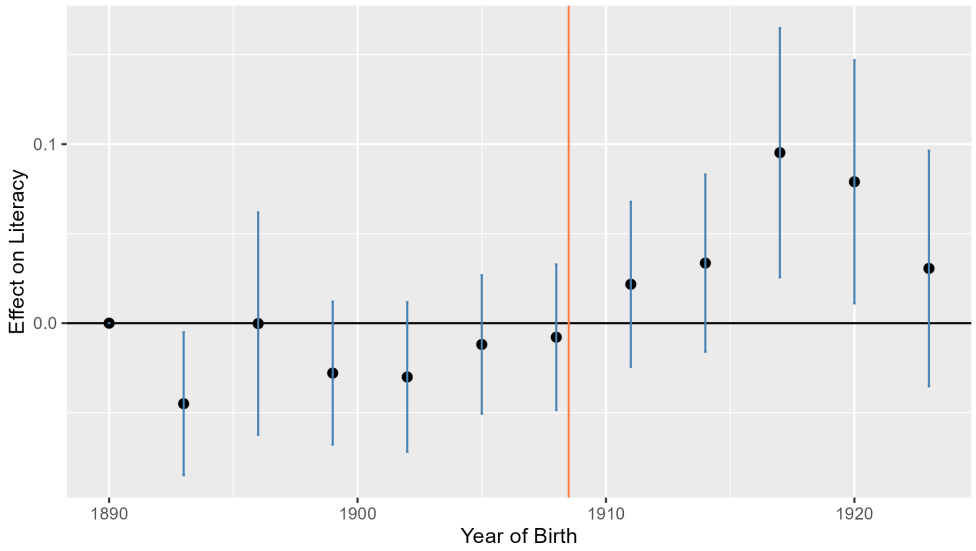


Figure 12: Dynamic DiD: Effect of Rural Schools on Literacy

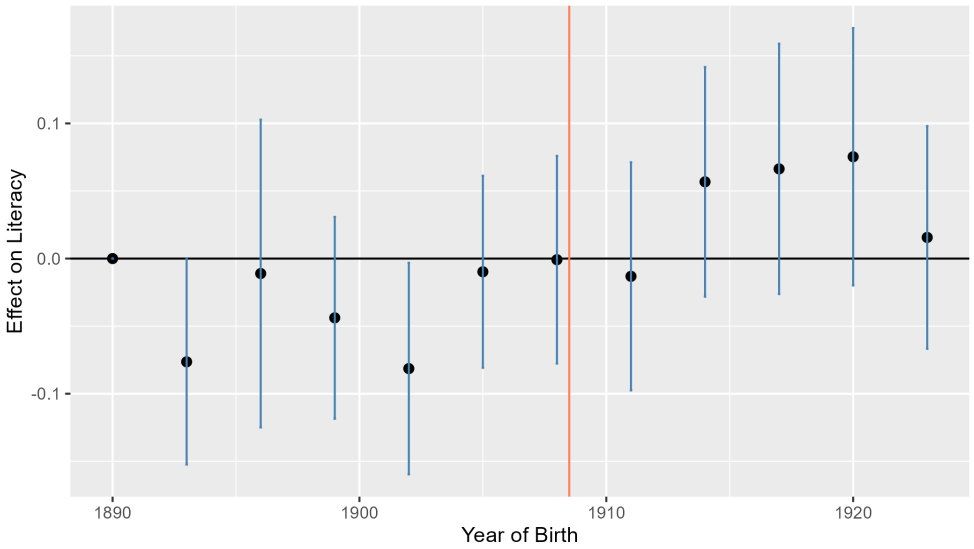


Figure 13: Dynamic DiD: Effect of Rural Schools on Literacy, Males

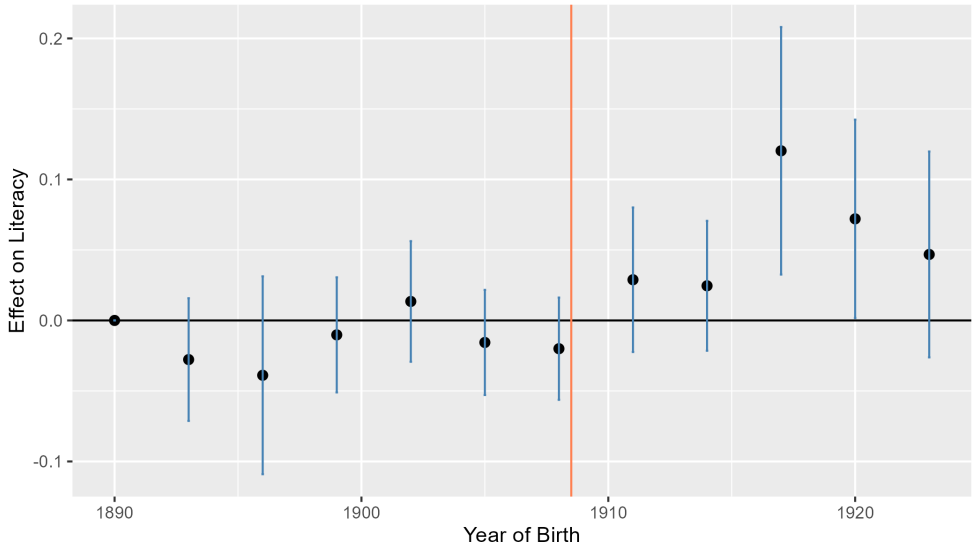


Figure 14: Dynamic DiD: Effect of Rural Schools on Literacy, Females

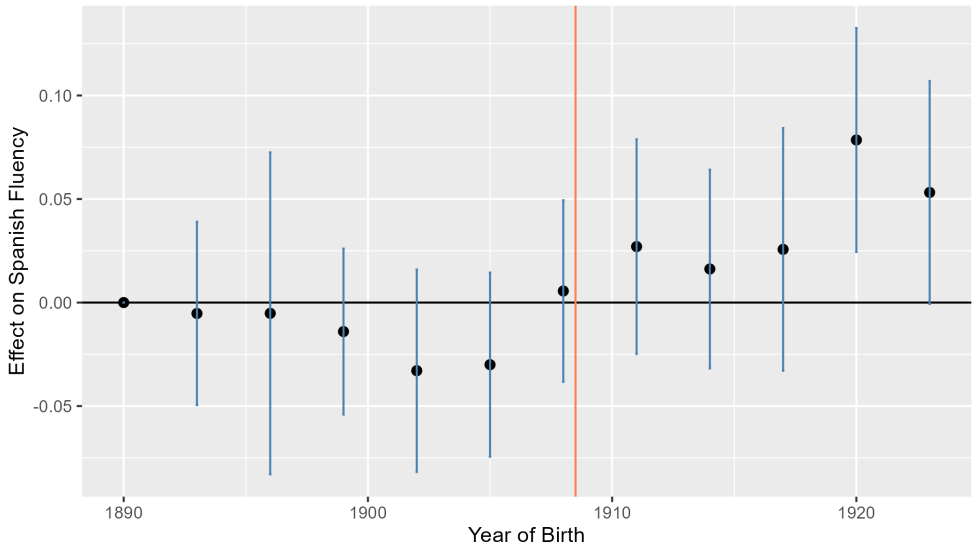


Figure 15: Dynamic DiD: Effect of Rural Schools on Spanish

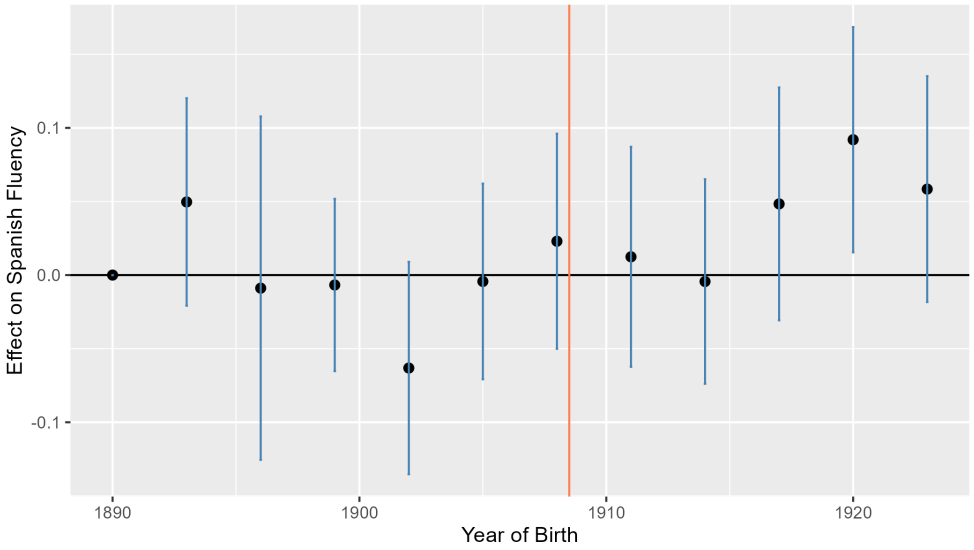


Figure 16: Dynamic DiD: Effect of Rural Schools on Spanish, Males

Figure 17: Dynamic DiD: Effect of Rural Schools on Spanish, Females

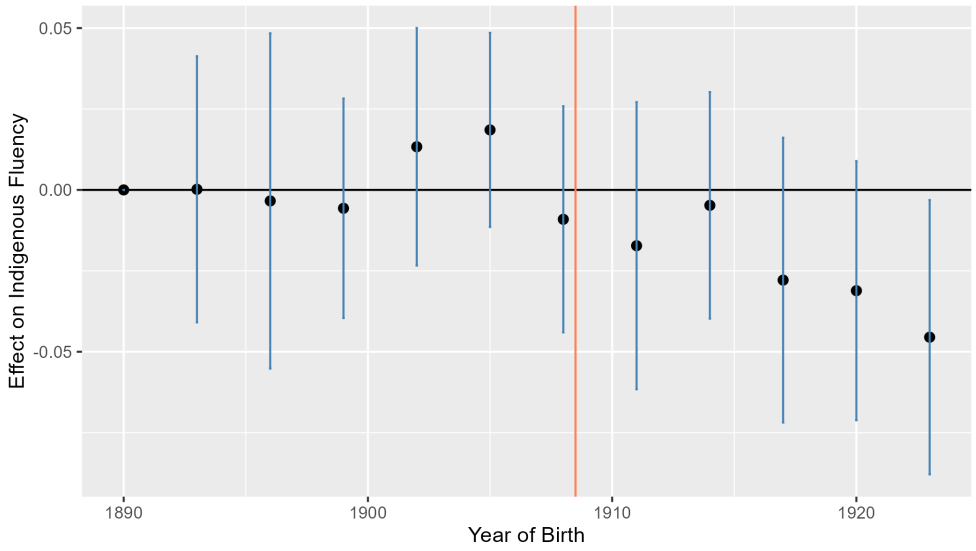


Figure 18: Dynamic DiD: Effect of Rural Schools on Indigenous Languages

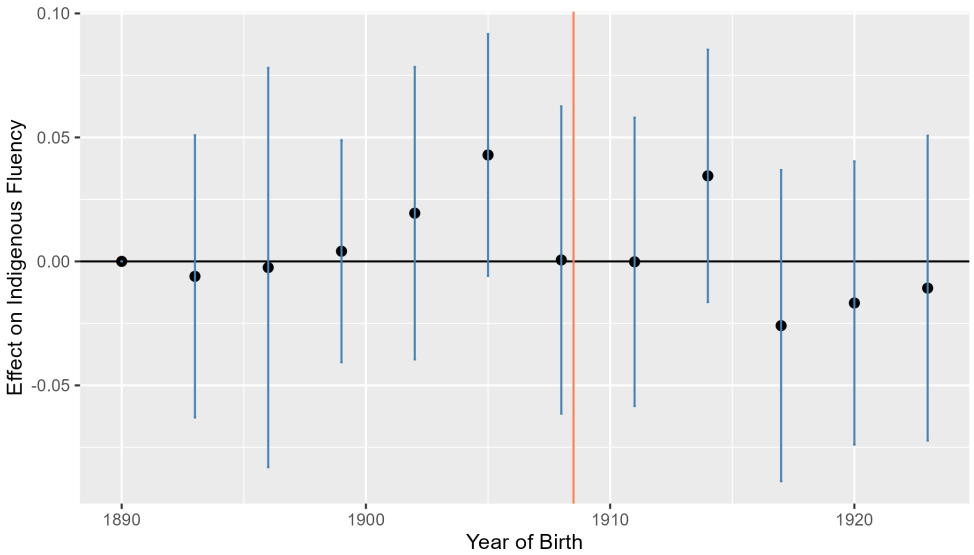


Figure 19: Dynamic DiD: Effect of Rural Schools on Indigenous Languages, Males

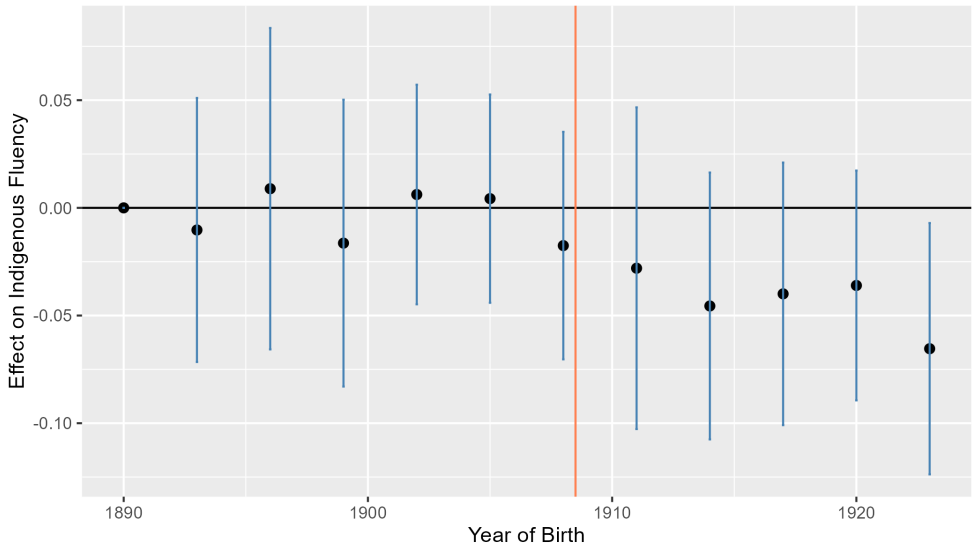


Figure 20: Dynamic DiD: Effect of Rural Schools on Indigenous Languages, Females