Rural Education, Nation-building, and Ethnic Assimilation in Post-Revolutionary Mexico

Ariel Gómez

Version: 12/27/2022
Please click here for the most recent version

Abstract

Can non-democratic regimes nation-build through investments in education? I investigate Mexico's efforts to homogenize its population and legitimize the post-Revolutionary state through a 1920s expansion in rural schooling. I first show that rural schools, which promoted Spanish literacy and fluency, were associated with higher school attendance and literacy rates in 1930. Using a difference-in-differences design, I then find that in predominantly indigenous localities, rural schools induced language homogenization: children were more likely to speak Spanish and less likely to speak indigenous languages in 1930. I then explore the relationship between rural schools and land redistribution, a central demand of the Mexican Revolution. I find that localities with schools were more likely to receive a land grant in this period. This result is consistent with historical evidence that schools facilitated the implementation of land reform. I offer historical examples suggesting that teachers advocated for local land redistribution and other social reforms.

1 Introduction

Non-democratic regimes have often made significant investments in expanding their schooling systems Paglayan (2021, 2022), despite evidence that education may pose challenges for centralized states by increasing political agency (Glaeser et al., 2007). Why then would non-democratic regimes provide public education? A theoretical explanation put forth by recent literature is that by assimilating minority groups and fostering nationalism, education can be an effective nation-building policy (Alesina et al., 2021). Governments have historically used nation-building programs more broadly to stimulate economic development and increase political support (Alesina and La Ferrara, 2005, Weber, 1978). However, whether education expansions are effective in their primary nation-building goals remains an open question.

I investigate the expansion of the federal education system in post-Revolutionary Mexico, which led to the opening of 2,600 rural schools 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 and may have

helped consolidate power in the federal government. Moreover, the government largely established *mestizaje*, or the cultural and racial mixing of indigenous and Spanish heritage, as the basis of a Mexican national identity during this period. I evaluate whether Revolutionary intellectuals and policy makers successfully induced cultural assimilation through language homogenization among indigenous populations. I also explore whether schools and teacher advocacy facilitated the implementation of land redistribution, a policy central in establishing the legitimacy of the newly established, Revolutionary government.

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 language homogenization. To evaluate the impacts of the rural schools 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 that reports whether each individual is literate, speaks Spanish, and speaks any indigenous language. With this data, I account for potential selection in access to rural education with a difference-in-differences design, in which I interact geographic variation in the locations of federal rural schools with cross-sectional variation between cohorts. Through this analysis, I find that the rural education program largely achieved its aims of inducing language homogenization: in predominantly indigenous areas, children exposed to these schools were more likely to be literate and fluent in Spanish and less likely to speak any indigenous language.

In the second part of the paper, I assess the relationship between the presence of schools and land redistribution, a key demand of the Revolution. I use yearly data indicating when each location first experienced land reform and interact it with geographic variation in the locations of schools. I estimate the relationship between these two policies and find a positive correlation between school locations and land redistribution. This result is consistent with the hypothesis that schools facilitated the implementation of land reform, possibly through the role of a new group of educated government representatives in rural communities—teachers. Overall, my findings suggest that schools formed an important part of post-Revolutionary Mexico's nation-building efforts, by inducing language homogenization in indigenous communities, while also playing a role in the implementation of an important Revolutionary policy, land reform.

I mainly contribute to work on nation-building and language homogenization (Alesina et al., 2020, Almargo and Andrés-Cerezo, 2019, Anderson, 1984, Assouad, 2021, Bazzi et al., 2021, Weber, 1978) and on why non-democratic regimes may choose to educate their populations (Alesina et al., 2021, Paglayan, 2021, 2022). My work is most closely related to Blanc and Kubo (2021), who show that state-sponsored education in France led populations to adopt the modern French dialect. While Blanc and Kubo (2021) study the medium and long-run impacts of this policy on the homogenization of related languages, I focus on the contemporary effects of schooling on literacy, the adoption of Spanish, and the loss of native languages among indigenous groups in

Mexico. Given that my short-run impacts are driven by children in 1930, my findings serve as a leading indicator of ethnic assimilation in the twentieth century.[1]

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).[2] 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.

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.[3]

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 results on two sets of outcomes, along with empirical tests for my identification assumptions. I conclude by suggesting future directions for research on education policy in post-Revolutionary Mexico.

2 Historical Context

2.1 The Mexican Revolution and It Aftermath

The rural education program in 1920s Mexico was intimately linked with the Mexican Revolution. The Revolution began in 1910 and escalated into a multi-sided civil war by 1914. Although the capitalist Constitutionalist faction ultimately defeated the more

[1] Based on calculations from census tabulations, I find that between 1930 and 1970, the proportion of individuals who spoke indigenous languages in Mexico fell from 16% to 7.8%.

[2] I also touch on the causes and consequences of land reform in Mexico (Dell, 2012, Elizalde, 2020, Garfias and Sellars, 2021)

[3] 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 historians have suggested that the growth in public secular education was one of the underlying causes of the *Cristero* War from 1926 to 1929 (Meyer, 1976, Vaughan, 1982).

radical Agrarian faction, political pressures led to the inclusion of social reforms in the 1917 constitution, including workers rights, limits on the Catholic Church, and land redistribution (Vaughan, 1982). Of relevance here is Article 3 of the Constitution, which mandated primary schooling, prohibited clerical primary education, and devolved the remaining authority over education to the local level (Vaughan, 1982). In 1920, generals Álvaro Obregón and Plutarco Elias Calles staged a successful coup, and the subsequent election of Obregón as president later that year marked the end of the Revolution.

Even so, the 1920s was a period of continued instability, especially in rural areas.[4] Calles, who was elected president from 1924 to 1928, shared influence and power with Obregón while trying to reconstruct a severely weakened central state. Their centralization efforts led to the foundation of new government bodies to address the demands of radical groups while bringing them under the influence of the federal government (Meyer, 1976). To retain control over the expanding state after his presidency, Calles formalized the co-optation of revolutionary rhetoric through the establishment of the National Revolutionary Party (PNR) in 1929 (Meyer, 1976).[5]

Calles and Obregón both understood that the survival of their regime hinged on the support of the rural population.[6] The most important rural policy at the time was land redistribution, one of the central demands of the Revolution.[7] In the formal process established in 1922, communities first had to organize as a unit and file a formal petition for land. The petition would go through a bureaucratic process, and ultimately the president himself would approve or deny each petition (Krauze et al., 1981). Land was granted to communities in the form an *ejido*,[8] a semi-communal land tenure system in which each land-holder, known as an *ejidatario*, was assigned an individual plot to work, with an additional plot set aside for communal use. *Ejidatarios* only had use-rights over their parcels and could not sell the land or put it up for collateral (De Janvry et al., 2014, 2015). Historians have noted that land reform petitions were selectively approved, carefully avoiding the expropriation of the most powerful landlords (Cárdenas, 2015, Hall, 1980, Krauze et al., 1981).

Calles' influence ended when he forced into exile under Lázaro Cárdenas, who was president from 1934 to 1940. The Cárdenas administration saw the consolidation of the

[4]This included an attempted coup in 1923, the *Cristero* War against Catholic rebels (1926-1929), and the assassination of Obregón in 1928.

[5]This was the precursor of the Institutional Revolutionary Party (PRI), which would control the presidency until 2000.

[6]By the author's calculation based on census tabulations, this was about 66.5% of the population in 1930.

[7]Although Obregón and Calles were ideologically opposed to land reform, in a personal correspondence with John Womack, the historian described as "cold-blooded politicians." Moreover, Krauze et al. (1981) show that Calles clearly understood the risk of another revolution if landless peasants did not gain access to land

[8]A different land tenure system, the *communidad*, was meant mainly for indigenous people whose land was expropriated under the previous regime. If an indigenous community could show proof of expropriation, then they would have the land returned as restitution. Given the difficulty in ascertaining previous land arrangements, the government often allocated land to indigenous people in the form of *ejidos* (Hall, 1980, Krauze et al., 1981). I do not distinguish between these types of land tenure and, instead, refer to both as *ejidos*.

state and a second wave of social and economic reforms, including a massive expansion of land reform (Vaughan, 1982).

2.2 *Mestizaje* and Education Policy

Education policy in the 1920s was spearheaded by supporters of the *indigenista* movement. These revolutionary intellectuals rejected theories of European racial superiority and believed that *mestizaje*, the cultural and racial mixing of indigenous and Spanish heritage, was at the core of Mexico's economic and social potential. *Mestizaje* and cultural assimilation were seen as inevitable and necessary processes to be hastened through education (Knight, 1990). Schooling, they hoped, would encourage indigenous people to adopt Spanish, Western norms and values, and a Mexican, *mestizo*, national identity (Vaughan, 1982).

However, without the support of the central government, as per the 1917 constitution, state and municipal schools began to close under budgetary pressures in 1920. José Vasconcelos, a leader of the *indigenista* movement, advocated for a constitutional amendment in 1921 that returned authority over public education to the federal government. This amendment established a new education department, the *Secretaría de Educación Publica* (SEP).[9] As the director of the SEP from 1921 to 1923, Vasconcelos led a wave of educational and cultural initiatives[10] to promote a his vision of Mexican nationalism.

The nationalist and revolutionary rhetoric of the *indigenistas* provided ideological cover for the leaders of the central state, which supported the SEP's education expansion as part of a broader nation-building effort. Between 1922 and 1926, the SEP established 2,600 rural schools throughout Mexico (Vaughan, 1982).[11] In many places, rural schools allowed the state to establish a presence in places where it had little control, as SEP schoolteachers were often the only nearby representatives of the federal government.

2.3 The Rural School Program

Starting in 1922, the Department of Indigenous Culture, initially seen only as a temporary branch of the SEP, began expanding rural education through cultural missionaries, teachers tasked with locating villages with indigenous populations 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"). As the name

[9]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).

[10]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.

[11]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).

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)

suggests, this new institution was established to teach children to read and serve as a community center. As local government representatives, SEP teachers were expected to be not only educators, but also social workers and community organizers. According to official SEP documents, teachers were to

> "...congregate all community members, even those who for some reason are not on good terms...[such that] friendships [are] strengthened, and enmity and disagreements [fade]...[teachers will] resolve the region's problems..." (Secretaría de Educación Pública, 1927)

In addition to creating a literate populace, these schools were meant to help resolve tensions in rural areas. Thus, these schools could be viewed as part of a broader nation-building effort, meant to incorporate indigenous people into a broader "imagined community" (Anderson, 1984) and reduce the risk of violence or revolt against the central state.

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.[12] Teachers were to harmonize and uplift rural communities while advocating for the central government's benevolence and its expanding social role (Secretaría de Educación Pública, 1927). This was especially important under Calles, who began to enforce the anti-clerical articles of the constitution to displace the Catholic Church as the center of rural life. 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)[13] [emphasis added]

SEP school inspectors also began collecting data on civic activities carried out by teachers and schools, such as vaccination campaigns and community initiatives organized through teacher activism. Based on reports from a federal school inspector, one government official in 1927 wrote about a teacher's role in the community:

> "David Muñoz...is an indefatigable teacher, a man of action and enthusiasm...He has completely organized [the village], such that today there are now streets formed, a small plaza, and a magnificent building functioning as a school...And that's not all. Muñoz organized the fieldworkers and as their head handled the granting of *ejidos* to [the village] and the surrounding areas." (Secretaría de Educación Pública, 1927)

Based on this report, it appears that teachers may have facilitated the implementation of land reform, by helping villagers, many of them illiterate, with the bureaucratic process of petitioning for redistribution. Vaughan (1982) suggests that some rural teachers, influenced by Revolutionary leaders, may have supported social reforms like land redistribution for ideological reasons. With currently available data, it is difficult

[12] Vasconcelos strongly disagreed with Obregón's choice of Calles as his successor and went into temporary exile after 1923.

[13] 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).

to precisely know what role teachers played in the process of land reform. We shall see there that the evidence is consistent with a schools playing a role in the implementation of land reform.

For context, Table 2 reports state-level data on federal rural school attendance in September 1926,[14] showing that a total of 143,661 children[15] were enrolled in rural schools, with an average attendance of 114,015.[16] This can be benchmarked against either the rural population,[17] 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.[18]

Table 2: Federal Rural School Average Attendance, September 1926

State	Average Attendance				Adult to Child Attendance	
	Boys	Girls	Men	Women	Male	Female
Aguascalientes	461	419	173	11	0.38	0.03
Campeche	503	293	186	21	0.37	0.07
Coahuila	1,190	1,146	675	106	0.57	0.09
Colima	875	844	380	45	0.43	0.05
Chihuahua	1,248	1,015	183	46	0.15	0.05
Chiapas	3,731	1,206	616	29	0.17	0.02
Distrito Federal	44	34	100	42	2.27	1.24
Durango	1,443	1,170	609	84	0.42	0.07
Guanajuato	3,392	2,528	1,482	320	0.44	0.13
Guerrero	3,359	1,944	871	92	0.26	0.05
Hidalgo	4,571	2,460	1,591	140	0.35	0.06
Jalisco	5,638	5,536	2,456	926	0.44	0.17
México	6,202	4,131	2,608	189	0.42	0.05
Michoacán	5,108	4,094	1,931	420	0.38	0.10
Morelos	732	694	332	20	0.45	0.03
Najarit	1,429	1,507	355	100	0.25	0.07
Nuevo León	2,357	1,988	953	275	0.40	0.14
Oaxaca	5,363	1,658	1,278	91	0.24	0.05
Puebla	5,060	2,718	1,623	179	0.32	0.07
Querétaro	1,289	772	607	35	0.47	0.05
San Luis Potosí	4,687	3,116	2,253	147	0.48	0.05
Sinaloa	840	921	373	46	0.44	0.05
Sonora	1,079	1,065	298	93	0.28	0.09
Tabasco	428	123	161	17	0.38	0.14
Tamaulipas	696	446	457	23	0.66	0.05
Tlaxcala	1,161	1,031	449	156	0.39	0.15
Veracruz	3,170	803	879	44	0.28	0.05
Yucatán	19	20	0	0	0.00	0.00
Zacatecas	2,117	2,181	1,039	66	0.49	0.03
Total	**68,102**	**45,913**	**24,918**	**37,63**	**0.37**	**0.08**

Source: Secretaría de Educación Pública (1927)

[14] See Table 16 in the Appendix for state-level enrollment data.

[15] These schools also offered night classes for older teenagers and adults, with 40,200 enrolled and an average attendance of 28,681.

[16] Calculated by rounding the ratio of total number of days attended by all students to the total number enrolled.

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

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

3 Rural Schools and Ethnic Assimilation

3.1 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). Based on lists, I have digitized the localities hosting 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.

My main outcomes are literacy, a measure of educational attainment, and fluency in Spanish and indigenous languages, which capture two dimensions of cultural assimilation. The most basic goal of rural schooling was to increase literacy rates, which officials viewed as a pre-requisite 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 education contributed to the loss of native languages.

To test for effects on these outcomes, I use a 10% micro-sample of surviving documents from the 1930 Mexican census, 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,[19] disability status, and state of birth as well as information on literacy, contemporaneous school attendance,[20] language spoken, and occupation. The data also includes information on the state, municipality, and locality[21] 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.[22]

[19]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).
[20]I cannot observe educational attainment or years of schooling as they were not collected in 1930.
[21]These are the three main administrative divisions in Mexico, in descending order. Localities can be thought of as villages, cities, or other population centers.
[22]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.

3.2 Census Sample

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.

In my analysis, I make four main restrictions on the census sample. First, I include only individuals born between 1890 and 1924. Following Duflo (2001), I define a set of "young," fully 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. I have excluded them. 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. The remaining cohorts, born between 1909 and 1919, were 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.[23] 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[24] and which were not classified as municipal capitals or cities in 1930 or earlier.[25] Given changes over time in the official definition of a rural locality,[26] 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.

[23]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.

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

[25]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.

[26]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.

In my main specification, I restrict the sample 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 13 plots the cumulative distribution of this proportion. The vast majority of localities have zero observations of older indigenous 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.

3.3 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

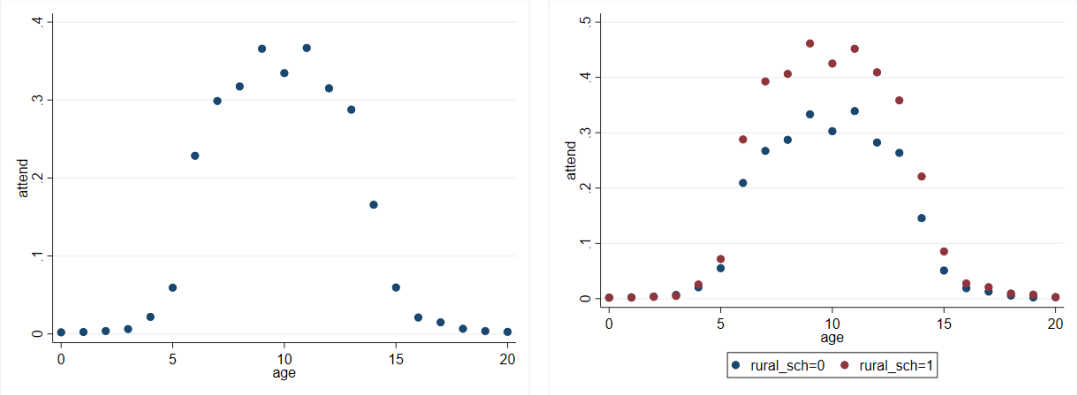
Data is restricted to localities that were not classified as municipal capitals, cities, or railroad stations in 1930 or earlier. I also exclude localities that are part of municipalities which did not have any federal rural schools in either 1925 or 1926, as well as localities with populations greater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

Table 3 shows descriptive statistics for individuals in my census sample. The proportion of the population that speaks Spanish is lower than the reported national value,[27] 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

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

1930. In Figure 1a, I plot a binned-scatter plot for school attendance by age for people twenty 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 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.

3.4 The Effect of Rural Schooling on Literacy and Language

3.4.1 Endogeneity in Rural School Allocation

It is useful to consider the identification problems which rule out a simple OLS regression of literacy and language on the presence of a rural school. The primary concern is the possibility of positive selection in the allocation of schools across locations. Even if the SEP, in practice, allocated these schools only to rural areas, some treated localities may have been closer to cities or the transportation network. Individuals in such a locality may been more likely to interact with the Spanish-speaking urban population. This proximity may have incentivized treated individuals to learn Spanish and become literate to take advantage of economic opportunities. If, instead, the government allocated schools to more remote areas, intending to target populations least likely to assimilate on their own, this would generate negative selection with respect to literacy

and assimilation. Another possibility is that communities selected themselves into treatment by petitioning the government for schools.[28]

Nevertheless, some objective criteria may have influenced whether a locality ultimately received a school or not. One of the many roles of cultural missionaries, which I briefly described in the historical context section, was to visit rural communities and report back to their supervisors, which they would provide with a list of information about each place and a recommendation for whether it should be included in the federal rural schooling program. The reports were relatively standardized, usually 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.

If these criteria were in fact generally followed, then the localities with schools should be expected to be smaller, have higher proportions of indigenous speakers, have few if any pre-existing schools, and be less likely to experience social dysfunction or local conflict. With my current data, the former two characteristics are observable. As discussed in the data section, I restrict my sample to rural, predominantly indigenous localities. Even combined with fixed-effects and controls, however, a simple OLS regression may not account for other sources of selection bias.

3.4.2 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, the first year of the rural schooling program, and thus likely too old to be affected, to cohorts born between 1920 and 1924, young enough to have benefited but old enough to be in school in 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

[28] 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.

a federal rural school in either 1925 or 1926. X_i is a vector of individual characteristics, sex/gender, and disability status.[29]

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[30] m with cohort fixed-effects to control for relatively 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. 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

[29]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.

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

born in year t residing in a treated locality l in 1930. Finally, in addition to the main specification that pools results by gender/sex, I run regressions separately for males and females, given that they had different base rates for literacy and language.

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.

3.4.3 Results

As explained in previous sections, literacy, which the SEP hoped would lead to a "spiritual awakening" of the rural population, is an important dimension of nation-building because of its role in constructing "imagined communities" (Anderson, 1984). Language homogenization through the acquisition of Spanish fluency was an important pre-requisite of this process. While forgetting an indigenous language is not necessary for the creation of a Spanish-speaking nation, a decrease in native-language fluency may be a leading indicator of a broader cultural shift away from indigenous culture toward a Mexican national identity. Moreover, any language loss in children may have inhibited the transmission of indigenous culture, making it more important for children to adopt norms from their Spanish-speaking environment.

My first set of results demonstrates that, in predominantly indigenous localities, the rural education program was successful in increasing literacy rates among school-aged children. Although I do not use the changes in literacy as a first stage, the results suggest that education is the likeliest mechanism through which language may have changed. I then present results on fluency in Spanish and indigenous languages, showing that schools increased Spanish fluency among children but decreased the likelihood that they spoke any 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. Although indigenous localities contain the sample of interest, the results for non-indigenous localities in Columns 3 and 4 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 the propensity to speak an indigenous language.

Literacy. Table 4 shows that children exposed to federal rural schools were more likely to be able read and write than older, unexposed cohorts. In Columns 1 and 2, I find

Table 4: Results: Literacy

	(1)	(2)	(3)	(4)
	Respondent Can Read and Write			
VARIABLES	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) X	0.0751***	0.0791***	0.0448**	0.0441**
Rural School	(0.0290)	(0.0291)	(0.0192)	(0.0190)
	[2.593]	[2.719]	[2.338]	[2.313]
Born 1909-1919 (Partial) X		0.0574***		0.0447***
Rural School		(0.0148)		(0.00927)
		[3.870]		[4.824]
Outcome Mean	0.146	0.155	0.273	0.289
Observations	22,120	31,895	74,804	114,345
R-squared	0.438	0.431	0.380	0.377
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	521	523	2110	2120
F-stat	6.725	7.929	5.468	11.73
R2 w/in adj.	0.00216	0.00188	0.000476	0.000479

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.

Table 5: Results: Female Literacy

VARIABLES	(1)	(2)	(3)	(4)
	Respondent Can Read and Write			
	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) X	0.0742**	0.0713**	0.0708***	0.0704***
Rural School	(0.0308)	(0.0313)	(0.0213)	(0.0212)
	[2.411]	[2.277]	[3.324]	[3.324]
Born 1909-1919 (Partial) X		0.0449***		0.0721***
Rural School		(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.

Table 6: Results: Male Literacy

VARIABLES	(1)	(2)	(3)	(4)
	Respondent Can Read and Write			
	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) X	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) X		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.

that literacy is about 8% higher in 1930 among exposed children compared to the older cohort. In Column 2, the interaction for the partially exposed cohorts shows an increase of about 6%. These estimates are all statistically significant at the 1% level. Compared to the mean literacy rate in the indigenous sample, these effects suggest an increase of about 50%. 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 attend rural schools become literate. Columns 3 and 4 show that schools may have increased literacy in non-indigenous localities as well, though the effects are smaller in both absolute and relative magnitude. This may be due to the fact that these populations already have relatively higher literacy rates, just under 30%.

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.

Table 7: Placebo Check: Literacy

	(1)	(2)	(3)
	Respondent Can Read and Write		
VARIABLES	Pooled	Female	Male
Born After 1900 X	-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.

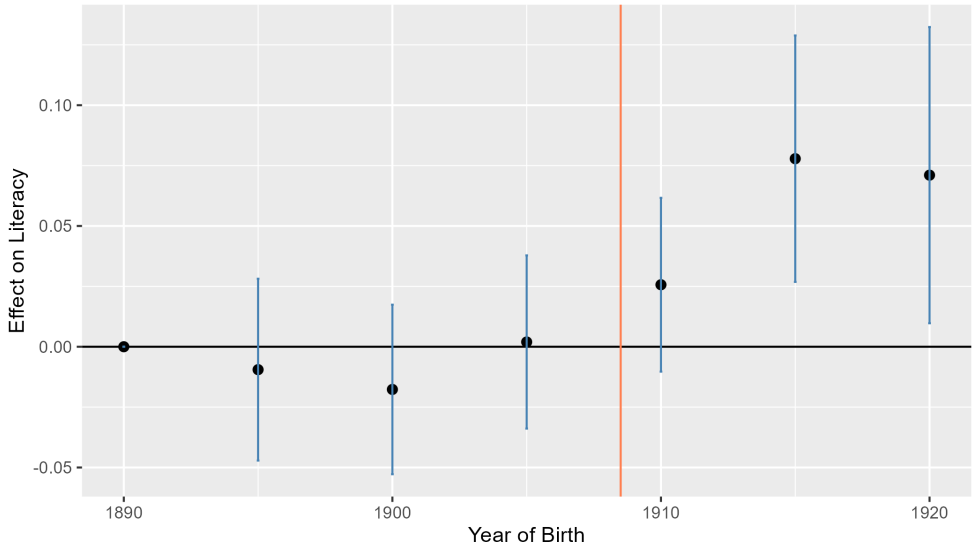


Figure 2: Dynamic DiD: Effect of Rural Schools on Literacy

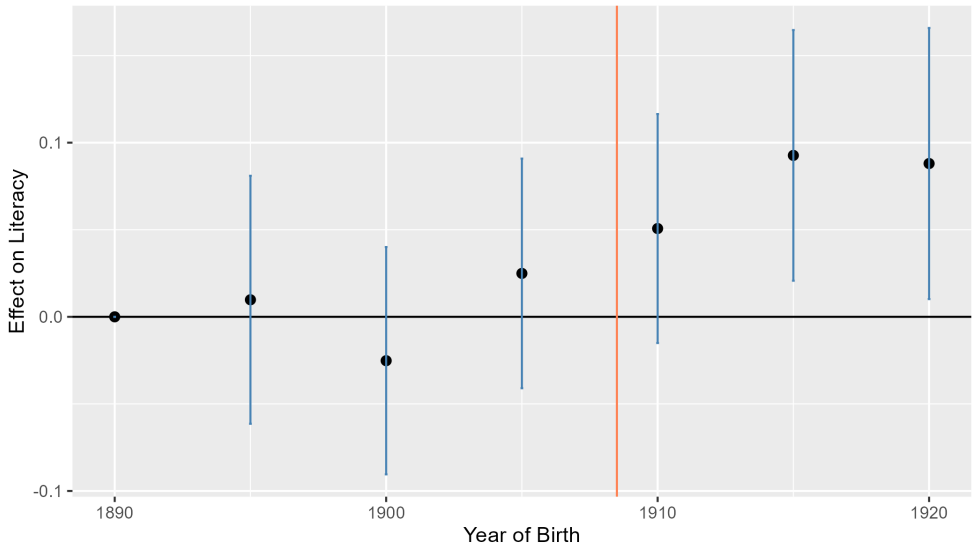


Figure 3: Dynamic DiD: Effect of Rural Schools on Literacy, Males

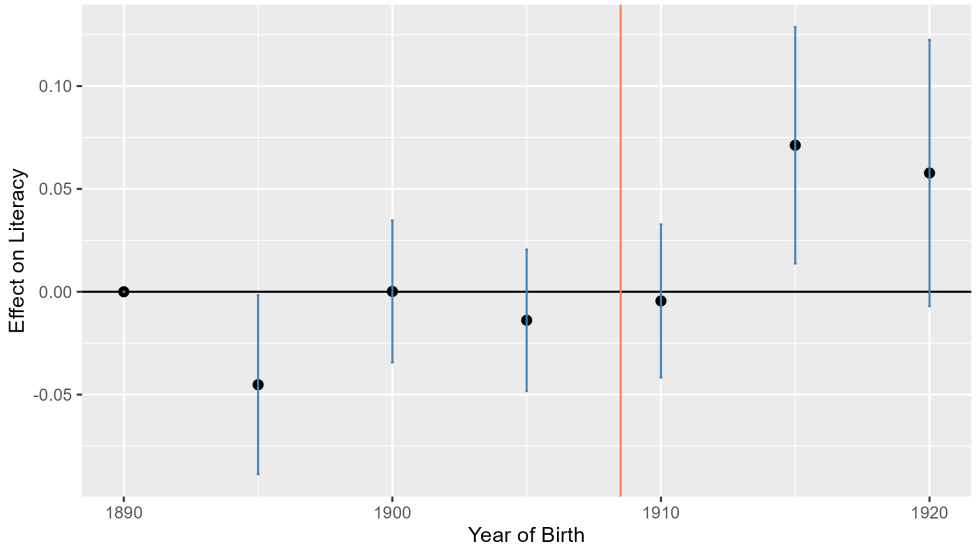


Figure 4: Dynamic DiD: Effect of Rural Schools on Literacy, Females

Splitting the sample by gender/sex, I find that that the effects on literacy for both females and males are qualitatively consistent with the pooled regressions. Columns 1 and 2 in Table 5 and Table 6 show that effects for males and females are similar in magnitude, about 7-9%. However, given that females have half the literacy rate of males, about 10%, this suggests a doubling of literacy rates for females and an increase in 50% for males. The effects on partially exposed cohorts are similar to those in the Table 4. In non-indigenous localities, the effect of schools on literacy appears to be driven primarily by females, who experienced an increase in literacy of about 7%, compared to virtually no increase for males.[31]

As a check on the plausibility of the parallel trends assumption, in Table 7, 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 test for pre-tends, Figure 2 plots the coefficients from my dynamic specification, allowing for a visual check for changes before the treatment period. Figure 3 and Figure 4 plot coefficients from the same specification, splitting the sample by gender/sex. In all three 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 three plots show that there is little evidence for significant changes in the period before 1909, with significant increases in literacy for exposed cohorts. These plots, along with my formal placebo checks in Table 7, suggest that the parallel trends assumption is plausible in this context, 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 8 shows the estimated effects of exposure to rural schools on the propensity for children to speak Spanish in 1930, pooling both male and female populations. In indigenous localities, Columns 1 and 2 show that exposure to schools increases the probability that a child speaks Spanish by about 8%. Compared to a mean Spanish fluency rate of about 50%, this is an increase of about 20%. In Table 9

[31] 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. The broad literacy campaign of the 1920s may have fostered gender equity in education.

Table 8: Results: Spanish Fluency

	(1)	(2)	(3)	(4)
	Respondent Speaks Spanish			
VARIABLES	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) X	0.0834***	0.0852***	-0.00231	-0.00208
Rural School	(0.0232)	(0.0233)	(0.00157)	(0.00153)
	[3.592]	[3.648]	[-1.476]	[-1.357]
Born 1909-1919 (Partial) X		0.0391***		-0.00275
Rural School		(0.0134)		(0.00169)
		[2.922]		[-1.628]
Outcome Mean	0.501	0.503	0.990	0.990
Observations	22,120	31,895	74,804	114,345
R-squared	0.643	0.642	0.386	0.378
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	521	523	2110	2120
F-stat	12.90	7.637	2.179	1.737
R2 w/in adj.	0.00210	0.00153	1.11e-05	1.50e-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.

Table 9: Results: Female Spanish Fluency

	(1)	(2)	(3)	(4)
	Respondent Speaks Spanish			
VARIABLES	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) X	0.0910***	0.0905***	-0.00248	-0.00270
Rural School	(0.0277)	(0.0279)	(0.00246)	(0.00239)
	[3.289]	[3.248]	[-1.006]	[-1.129]
Born 1909-1919 (Partial) X		0.0561***		-0.000801
Rural School		(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.

Table 10: Results: Male Spanish Fluency

	(1)	(2)	(3)	(4)
	Respondent Speaks Spanish			
VARIABLES	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) X	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) X		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.

Table 11: Placebo Check: Spanish

	(1)	(2)	(3)
	Respondent Speaks Spanish		
VARIABLES	Pooled	Female	Male
Born After 1900 X	-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.

and Table 10, Columns 1 and 2 show qualitatively similar results, though for females the estimated effects are slightly larger in absolute and relative magnitude. In all three 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 in non-indigenous areas.

To test for pre-trends, Table 11 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.

To visually check whether pre-trends may threaten my identification, Figure 5, Figure 6, and Figure 7 plot coefficients from my dynamic specification for the pooled, male, and female sample respectively. Figure 6 suggests that the significant decrease in Spanish fluency in Table 11 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.

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 12 presents 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, I find that children exposed 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. Table 13 and Table 14 suggest that these results are mostly driven by females. This may make sense given that the relative effect of school exposure on females in indigenous localities was higher for Spanish fluency and literacy compared to males. In all three tables, Columns 3 and 4 show no evidence of any change in indigenous language in non-indigenous localities.

Table 15 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.

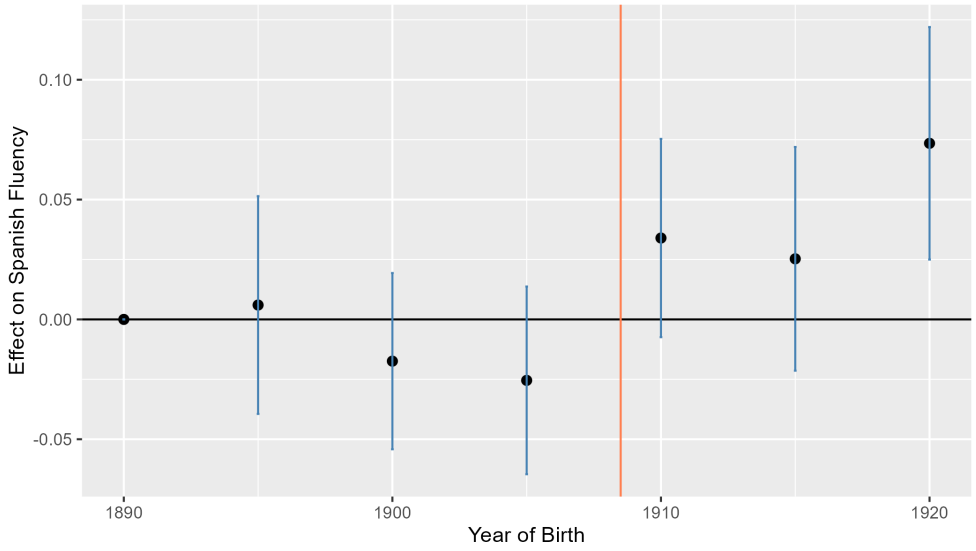


Figure 5: Dynamic DiD: Effect of Rural Schools on Spanish

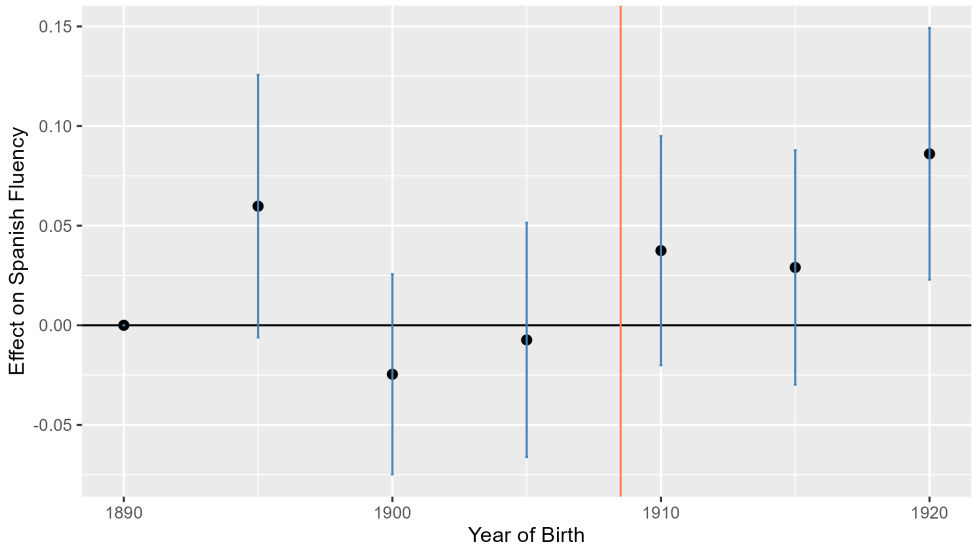


Figure 6: Dynamic DiD: Effect of Rural Schools on Spanish, Males

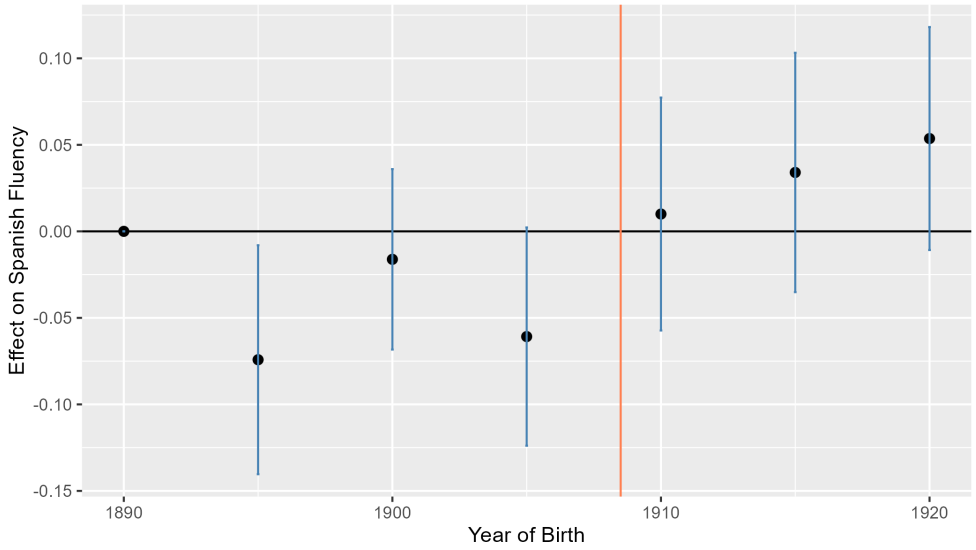


Figure 7: Dynamic DiD: Effect of Rural Schools on Spanish, Females

Plots for the coefficients in my dynamic specification may shed light on the different results for males and females in this outcome. In Figure 9, 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. The pooled and female plots, Figure 8 and Figure 10, are consistent with reductions in the use of indigenous language for females as a result of school exposure, with no evidence of significant pre-trends.

Although the relative magnitude of this decrease is much smaller than the relative increases in literacy and Spanish fluency, the high rate

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, 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.[32]

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.

4 Rural Schools and Land Reform

I now explore the role of schools on another important economic policy during my study period. Beginning in 1915, land reform was conducted through the expropriation of large landholdings, or *haciendas*. This land was redistributed through *ejidos*, a semi-communal land tenure system. In this section, I present evidence on and offer explanations for the relationship between school establishments and land grants.

[32]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 12: Results: Indigenous Fluency

VARIABLES	(1)	(2)	(3)	(4)
	Respondent Speaks an Indigenous Language			
	Indigenous Localities		Non-Indig. Localities	
Born After 1920 (Full) X	-0.0393***	-0.0410***	0.00276	0.00292
Rural School	(0.0146)	(0.0146)	(0.00211)	(0.00208)
	[-2.698]	[-2.803]	[1.308]	[1.402]
Born 1909-1919 (Partial) X		-0.0166		0.00183
Rural School		(0.0104)		(0.00225)
		[-1.592]		[0.814]
Outcome Mean	0.906	0.903	0.0204	0.0192
Observations	22,120	31,895	74,804	114,345
R-squared	0.413	0.428	0.488	0.466
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	521	523	2110	2120
F-stat	7.277	3.988	1.711	1.004
R2 w/in adj.	0.000802	0.000583	7.24e-06	-8.34e-07

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.

Table 13: Results: Female Indigenous Fluency

VARIABLES	(1) Respondent Speaks an Indigenous Localities	(2) Indigenous Language Localities	(3) Non-Indig. Localities	(4)
Born After 1920 (Full) X 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) X 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.

Table 14: Results: Male Indigenous Fluency

VARIABLES	(1) Respondent Speaks an Indigenous Localities	(2) Indigenous Language Localities	(3) Non-Indig. Localities	(4)
Born After 1920 (Full) X Rural School	-0.0226	-0.0243	0.00513	0.00487
	(0.0175)	(0.0172)	(0.00314)	(0.00304)
	[-1.288]	[-1.414]	[1.637]	[1.600]
Born 1909-1919 (Partial) X Rural School		-0.00346		0.00515
		(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.

Table 15: Placebo Check: Indigenous

	(1)	(2)	(3)
	Speaks an Indigenous Language		
VARIABLES	Pooled	Female	Male
Born After 1900 X	0.0128	0.000337	0.0339**
Rural School	(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.

4.1 Data

My analysis on the effects of federal rural schools on land reform uses the same data on schools as did the previous analysis on indigenous assimilation. To measure land redistribution at the locality level, I rely on the *Padrón e Historial de Núcleos Agrarios* (PHINA) database from the *Registro Agrario Naciónal* (RAN). This is the cadastral registry of *ejidos*, which, among other information, gives the name of each *ejido*, its municipality, and the year in which the land was granted. [33] Based on these dates, my data on land reform begins in 1914.

To link *ejidos* and localities, I combine shape files from RAN mapping the perimeters of each *ejido* with the geographic coordinates for each locality from the INEGI geographic database. For simplicity, I define the first year of a locality's exposure to land reform by taking the year of the initial land grant of the closest *ejido* within five kilometers. Then, the outcome variable of interest is a yearly indicator that is equal to one in the years after a locality is exposed to land reform and zero in the years before. [34]

As with my individual-level analysis, I exclude from my land reform data set localities that were classified as municipal capitals, cities, or train stations in 1930 or earlier, as well as localities with populations no greater than 3,000 in 1930. I further restrict my sample to *ejidos* founded between 1914 and 1940 because the bulk of total redistributions occurred by the end of President Cárdenas' administration in 1940.

I also use data from Garfias (2018) and Sanderson, which aggregates land reform data to the municipality level and combines it with municipality-level tabulations from the 1930 census. One advantage of using this data is that it includes information on whether each land reform petition was approved or denied. I then aggregate my data on rural schools to the municipality level and the number of schools per 10,000 individuals in the 1930 census as a measure of treatment. I combine these data for my municipality-level analysis.

[33] Because not all *ejidos* have been formally registered, approximately 28% do not report information on land grants. For those that do, this information includes the grant approval date and the date of its execution. Although I do not have data on the time it took from the initial petition to the official sanctioning of a land grant, the median *ejido* took about a year to move from the approval to the execution of the grant. At the tails of this distribution, some *ejidos* took longer than a decade, though it is unclear if the long gap in these cases is due to misreporting or difficulties in enforcing land redistribution.

[34] Restricting the sample to *ejidos* from the PHINA data that can be matched to a shapefile, I am left with approximately 70% of the original data set for which I have information on a land grant. A locality can be matched to an *ejido* in several ways using the data. In untabulated results (available upon request), I find that alternative approaches to matching localities to *ejidos* produce qualitatively similar results.

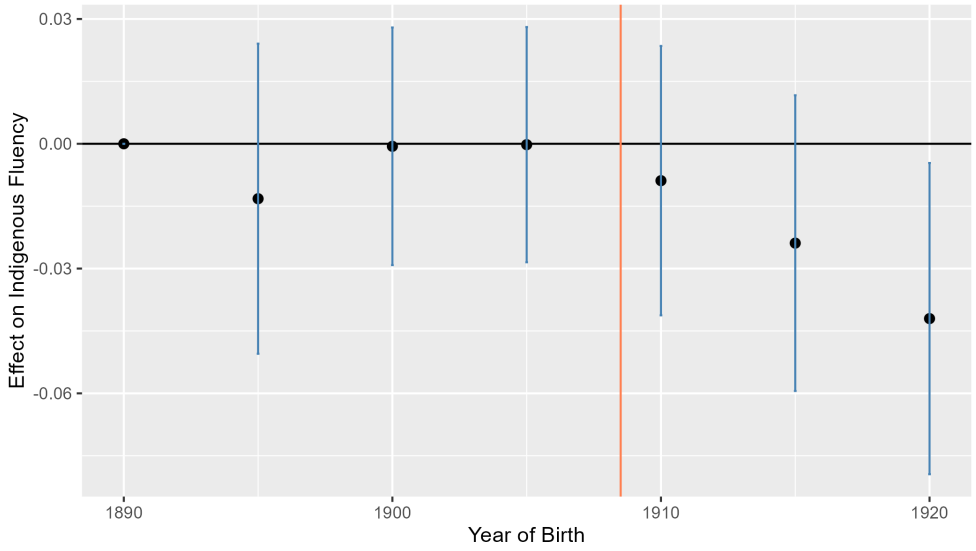


Figure 8: Dynamic DiD: Effect of Rural Schools on Indigenous Speaking

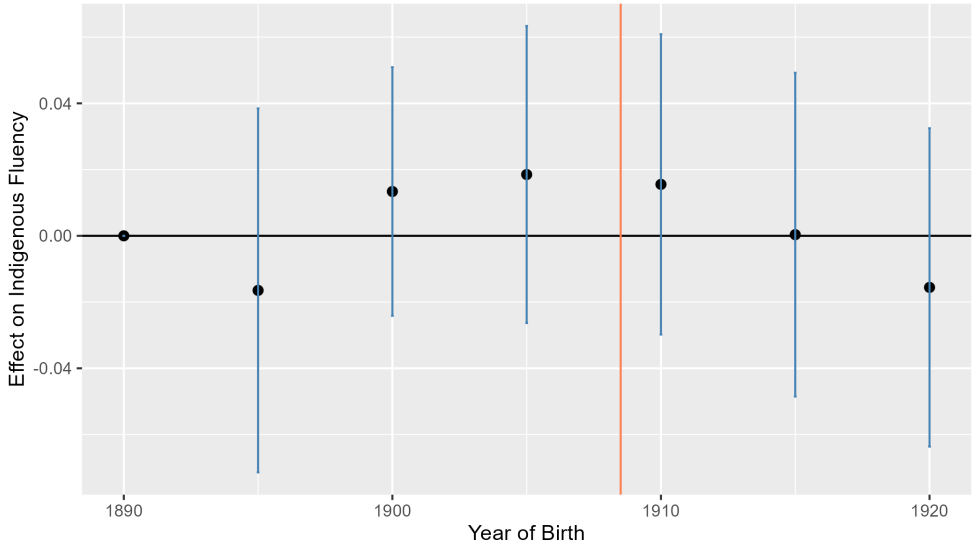


Figure 9: Dynamic DiD: Effect of Rural Schools on Indigenous, Males

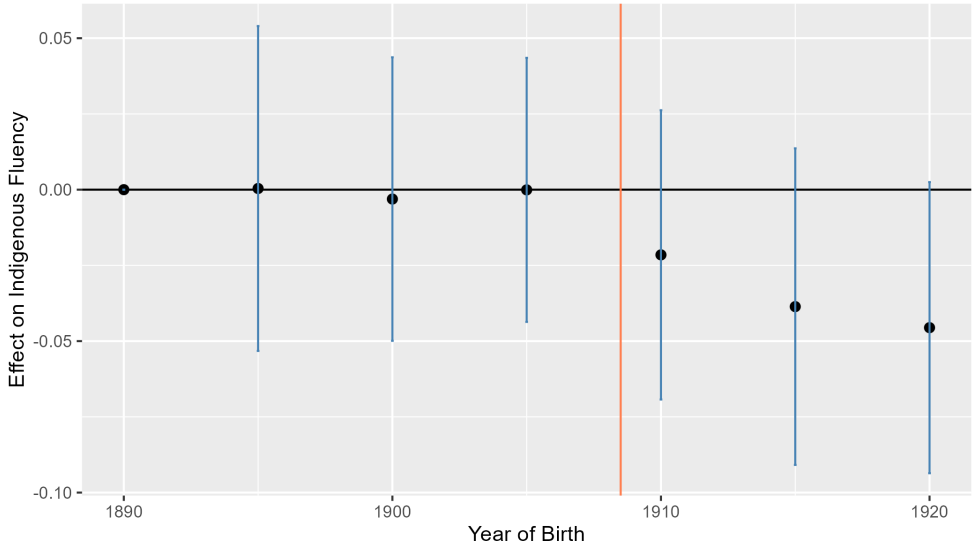


Figure 10: Dynamic DiD: Effect of Rural Schools on Indigenous, Females

4.2 The Relationship between Schools and Land Redistribution

4.2.1 Empirical Specification

My approach to assessing the impacts of rural schools on land reform is similar to my difference-in-difference strategy in Equation 3. Formally, my dynamic, locality-level specification for land reform is:

$$Y_{lt} = \alpha + \sum_{t} \beta_t \text{School}_{lt} + \gamma_t + \delta_l + \phi_{mt} + \epsilon_{lt} \tag{4}$$

where Y_{lt} is an indicator that is equal to one in the (approximate) year in which a locality first experiences land reform, remains one in all subsequent years, and is zero otherwise. Treatment_{lt} is an indicator equal to one in year t for treated locality l, and zero otherwise. The year 1922 is omitted. γ_t is a set of year fixed-effects, δ_l is a set of locality fixed-effects, and ϕ_{mt} is a set of municipality-by-year fixed-effects for municipality m.

For my munucipality-level analysis, I use the following specification:

$$Y_{mt} = \alpha + \sum_{t} \beta_t \text{SchoolsPer10k}_{mt} + \gamma_t + \delta_m + logPop_m \times \mathbb{1}_t + \phi_{st} + \epsilon_{lt} \tag{5}$$

where I regress Y_{mt}, a yearly measure of land redistribution in municipality m, on $\text{SchoolsPer10k}_{mt}$, an interaction between the number of schools per 10,000 people in 1930 and year fixed-effects, municipality fixed-effects δ_m, year fixed-effects γ_t, state (s) by year fixed-effects ϕ_{st}, and an interaction between the log of total population in m in 1930 and year fixed-effects. From the Garfias (2018) and Sanderson data, I define the following municipality-level measures of land reform: cumulative total land reform events, cumulative total land reform events in which irrigated land was redistributed, cumulative total *ejidos* created, and cumulative total *comunidades* created.[35]

4.2.2 Results

In this section, I present results from my dynamic specifications estimating the correlation between the presence of federal rural schools and the incidence of land reform. I first show results at the locality level, then at the municipality level. I then discuss under what conditions these results may be interpreted as a causal relationship.

Locality-level. My outcome of interest is a dummy variable indicating whether a locality was within five kilometers of the boundary of any *ejido* established up until the year in question. This is my proxy for whether a locality had ever received a land grant by that year. In Figure 11, I plot the coefficients of my dynamic specification, in which I regress my outcome on interactions between treatment and yearly indicators. This figure shows a clear divergence in the receipt of redistributed land between places

[35] In the Appendix I also use the proportion of land reform events that were approved in a given year to check for differential approval rates across years.

with and without rural schools. The difference between treated and untreated localities significantly increases in magnitude between 1922 and 1933.[36] The positive correlation between rural schools and land reform in this period is consistent with the hypothesis that schools may have played a role in the process of land redistribution. As discussed in the historical context section, contemporary accounts suggest that teachers may have played a role in organizing communities to petition for land.

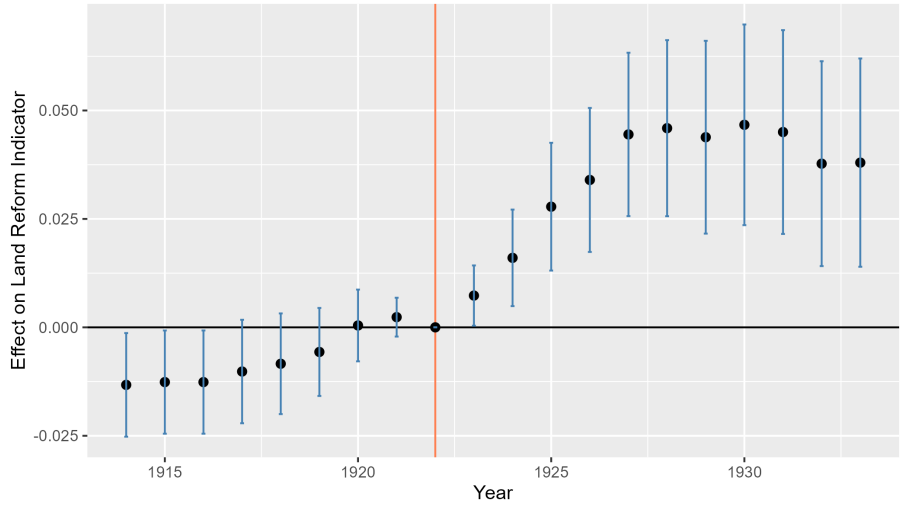


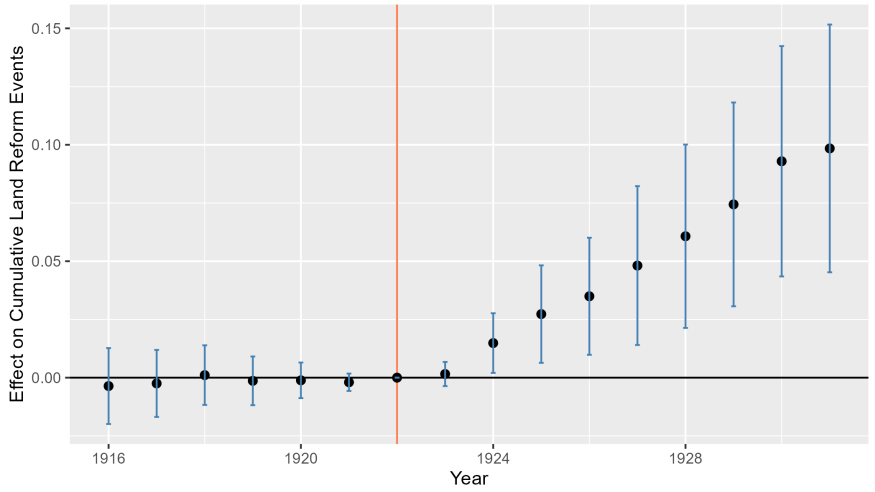
Figure 11: Locality-Level Correlation between Rural Schools and Land Reform in a Previous Year

Municipality-level. Aggregating the data to the municipality level, I lose some of the granularity of the analysis, but I can take advantage of additional municipality-level data collected by Garfias (2018) and Sanderson. In Figure 24, I plot the coefficients from my dynamic specification Equation 5 for my four main municipality-level outcomes. Following Garfias (2018), I restrict to municipalities with at least one *hacienda*, or large landholding, as these were the places with the most valuable land and the ones that were most likely to be the target for redistribution.[37]

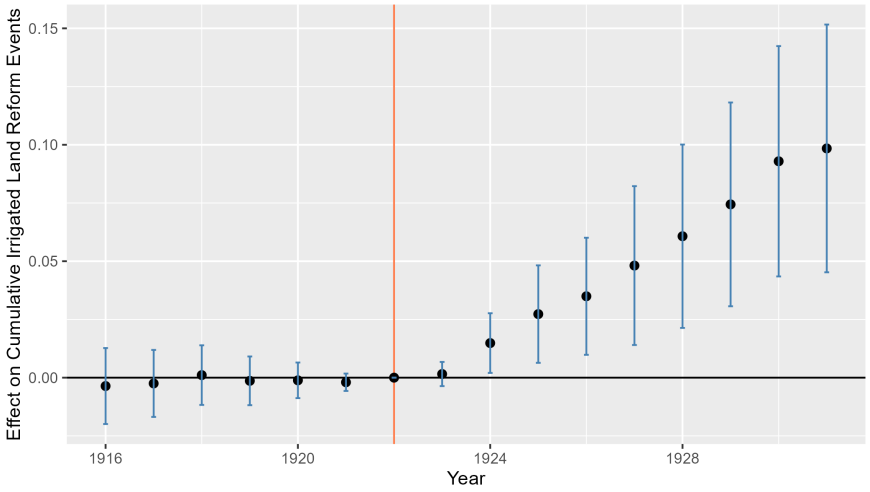
In the top two panels, Figure 24a and Figure 24b show the correlation between rural schools and the cumulative redistribution events in each municipality, controlling for the fixed effects and interacted covariates as specified in Equation 5. In Figure 24a, I include all land reform events, whereas in Figure 24b I include only events which included at least one hectare of irrigated land, which Garfias (2018) argues was the most productive land and the most contentious to redistribute. For both of these outcomes, an increase in the number of schools per capita is associated with a statistically significant increase in land reform events. The similarity in magnitudes of the coefficients shows that the correlation is primarily driven by irrigated land, suggesting that any effects of schools on land land may have worked through the redistribution of valuable land.

[36] In results not shown here, this difference falls to zero and then becomes negative between 1934 and 1940. A possible explanation for the latter convergence is that this period covers the Cárdenas administration. Cárdenas was one of Mexico's most left-wing presidents and led a campaign to rapidly expand land reform. Consequently, localities that did not receive a federal rural school during my study period were more likely to have received a land grant after 1934.

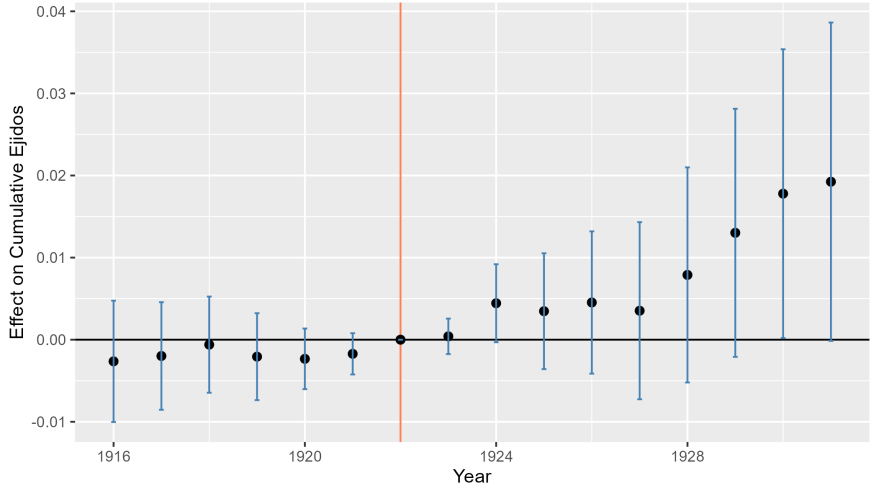
[37] See **??** in the Appendix for results using all municipalities.



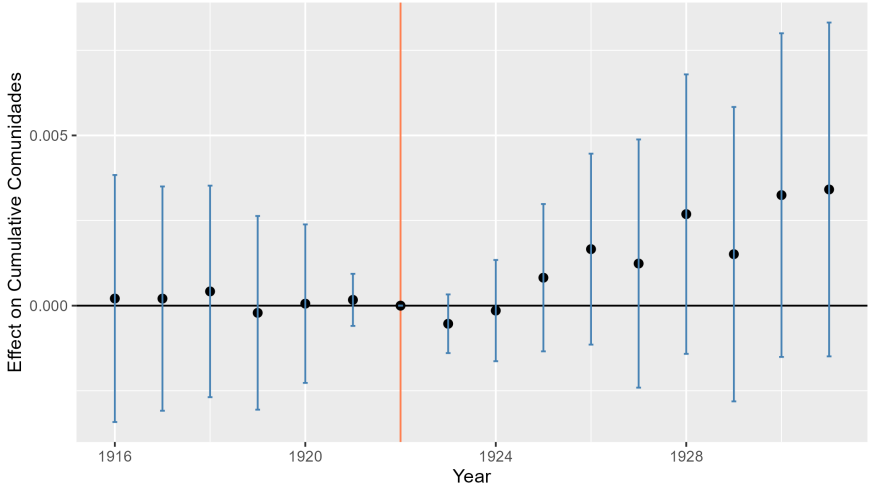
(a) All Events



(b) Irrigated Land



(c) *Ejidos*



(d) *Comunidades*

Figure 12: Municipality-Level Correlation: Rural Schools per 10000 and Land Reform

In the bottom panels, I split all land reform events by whether they led to the creation of *ejidos* (Figure 24c) or *comunidades* (Figure 24d). Recall that *comunidades* were a special redistribution process whereby indigenous communities petitioned for the restitution of ancestral lands, often through colonial-era land titles. Although the coefficients in these two panels are mostly insignificant in the period after the rural schools program began, their upward trajectory suggests that schools were associated with increases in the creation of both *ejidos* and *comunidades*. The latter results suggests that these gains were not restricted to non-indigenous communities.

Pre-trends and Concurrent Policy Changes. To interpret these correlations as a causal effect of rural school presence on the incidence of land reform, my analysis must assume that the rate of land redistribution would not have changed in the absence of treatment, i.e. the parallel trends assumption, and there must not be any other policy change in 1922 that could affect redistribution rates. In my locality-level analysis, the coefficients for years preceding the rural school expansion in Figure 11 show a slight but significant difference. Although the pre-trends are of a much smaller magnitude than the large divergence in the post-period, there was also a lower rate of overall redistribution before 1922. Given that land reform was increasingly more common in treated localities even before 1922, these results are evidence of differential pre-trends in my outcome, indicating the that trends in my treatment period may not be parallel.

In the municipality-level analysis, however, Figure 24 shows that once restricting the sample to municipalities with at least one *hacienda*, there is little evidence of pre-trends.[38]

Even if the parallel trends assumption is plausible after restricting to places with *haciendas*, 1922, the first year of the rural schools program, also coincided with an important change in the land reform process. Specifically, this was the year in which the president became the final word on the approval or denial of each petition. If this makes it easier for a president to implement land redistribution in politically sensitive locations, then we might expect an uptick in land reform events after this period. The fact that the coefficients in the top panels are not significant until 1924 is slightly reassuring, since this the Calles administration (1924-1928) saw the largest growth in rural schools in the 1920s. Moreover, in the Appendix, regressing the yearly approval rate from the Sanderson data onto the schools per capita/year indicators interactions does not show a significant change in approval rates after 1922.

One difficulty in interpreting these results is the fact that I am using the year of the initial land grant as my outcome, not the year a community first petitioned for land. The federal government's official periodical, in which presidential resolutions approving or denying land grant requests were originally published, reports the year each petition was initially filed. Based on a handful of issues I have reviewed, it is not

[38] **??** in the Appendix shows that, when including all municipalities, pre-trends appear again at the municipality level. In results not reported here, when I exclude municipalities with any *hacienda*, the trends before and after 1922 are much different from the results in Figure 24. This may be because these municipalities are quite different from those with *haciendas* and therefore experience land reform very differently.

uncommon for the lag between petitions and grants to be close to five years (it was even longer in some cases). If treatment and control groups experienced different lags, then it would be difficult to interpret these results without knowing the year of the petition. Petitions would be a more direct measure of collective action than the actual redistribution, as they did not require government action to submit. Given current data limitations, I refrain from making a causal interpretation of these results. However, they are consistent with and highly suggestive of rural schools playing an important role in the implementation of land reform.

4.2.3 Did Land Reform Increase Investments in Education?

At first glance, the pre-trends in Figure 11 may constitute a threat to identification in my results on cultural assimilation. When considering the potential effects of land reform on an outcome, the proper counterfactual to an *ejidatario* receiving an individual plot of land is that of a field hand working the land of a wealthy *hacendero*. Because land reform was meant to benefit landless or dispossessed peasants, it is possible that *ejidatarios* experienced income and wealth effects. These effects may have increased investments in education, specifically literacy and Spanish fluency. Moreover, *ejidatarios* may have become involved in the marketing of agricultural goods, incentivizing their acquisition of basic literacy and numeracy skills as well as the adoption of a common language.

The above suggests that adults in the 1920s would have had the strongest incentives to acquire language skills, but my findings on assimilation are driven primarily by children. Then in order for the bundling of land to confound my effect on literacy and language, it must be that adults invested in children but did not become more educated themselves. Recent work suggests, however, that by granting incomplete property rights over their land, this system led to labor misallocation and limited returns to scale.[39] This would have limited an *ejido*'s productivity and any resulting income effects. Historical evidence also suggests that the public provision of agricultural inputs and credit for *ejidos* was subject to political influence, further limiting the economic benefits of a land grant (Yates, 1981).[40] Moreover, even if land distribution had significantly increased incomes, the crisis in education that spurred the creation of the SEP suggests that without federal support, the supply of educational inputs may have been limited.

5 Conclusion

In this paper, I show that Mexico's first major effort to expand public education increased literacy rates and Spanish fluency, while decreasing the propensity for children to speak

[39] Specifically, (De Janvry et al., 2015) find that when *ejidatarios* were issued land titles after a reform in 1992, affected households were more likely to have a migrant member (consistent with labor misallocation and reallocation), whereas there was consolidation of farming units (suggesting the exploitation of returns to scale).

[40] Dell (2012) also finds that *ejidos* are associated with a reduction in the provision of public goods.

indigenous languages. These constitute two important dimensions of nation-building. On the one hand, literacy may allow for individuals to conceive of themselves as part of a broader "imagined community" by reading about fellow nationals in print media. On the other hand, language homogenization increases the possibility for fruitful communication with strangers, possibly leading individuals to identify more with the nation and less with their ethnic groups. Given that I observe these effects in children in 1930, my findings serve as a leading indicator for the process of indigenous assimilation across the twentieth century. I also find that places with rural schools experienced increases in land redistribution, an important Revolutionary reform. I offer historical examples suggesting that teachers may have played a role in legitimizing the post-Revolutionary government by helping communities apply for land reform. In sum, by increasing literacy rates, inducing language homogenization, and helping implement land reform, the Mexican government's 1920s rural schools program appears to have been an effective nation-building policy.

References

Alesina, Alberto and Eliana La Ferrara. Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800, 2005. ISSN 00220515.

Alesina, Alberto, Bryony Reich, and Alessandro Riboni. *Nation-building, nationalism, and wars*, volume 25. Springer US, 2020. ISBN 0123456789.

Alesina, Alberto, Paola Giuliano, and Bryony Reich. Nation-Building and Education. 131(January):2273–2303, 2021.

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

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

Angeles, Luis and Aldo Elizalde. Pre-colonial institutions and socioeconomic development: The case of Latin America. *Journal of Development Economics*, 124:22–40, 2017. ISSN 03043878.

Assouad, Lydia. Charismatic Leaders and Nation-Building. 2021.

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

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

Cárdenas, Enrique. *El largo curso de la economía mexicana. De 1780 a nuestros días*. Fondo de Cultura Económica el Colegio de México, México, 1st edition, 2015. ISBN 978-607-462-783-1.

De Janvry, Alain, Marco Gonzalez-Navarro, and Elisabeth Sadoulet. Are land reforms granting complete property rights politically risky? Electoral outcomes of Mexico's certification program. *Journal of Development Economics*, 110:216–225, 2014. ISSN 03043878.

De Janvry, Alain, Kyle Emerick, Marco Gonzalez-Navarro, and Elisabeth Sadoulet. Delinking land rights from land use: Certification and migration in Mexico. *American Economic Review*, 105(10):3125–3149, 2015. ISSN 19447981.

Dell, Melissa. Path dependence in development: Evidence from the Mexican Revolution. 2012.

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

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

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

Duflo, Esther. 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.

Elizalde, Aldo. On the economic effects of Indigenous institutions: Evidence from Mexico. *Journal of Development Economics*, 147(August):102530, 2020. ISSN 03043878.

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

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

Feir, Donna L. 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.

Feir, Donna L. 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.

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

Garfias, Francisco. Elite competition and state capacity development: Theory and evidence from post-revolutionary Mexico. *American Political Science Review*, 112(2): 339–357, 2018. ISSN 15375943.

Garfias, Francisco and Emily A. Sellars. From conquest to centralization: Domestic conflict and the transition to direct rule. *Journal of Politics*, 83(3):992–1009, 2021. ISSN 14682508.

Glaeser, Edward L., Giacomo A.M. Ponzetto, and Andrei Shleifer. Why does democracy need education? *Journal of Economic Growth*, 12(2):77–99, 2007. ISSN 13814338.

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

Hall, Linda B. Alvaro Obregón and the Politics of Mexican Land Reform, 1920-1924. *The Hispanic American Historical Review*, 60(2):213–238, 1980.

Instituto Nacional de Estadística y Geografía. Quinto Censo de Población 1930.

Knight, Alan. Racism, Revolution, and Indigenismo: Mexico, 1910-1940. In *The Idea of Race in Latin America, 1870-1940*, pages 71–113. Austin, TX, 1990.

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

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

Paglayan, Agustina S. The Non-Democratic Roots of Mass Education: Evidence from 200 Years. pages 179–198, 2021.

Paglayan, Agustina S. Education or Indoctrination? The Violent Origins of Public School Systems in an Era of State-Building. pages 1–16, 2022.

Roth, Jonathan, 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.

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

Sanderson, Susan Walsh. Land Reform in Mexico 1910-1976.

Secretaría de Economía, Dirección General de Estadística. *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.

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

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

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

Yates, P. Lamartine. Mexican Land Reform, 1959-1969: A Comment. *The Economic Journal*, 91(363):745–752, 1981.

Zamudio Sánchez, Francisco José, 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.

Zamudio Sánchez, Francisco José, 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.

A Appendix

A.1 Appendix Tables

Table 16: 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 17: 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 13: Cumulative Distribution of Older Indigenous Population

A.2.2 Dynamic DiD Plots, 3-Year Bins

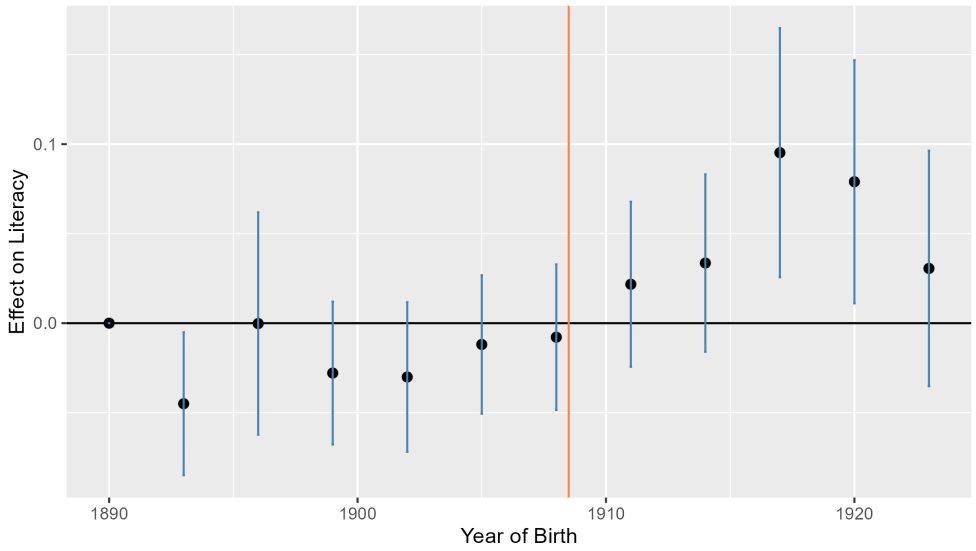


Figure 14: Dynamic DiD: Effect of Rural Schools on Literacy

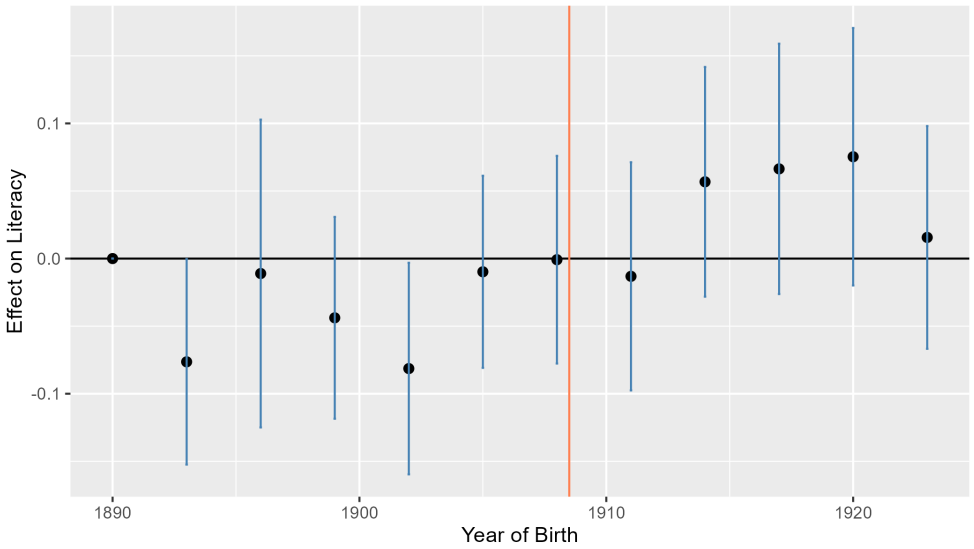


Figure 15: Dynamic DiD: Effect of Rural Schools on Literacy, Males

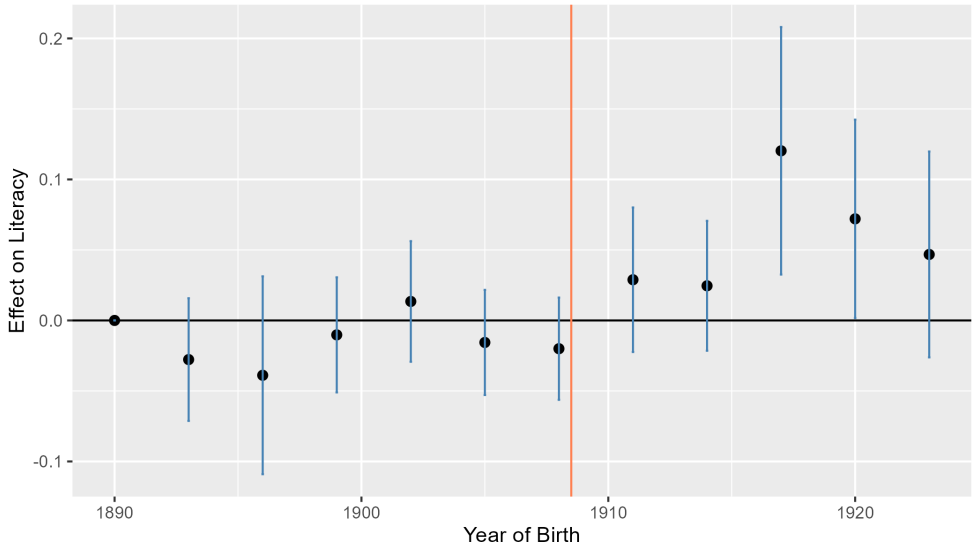


Figure 16: Dynamic DiD: Effect of Rural Schools on Literacy, Females

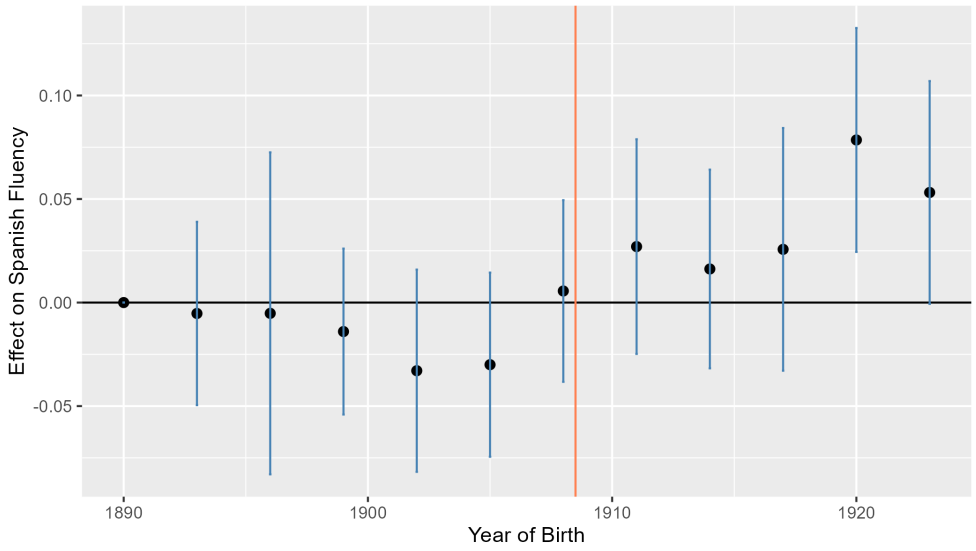


Figure 17: Dynamic DiD: Effect of Rural Schools on Spanish

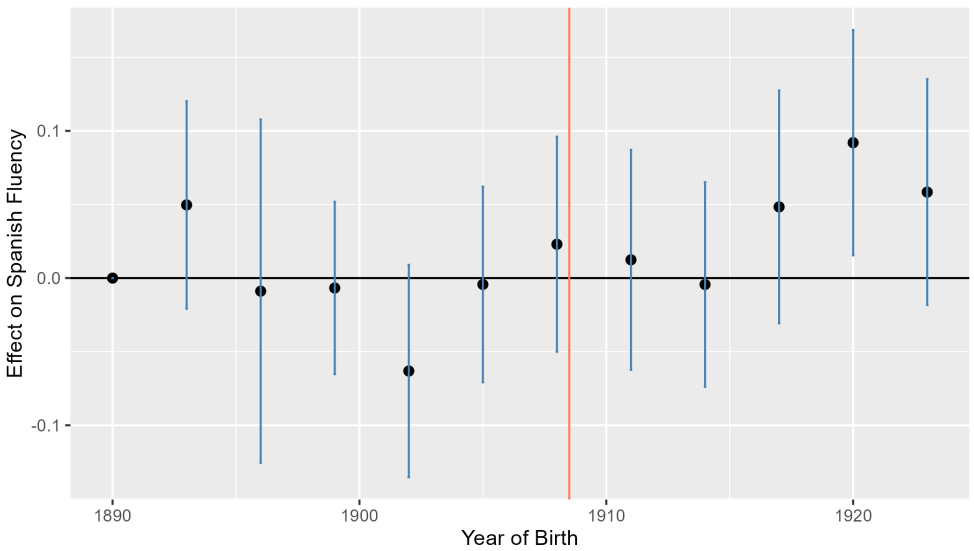


Figure 18: Dynamic DiD: Effect of Rural Schools on Spanish, Males

Figure 19: Dynamic DiD: Effect of Rural Schools on Spanish, Females

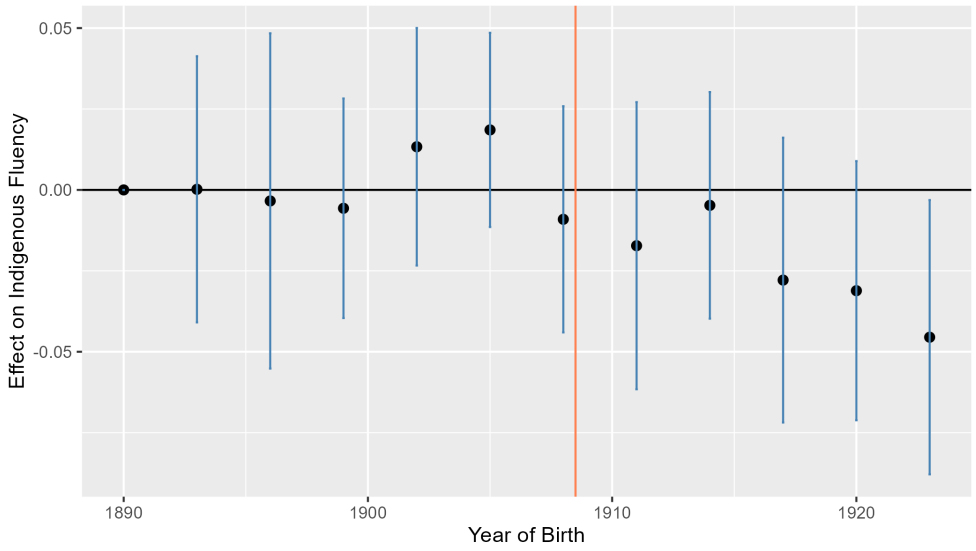


Figure 20: Dynamic DiD: Effect of Rural Schools on Indigenous Languages

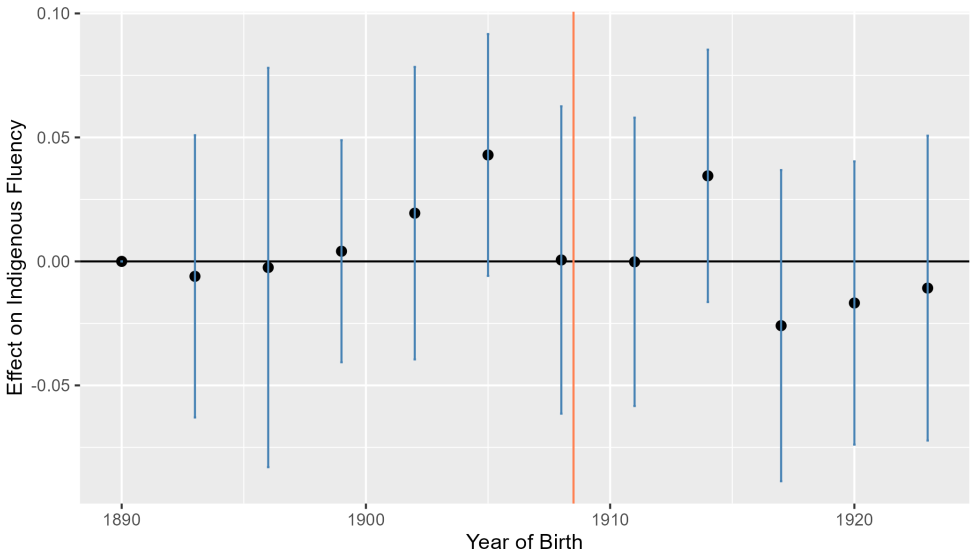


Figure 21: Dynamic DiD: Effect of Rural Schools on Indigenous Languages, Males

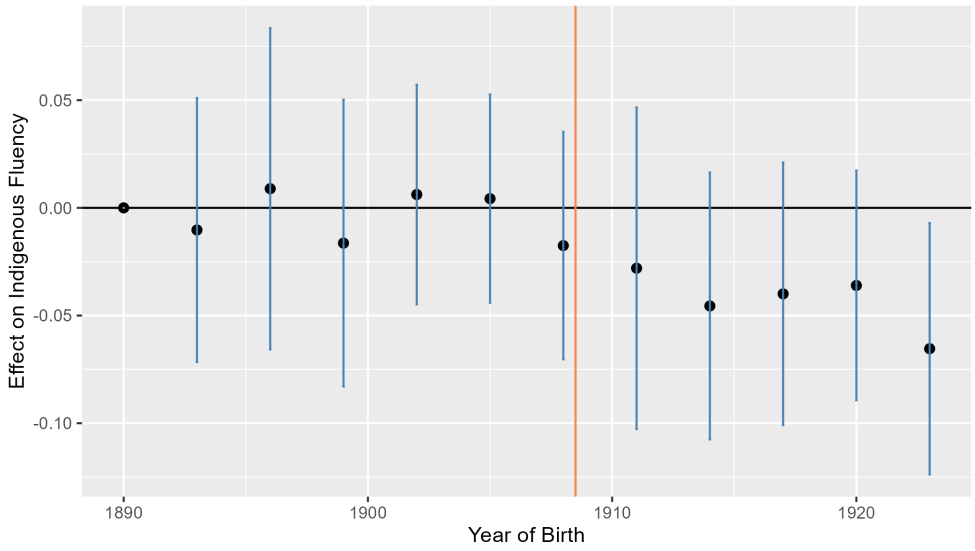
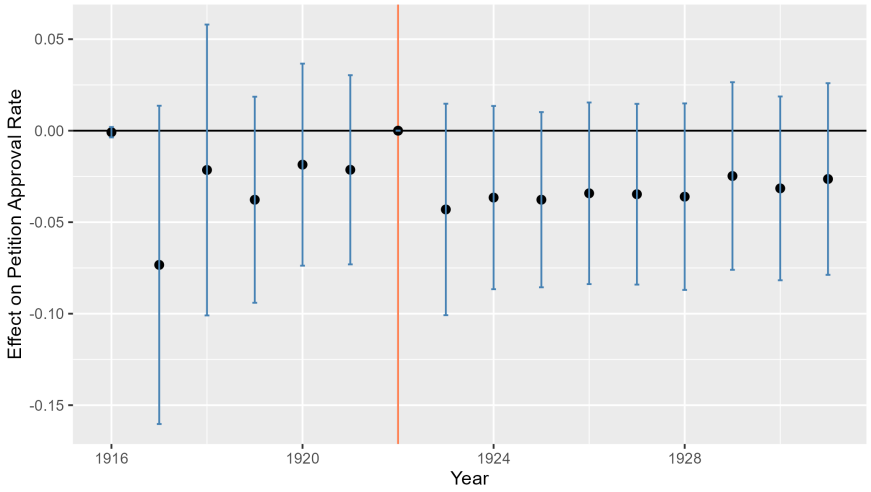
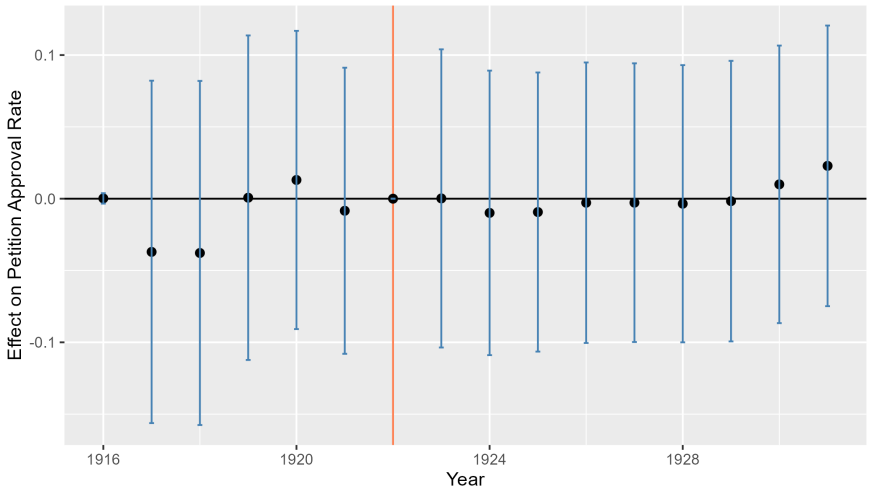


Figure 22: Dynamic DiD: Effect of Rural Schools on Indigenous Languages, Females

A.2.3 Additional Land Reform Figures

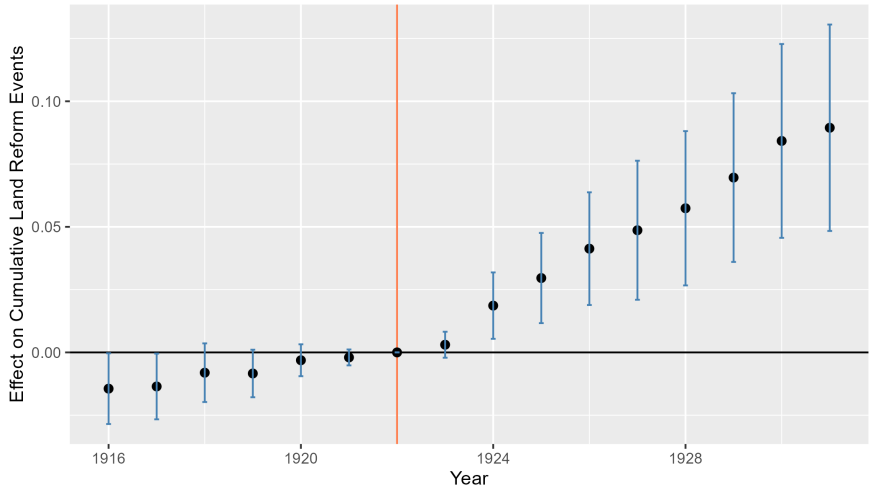


(a) All Municipalities

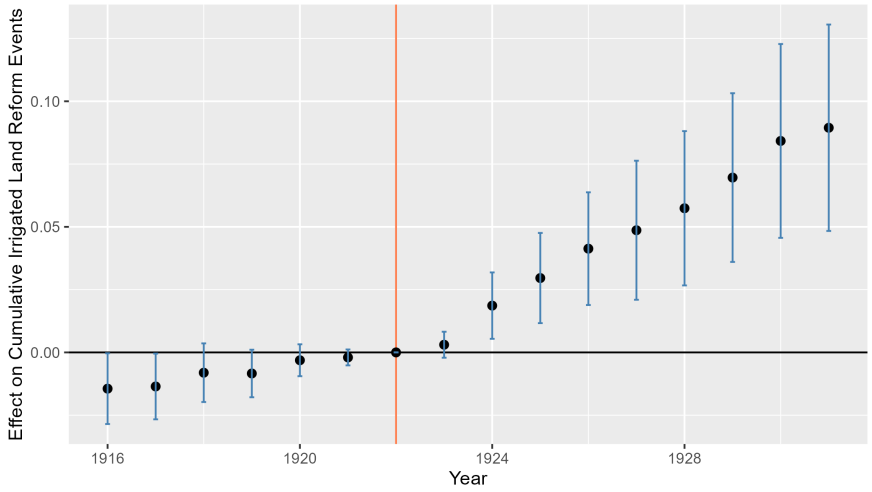


(b) Municipalities with Haciendas

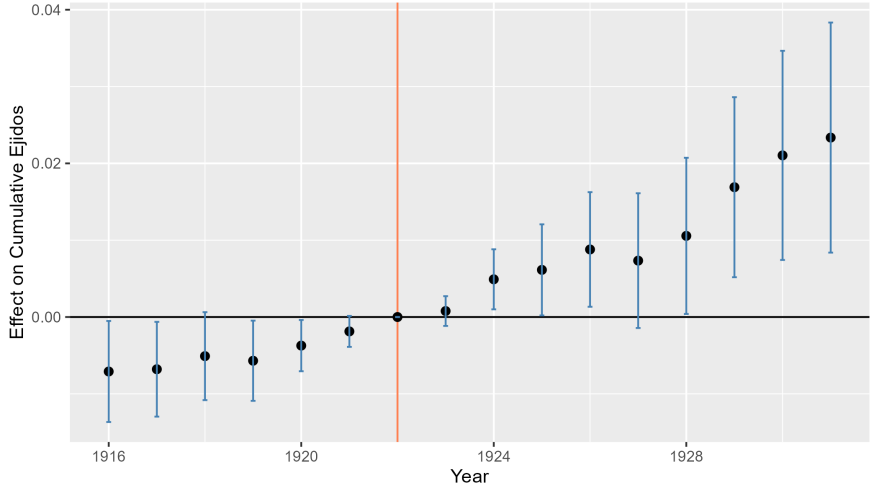
Figure 23: Municipality-Level Correlation: Rural Schools per 10000 and Petition Approval Rate



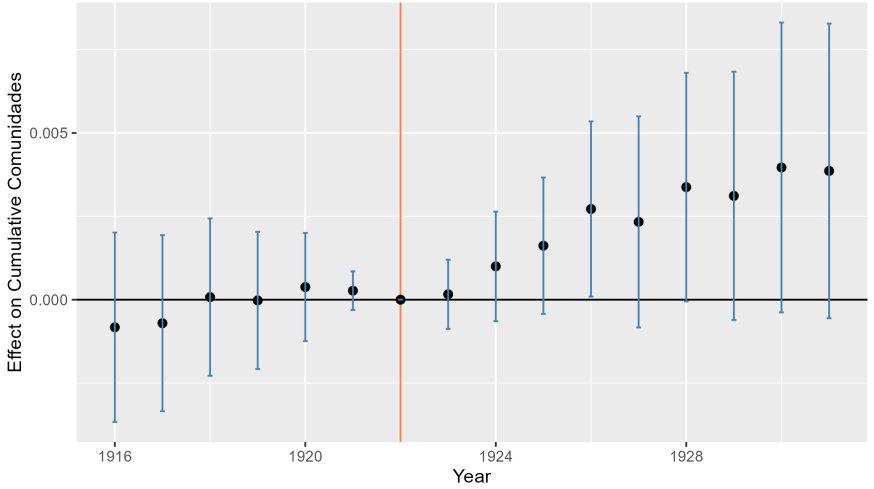
(a) All Events



(b) Irrigated Land



(c) *Ejidos*



(d) *Comunidades*

Figure 24: All Municipalities: Rural Schools per 10000 and Land Reform