Assimilation Policy, Integration, and Identity
Evidence from American Indian Boarding Schools

Christian Maruthiah [†]

November 14, 2025

Click here for latest version

Abstract

This paper examines the effects of coercive assimilation policies on integration and cultural attachment among ethnic minorities. I study the off-reservation school system, a historical policy under which U.S. authorities removed Native American children from their communities to distant boarding schools for extended periods of time. I exploit the staggered recruitment patterns of schools and variation in cohort exposure to facilitate causal identification. I show that the off-reservation school system led to cultural assimilation and—to a lesser extent—economic integration among cohorts exposed to the programme. However, I also provide evidence that these effects did not come at the expense of identity or cultural attachment, as the same cohorts maintained an ongoing connection to their home communities. In line with the historical literature, my results suggest that the policy may have strengthened the identities it sought to displace, highlighting the agency of Native American individuals and communities at the time.

JEL Codes: I20, J15, J18, N31, Z1.

[†]Trinity College Dublin. Email: christian.maruthiah@tcd.ie. I am grateful to Monica Martinez-Bravo for guidance and advice on this project. I have also benefited from discussions with and feedback from Dmitry Arkhangelsky, Ann Carlos, Eric Chyn, Donn Feir, Claudio Ferraz, Rob Gillezeau, Kareem Haggag, Gaia Narciso, Nathan Nunn, Alexander Persaud, Diego Ramos-Toro, Munir Squires, Liyang Sun and Tom Zohar. I am grateful for feedback from participants at the 2024 NBER Summer Institute (Development of the American Economy), the 2024 NBER Meeting on the Economics of Mobility, the 2024 CEPR Economic History Annual Symposium, the Virtual Economic History Seminar (University of Warwick), and the Workshop in Economic History (Uppsala University). Significant progress was made on this project during my stay at the University of British Columbia, and I thank staff, faculty and graduate students for their hospitality. I thank Evelina Angelopoulou, Edoardo Maifreni, Alejandro Mizrahi and Francesca Pagnotta, for their assistance in digitising various historical data sources. I acknowledge financial support from the Maria de Maeztu Unit of Excellence CEMFI MDM-2016-0684, funded by MCIN/AEI/10.13039/501100011033, and CEMFI.

A great general has said that the only good Indian is a dead one... I agree with the sentiment, but only in this: that all the Indian there is in the race should be dead. Kill the Indian in him, and save the man.

– Richard Henry Pratt, Nineteenth Annual Conference of Charities and Correction, 1892

We see a monument of the Indian in New York harbor as a memorial of his vanishing race. The Indian wants no such memorial monument, for he is not yet dead.

– Chauncey Yellow Robe, Annual Conference of the Society of American Indians, 1914

1 Introduction

Governments, colonial powers, and other state actors have frequently sought to reshape the attitudes, behaviours, and identities of their subjects—characteristics that are key inputs in economic and social decision-making (Akerlof and Kranton, 2000). These efforts have involved varying degrees of coercion—from curriculum changes (e.g., Cantoni et al., 2017), to language policy (e.g., Clots-Figueras and Masella, 2013), to restrictions on cultural practices (e.g., Abdelgadir and Fouka, 2020). Such efforts have typically targeted vulnerable, disenfranchised, or otherwise marginalised groups, such as immigrants and Indigenous populations. While there is a large literature in economics that explores the socioeconomic and cultural trajectories of immigrants (Abramitzky and Boustan, 2024) and the effects of state policies on immigrant integration (Fouka, 2024), we still have a relatively limited understanding of how assimilation efforts a century or more earlier were experienced by Indigenous populations in the same countries. This is a critical gap in our knowledge, given that these populations were targeted by policies that were among the most coercive assimilation efforts in modern history.

One such policy—implemented in Australia, Canada, and the United States, among other countries—has been the removal of Indigenous children from their families and communities for (re-)education in boarding schools.[1] These policies generally sought to

[1]Indigenous children were removed to missions or adoptive families in Australia, to Church-run "Indian residential schools" in Canada, and to state-run "off-reservation schools" in the United States. On Australia

replace Indigenous languages, religions and customs with those of the dominant society, with the justification that assimilation would lead to economic self-sufficiency (Smith, 2009). In the economics literature, modern assessments of Indigenous boarding school policies have focused on the long-term effects of such programmes. Broadly, these studies have found that Indigenous boarding school policies generated economic and educational benefits, but also led to a loss of cultural connection. However—due to a lack of historical data—there has been no quantitative analysis on the responses of Indigenous individuals and communities to boarding schools at the time these programmes were operational. Based on the historical literature, these effects are likely to have been nuanced in a way that is hard to capture with long-run data. For example, Indigenous boarding schools in the United States are generally seen to have been successful in promoting assimilation, to the point of generating tension between returned students and their communities (Adams, 2020, p. 303). But some scholars have also posited that these schools served the unintended purpose of strengthening Native American identity (Hertzberg, 1971; Nagel, 1996). An empirical investigation of the effects of Indigenous boarding school programmes in their historical context is important for understanding the responses of the Indigenous individuals and communities that were targeted, the historical legacy of such policies, and the effects of coercive assimilation efforts more generally.[2]

and Canada see ABC News (2008) and CBC News (2008) respectively. Indigenous boarding schools were recently the subject of a federal investigation in the United States (U.S. Department of the Interior, 2024). Even today, it is alleged that policies of a similar spirit are being implemented in China (BBC News, 2019) and in Russian-occupied areas of Ukraine (BBC News, 2025).

[2]The nuanced response of Indigenous populations to coercive assimilation efforts is well illustrated by the case of Chauncey (born Canowicakte) Yellow Robe. A "full blood" Lakota Sioux from South Dakota, Yellow Robe entered the Carlisle Indian School in Pennsylvania in 1883, in "full Indian costume... not knowing a word of English, not having seen a book or a schoolhouse before" (cited in Weinberg, 2004, p. 17). He remained at Carlisle—where he took the name "Chauncey"—until 1894, subsequently settled in Rapid City, South Dakota, and married a White American. However, while seamlessly assimilating into Western society, Yellow Robe also maintained a strong connection with his Indigenous identity. He actively sought to familiarise his daughters with the Lakota language (Weinberg, 2004, p. 29), joined the Society of American Indians (the first Native American-run civil rights organisation in the United States), and maintained an ongoing connection with his home community (as measured through his appearance in local "Indian censuses" into the 1920s). Yellow Robe's eldest daughter, Rosebud, went on to become a prominent Native American folklorist and

This paper examines the effects of the above-mentioned programme in the United States on cultural and economic integration among Native Americans. The policy—known as the "off-reservation school system"—involved the removal of Native American children from their communities (reservations) to distant boarding schools for periods of at least 3 to 5 years. The first off-reservation boarding schools opened in the late 1870s, and were rolled out across the country over the next three decades. Once children entered the schools, educators sought to completely reshape their identities, by banning the use of tribal languages and the practice of tribal religions, and promoting Western cultural practices over their tribal alternatives. Off-reservation schools were considered to be more effective at achieving cultural assimilation (and by extension, economic integration) than existing schools located on or near reservations, since children at the latter schools could still be visited and influenced by their families and communities.

As a first step towards studying the effects of the off-reservation school system, I make three data contributions. First, I systematically digitise new, person-level data on Native Americans collected in the 1910 census (covering over 240,000 records), and link this information to individual identifiers in IPUMS' publicly-available version of the same census. This new dataset—which includes information on reservation, tribe, percentage of "Indian blood", and educational attainment—covers over 90 per cent of the Native American population in 1910. Second, using person-level information on tribes and historical locations, along with a newly-digitised map of reservations in the early twentieth century, I match around 90 per cent of the 1910 Native American population under the supervision of the Indian Office at the time (i.e., with reservations) to a unique reservation. Recent work in Indigenous economic history has measured outcomes at the tribe-level (Feir, Gillezeau and Jones, 2023), or approximated the boundaries of historical reservations with modern ones to compute reservation-level outcomes (Miller, 2024). The former strategy is not well

educator. Clearly, none of this is observable in modern-day, area-level data.

suited to studying the effects of the off-reservation school system, since the policy was applied at the (historical) reservation level which—due to forced relocations throughout the nineteenth century—often did not correspond to Native homelands. The latter strategy is also imprecise, given that modern-day boundaries have been significantly altered since the early twentieth century—both diminished (e.g., due to allotment) and expanded (e.g., due to the restoration or purchase of land)—and because a non-negligible share of the Native American population did not reside on their "home" reservation in 1910 (a fact that I am able to document, and address, with person-level information on tribal affiliation). Third, using a combination of historical school reports and attendance records, I determine the year (if any) that each reservation in my dataset was exposed to the off-reservation school system. This constitutes the first reservation-level dataset on off-reservation school recruitment patterns during their rollout.

A simple comparison between treated and non-treated reservations, or treated and non-treated cohorts, is unlikely to be informative about the causal effects of the off-reservation school system. For example, off-reservation schools may have targeted reservations that were more open to assimilation (confounding a simple cross-reservation comparison), and younger cohorts were likely to have greater exposure to White Americans (confounding a simple cross-cohort comparison). To address these concerns, I use an event study design that exploits two sources of variation: first, that reservations were exposed to off-reservation schools at different points in time, and second, that individuals already past schooling age were less likely to be recruited. Specifically, as I observe an individual's reservation and age in my data, as well as the year that their reservation was treated, I can infer exposure on the basis of age when an off-reservation school *first* started recruiting from their reservation.[3] This strategy allows me to account for unobservable reservation-level characteristics, as

[3]This approach is informed by recommended schooling ages from historical publications by the Indian Office, as well as attendance records for selected off-reservation schools.

well as time-varying trends common to all cohorts, through the inclusion of reservation and cohort fixed effects. I am therefore able to estimate the causal effects of exposure to off-reservation schools at the reservation-by-cohort level.

I examine the effects of the off-reservation school system on educational, cultural, and economic outcomes, typically measured in 1910. With respect to educational outcomes—and in part, as a validation of my empirical strategy—I show that the first generation of Native American children exposed to off-reservation schools was more likely to have graduated from such a school. Effects with respect to literacy and English proficiency—while positive—are muted, suggesting that the off-reservation school system did not offer substantial educational benefits to Native American communities already served by (on-)reservation alternatives. I then show that exposure to the off-reservation school system led to marked cultural assimilation. I use a combination of algorithmic string-matching techniques and large language model-based evaluation to classify the roughly 55,000 unique first names of Native Americans in the 1910 census as being of "Western" origin (e.g., "John"), "Mixed" origin (e.g., "Ha-Gah-Kah John"), or "non-Western" origin (e.g., "Naun Gaun Way We"), and show that treated cohorts were more likely to have adopted a "Western" first name. I also find that treated cohorts were more likely to be classified as being of mixed European and Native American ancestry (i.e., "mixed blood" rather than "full blood")—a change that cannot be explained by intermarriage or demographic factors. My results with respect to identity—discussed below—suggest that effects on blood quantum reporting reflected the perceptions of census enumerators rather than changes in self-identification among Native Americans. Being perceived as "mixed blood" is likely to have reflected a variety of assimilation-relevant factors (e.g., a person's accent and style of dress), and therefore provides perhaps the starkest evidence on the effects of the off-reservation school system on cultural assimilation. With respect to economic outcomes, there is suggestive evidence that treated cohorts were more integrated with the

(Western) labour market, but otherwise no clear indication that they obtained meaningful economic benefits from the schools.

While I find that off-reservation schools were effective in generating cultural and economic integration, an important remaining question is whether they were equally effective in displacing Indigenous identities—in the words of Richard Henry Pratt, to "kill the Indian". Here, my findings suggest that assimilation did not come at the expense of ethnic identity or cultural and community attachment.

First, I show that—despite no difference in linkage rates between the 1910 population census and the 1930 population census—treated cohorts were more likely to be linked to local "Indian censuses" taken at reservations in the early 1930s—indicative of ongoing attachment to or presence at their home communities. In line with insights from the sociological and Native American studies literature (e.g., Nagel, 1996; Frederick, 1993), I provide evidence that exposure to the off-reservation school system strengthened ethnic identities in communities where these identities were likely to be more malleable at baseline. In particular, I find that the effects with respect to appearance in 1930s Indian censuses were stronger on ethnically diverse reservations (where intermarriage between tribes plausibly blurred ethnic boundaries), and on reservations with comparatively high (pre-treatment) mixed-race populations. Given that tribal constitutions written from the 1930s onwards regularly used Indian censuses to form "base rolls" for membership, my findings have implications for tribal membership in the past and today.[4]

Second, I digitise 1911 to 1915 membership lists of the Society of American Indians—the first Native American-run civil rights group in the United States (Clark, 2004). Founded in 1911—and with over 200 members at its peak—the Society of American Indians

[4]For example, Article II of the Constitution of the Oglala Sioux Tribe provides that "Membership of the Oglala Sioux Tribe shall be automatic when: (a) The person's name appears on the official roll of the Oglala Sioux Tribe of the Pine Ridge Reservation as of April 1, 1935... (b) A child is born to any member of the Oglala Sioux Tribe."

campaigned for improved educational opportunities, living conditions and civil rights for Native Americans. After hand-linking individuals from membership lists to the 1910 census, I find that treated cohorts were—at least suggestively—more likely to have been members of the organisation. Taken together, my findings suggest that off-reservation schools may have strengthened the identities they sought to displace, highlighting the agency and resilience of Native American individuals and communities at the time.

This paper contributes to three strands of literature. First, my work builds on research related to the socioeconomic and cultural consequences of Indigenous boarding school programmes.[5] This topic has received more attention in the historical and sociological literature (e.g., Adams, 2020; Lomawaima, 1994; Vuckovic, 2008), and has been the subject of numerous government reports (e.g., Australian Human Rights Commission, 1997; U.S. Department of the Interior, 2024). There are fewer studies that estimate the causal effects of exposure to Indigenous boarding schools. The most closely related work is Gregg (2018), which finds that modern-day reservations with greater historic exposure to off-reservation schools are more culturally assimilated (e.g., smaller family sizes and a greater proportion of individuals that exclusively speak English) and economically prosperous. Another two closely related studies are Feir (2016) and Jones (2024), which both examine the long-term effects of Canada's Indian Residential School system. All three studies are important and foundational works in the literature, but—given their focus on long-term outcomes—cannot speak to the effects of Indigenous boarding school programmes in their historical context. Furthermore, these studies measure exposure to boarding school programmes relatively late in the policy lifecycle. For example, the off-reservation boarding school shares used by Gregg (2018) only begin in the 1910s, 30 years after the start of the programme, and by

[5]There is also a growing literature in economics that studies the effects of other historical policies targeting Indigenous populations (e.g., reservation formation, the elimination of traditional food sources, and self-governance) on contemporary outcomes (respectively, Dippel, 2014; Feir, Gillezeau and Jones, 2023; Frye and Parker, 2021).

which time the focus of Indian education policy had shifted to on-reservation and public schools (Adams, 2020, p. 348). By focusing on the rollout of the off-reservation school system, I am also able to starkly document the cultural effects of the programme on cohorts with varying levels of exposure using an event study design, which is not possible with long-run data.

Second, my work is related to the large literature in economic history that studies the cultural assimilation and economic integration of immigrants to the United States from the mid nineteenth to early twentieth century (e.g., Abramitzky, Boustan and Eriksson, 2014; Abramitzky, Boustan and Erikkson, 2020), and the role of state policies in helping or hindering this process (e.g., Fouka, 2019). A key barrier to studying assimilation patterns in this literature is that information on many relevant dimensions of cultural assimilation, such as a person's accent or form of dress, has not been systematically collected, which is a particular limitation when studying the assimilation of first generation immigrants (Abramitzky, Boustan and Erikkson, 2020). In this paper, I propose a measure of cultural assimilation that arguably captures such unobserved dimensions of assimilation: whether a Native American individual was perceived (and therefore recorded in the census) as "mixed blood" by census enumerators. Furthermore, I focus on a population that—in contrast to European immigrants—was largely unfamiliar with Western cultural practices and traditions (Adams, 2020, pp. 109-121), and which experienced a degree of coercion—from forced relocations, to the confiscation of land, to the removal of children to boarding schools—well beyond that experienced by immigrants at the time.

Third, this paper adds to work on racial identification in historical U.S. censuses (e.g., Abramitzky et al., 2023; Adukia et al., 2025). Among these studies, research by Dahis, Nix and Qian (2020) examines the phenomenon of "passing" by African Americans in the early twentieth century. Linking across census years from 1880 to 1940, they find that a non-negligible share of African American men "passed" (i.e., switched race from "Black"

to "White"), and that the tendency to pass was influenced by discrimination, educational possibilities, and employment opportunities. The present paper builds on this and similar analyses of the construction and malleability of race in administrative data by estimating the causal effects of historical policies on racial classification in such data, and highlighting the potential endogeneity of categorisations *within* racial groups (in my setting, degree of Indian ancestry) during the early twentieth century.

The remainder of this paper is structured as follows. Section 2 provides context on the off-reservation school system and details on the enumeration of Native Americans in historical censuses. Section 3 outlines the main data sources used in my analysis. Section 4 presents my empirical strategy. Section 5 outlines my results, along with robustness analysis. Section 6 concludes.

2 Context

When Europeans first arrived in North America, the Indigenous population is estimated to have been between 2 and 5 million, with more than 1000 distinct communities (Nagel, 1996, p. 4). Over the next two centuries, diseases, wars, the loss of traditional food sources and forced relocation decimated the Native American population. By the mid nineteenth century, Native American military resistance had been overcome and populations were largely confined to reservations under the administration of federal "Indian agents".[6] Native Americans were now considered "wards of the government; the duty of the latter being to protect them, to educate them in industry, the arts of civilization... to sustain and clothe them until they can support themselves" (Board of Indian Commissioners, 1869, p. 10).

[6]Indian agents were federal employees responsible for the day-to-day administration of Native American reservations. Among other things, Indian agents were responsible for disbursing food rations, building and maintaining infrastructure, and enrolling children at schools. On average, agencies had jurisdiction over 1.5 reservations (Gregg, 2018).

2.1 *Indian Education Policy Prior to the Off-reservation School System*

From the second half of the nineteenth century, the federal government became increasingly involved in the provision of education to Native Americans (Vuckovic, 2008, p. 12). Education policy had four main aims: to provide a basic academic education, to give Native Americans training in practical skills and trades, to encourage cultural assimilation, and to promote Christianity over tribal religions (Adams, 2020, pp. 24-29).

Two forms of schools existed prior to the development of the off-reservation school system. The first form was the reservation day school. These schools were located near villages, with children returning to their families at the end of each school day. Policymakers soon came to the view that day schools did not sufficiently separate children from the influence and practices of their families and communities. In 1878, for example, the Indian agent of the Shoshone and Bannock Agency (Wyoming) opined that placing Native American children "under a teacher's care but four or five hours a day... to spend the other nineteen in the filth and degradation of the village, makes the attempt to educate and civilize them a mere farce" (Office of Indian Affairs, 1878, cited in Adams, 2020, p. 34). The second form of schools, developed in the 1870s in response to the above concerns, were reservation boarding schools. These were also located on or near reservations. Children lived at reservation boarding schools during the school term, but returned home for vacations. By 1879, when the first off-reservation schools began taking students, there were 107 day schools and 52 reservation boarding schools in 14 states. The total attendance across all schools was 4,448 students, or between 7 - 10 per cent of school-age Native American children (Office of Indian Affairs, 1909, p. 87).[7] A breakdown of attendance between boarding and day schools is only available from 1882 onwards. In 1882, reservation

[7]School-age population is calculated from agency-level statistical tables in Office of Indian Affairs (1879). Since some agencies did not report school-age population, I estimate school-age population as 20 per cent of total agency population.

boarding schools accounted for 50 per cent of total school attendance, reservation day schools accounted for 30 per cent, and off-reservation schools (at the time the Carlisle Indian School, and the Chemawa Indian School) accounted for 5 per cent. The remainder attended non-government operated contract schools.

While children remained in reservation boarding schools for around nine months of the year, policymakers eventually concluded that these schools also failed to sufficiently remove children from tribal influences. Children were found to "relapse" into tribal ways during vacation periods, and parents were still able to easily visit their children (Adams, 2020, p. 37).

2.2 The Rise of the Off-reservation School System

The first off-reservation school, the Carlisle Indian School (Pennsylvania), was opened in 1879. Carlisle was the culmination of efforts by an Army officer, Captain Richard Henry Pratt, to develop a new model of education that could rapidly assimilate Native American children into mainstream society. Under Pratt's vision, children would be removed from the influence of the reservation, where they could be completely immersed in Western society for an extended period of time. Policymakers saw promise in this model, and new schools opened rapidly over the next 20 years, stabilising at 25 schools in the early 1900s (Figure 1). By 1912, when the last off-reservation school was opened in Tacoma (Washington), there were 27 schools in operation. Figure 2 shows the distribution of off-reservation schools with respect to Native American reservations as at the early twentieth century. Appendix Table A.6 shows the locations and opening years of off-reservation schools from 1879 until 1912.

2.3 Off-reservation School Recruitment Practices

The recruitment practices of off-reservation schools varied from school to school and across time, but did follow some general principles. With respect to coercion, some (but

not all) children were forcibly taken in the early years of the off-reservation system. This practice was banned in 1893, when the Commissioner of Indian Affairs ruled that parental consent was required to send children to schools outside the reservation (Adams, 2020, p. 71).[8] In theory, coercion was further restricted in 1894 when Congress banned Indian agents and other government employees from obtaining parental consent "by withholding rations or by other improper means" (cited in Prucha, 1984, p. 905). However, the extent to which these laws were respected is unclear. Indeed, the fact that rules prohibiting forced removals needed to be re-issued in 1917 suggests that such practices continued well after 1893 (Lomawaima, 1994, p. 36).

The school age designated for Native American children was set at 6 to 16 years (Office of Indian Affairs, 1890, p. 452).[9] Off-reservation school administrators were encouraged to target individuals at the upper end of this band, who had already received some education. That said, individuals above and below these bands were recruited by off-reservation schools. The Commissioner of Indian Affairs recommended an upper limit of 20 years in 1902 (Office of Indian Affairs, 1902).[10]

Off-reservation schools were also encouraged to target "full blood' children (i.e. those without, or with very little, European ancestry), as they were considered most in need of assimilation. However, administrative reports from the time suggest that a substantial number of students had at least some European ancestry (e.g., Office of Indian Affairs, 1897, p. 319). A preference for students with a low percentage of "White blood" and/or those "living in Indian fashion" was formalised in 1902. Finally, with the exception of two large schools that recruited across the entire country (Carlisle and the Haskell Institute, Kansas),

[8]At the time, the Commissioner justified the ruling on the basis that "even ignorant and superstitious parents have rights" (cited in Adams, 2020, p. 71).

[9]This band was widened to 5 to 18 years in 1891 (Office of Indian Affairs, 1891, p. 511).

[10]The Circular states: "The enrollment of young Indian men and women... will only be permitted when valid reasons are assigned. When such persons have reached the age of 20 years, it is ordinarily time for them to cease leaning on the arm of government and endeavor to make a living for themselves".

off-reservation schools tended to recruit from within their state or from adjacent states. In part, this practice was motivated by concerns about the costs of transporting children between their communities and off-reservation schools (Gregg, 2018). By the mid-1890s, the Office of Indian Affairs began to formally restrict off-reservation school recruiting zones, which typically covered the school's state, and for larger schools, surrounding states.

2.4 Decline of the Off-reservation School System

While enrolment in off-reservation schools peaked in 1915 and remained relatively stable for the next 15 years, the model began to fall out of favour in the early twentieth century. Policymakers came to see the goal of rapid assimilation as unfeasible, and the practice of separating children from their families as cruel (Adams, 2020, pp. 338-339). In 1907, then-Commissioner of Indian Affairs Francis Leupp argued for closing off-reservation schools and enlarging the day schools system (Adams, 2020, p. 348), and from 1908 off-reservation schools were banned from directly recruiting students at reservations (though parents that wished to send their children to such schools could still do so) (Prucha, 1984, p. 820). Between 1900 and 1925 the number of off-reservation schools fell from 25 to 18 (including the closure of Carlisle). A comprehensive review of federal Indian policy published in 1928 was highly critical of off-reservation schools, noting a lack of nutrition, poor sanitary conditions, and the fact that the industrial training offered at the schools was irrelevant to students that returned to their reservations (Institute for Government Research, 1928). In response to the review, another round of school closures occurred over the next decade. While this did not signal the end of the off-reservation school system, it did represent the end of their role in the aggressive assimilation of Native American children.

2.5 *Off-reservation Schools and Their Alternatives*

In order to understand how off-reservation schools may have affected Native American children differently from on-reservation alternatives, it is helpful to highlight the similarities and differences between off-reservation boarding schools and alternatives on or near reservations. As discussed above, there were two types of schools on reservations: day schools and reservation boarding schools.

Due to their isolation, small size, few teachers (generally one teacher and their assistant) and lack of furnishings, day schools typically offered a poorer academic education than reservation and off-reservation boarding schools, with instruction mostly at the primary grades (Office of Indian Affairs, 1890, p. XIII and p. CLIV). Initially, the academic education provided at reservation and off-reservation boarding schools was generally at a similar level. As government-operated institutions, off-reservation schools followed a similar curriculum to reservation boarding schools.[11] While the largest off-reservation schools, such as Carlisle, eventually offered a post-primary education, this was not significantly more advanced than the education available at large reservation schools.[12] Pratt himself stated that Carlisle "would not attempt even a high school education... graduation was fixed between the grammar and high school grades of our public schools" (Pratt, 1912, p. 13). Off-reservation schools only began to offer a high school curriculum in the 1920s (Gregg, 2018).

Historical administrative data support the view that the quality of education at reservation boarding schools and off-reservation boarding schools was similar. School statistics

[11]The curriculum for Native Americans, known as the "course of study" was formalised in 1890 (Office of Indian Affairs, 1890, pp. CLVI-CLX).

[12]The 1890 Annual Report stated that off-reservation schools "are not universities, nor colleges, nor academies nor high schools. In the best of them the work done is not above that of an ordinary grammar school, while in most it is of the primary or intermediate grade. The pupils come to them for the most part ignorant of the English language, unaccustomed to study, impatient of restraint, and bringing, with them many of the vices and degraded habits of camp life" (Office of Indian Affairs, 1890, p. IX).

in the 1900 Annual Report to the Commissioner of Indian Affairs indicate that the cost per pupil at off-reservation boarding schools was around $148, while the cost per pupil at reservation boarding schools was slightly higher, at $151 (Office of Indian Affairs, 1890, p. 635). According to the same statistics, off-reservation schools reported 10 pupils per employee, while the corresponding figure at reservation boarding schools was 6 pupils per employee (the statistics do not distinguish between teachers and other employees). While these statistics are imperfect proxies of school quality, they at least suggest that the quality of education at off-reservation schools was not significantly higher than that at reservation boarding schools.[13]

The main differences between off-reservation and reservation boarding schools were geographical and operational. Firstly, while some off-reservation schools were closer to reservations than others, they were invariably located outside Native American communities. This limited the ability of families and communities to influence the education of their children. Secondly, while children at reservation boarding schools returned home at least once a year, children at off-reservation schools typically remained there, continuously, for three to five years. Thirdly, while reservation boarding schools recruited directly from reservations, and were therefore ethnically homogeneous, children from different tribes were deliberately mixed at off-reservation schools. This policy was, in part, intended to promote the use of English: while Indigenous languages were banned or discouraged in both reservation and off-reservation boarding schools, it is likely that such a rule was harder to enforce in schools where the same language was spoken by the entire student body (Adams, 2020, p. 154).[14]

[13] In line with the discussion in the preceding paragraph, both measures of school quality were lower at reservation day schools. Days schools operated at a cost per pupil of $48, and had 13 pupils per employee.

[14] Policymakers and school administrators were aware of the effects of combining children from linguistically diverse tribes. In 1896, the Superintendent of Wittenberg Indian School (Wisconsin) reported that "it has been and is one of the main principles of the school to have children from different tribes about equally divided... forcing the use of the English language" (Office of Indian Affairs, 1896, p. 47). Similarly, in 1898 the Superintendent of Phoenix Indian School (Arizona) opined that "the intermingling of different tongues is the

Another practice specific to off-reservation schools, and likely to promote cultural assimilation, was the "outing programme", which placed students with White families for up to a year (Adams, 2020, pp. 174-175). This programme was first implemented at Carlisle, and subsequently rolled out to other schools.[15]

Given the differences described above, one might expect that off-reservation schools were more effective than reservation boarding schools in breaking the cultural connection between Native American children and their families and communities.[16] This hypothesis is supported by historical reports from the Indian Office, as well as the personal accounts of returned students. For those that did return to reservations, there were sometimes clear cultural barriers between students and their families (Adams, 2020, p. 303). Students that had spent three to five years speaking English sometimes found it difficult to communicate with their families in their own languages (Child, 1993, p. 76). Cultural differences between returned students and their communities were also evident in episodes where tribes enforced compliance with rituals by way of economic sanctions and / or corporal punishment (Office of Indian Affairs, 1887, p. 168). At the same time, it is also possible that coercive efforts to suppress Native American identities at off-reservation schools strengthened the resolve of individuals and their communities to maintain these identities. There is evidence of such responses in the personal accounts of returned students. For example, upon returning to their reservations and realising they could not communicate comfortably with their parents, some returned students vowed to relearn their languages

surest and best way to teach English and broaden the tribal view" (Office of Indian Affairs, 1898, p. 365).

[15]Pratt considered the outing programme as "the best possible means of inducting Indian boys and girls into our civilized family and national life" (Office of Indian Affairs, 1897, p. 30).

[16]Off-reservation schools were generally seen as ineffective in promoting economic integration. After attending off-reservation schools, students typically returned to their home reservations, where they often had difficulties finding work (Office of Indian Affairs, 1898, p. 339). While students had received training in farming and various trades as part of their vocational education, the land on reservations was often unsuitable for the former, and there was limited demand for the latter (Adams, 2020, pp. 308-309). Authorities eventually sought to address these problems by providing returned students preferential access to clerical jobs, as well as establishing the Indian Employment Bureau in 1905 (Adams, 2020, p. 325).

(e.g., Rogers, 1974, cited in Child, 1993). Others actively sought to educate their own children in the language, customs and history of their tribes (e.g., Weinberg, 2004, p. 29). Furthermore, some graduates of off-reservation schools went on to actively campaign for Native American civil rights as members of the first such organisation in the United States, the Society of American Indians.

2.6 The Society of American Indians

The Society of American Indians (SAI) was formed in 1911 when Fayette A. McKenzie, a professor at the Ohio State University, arranged an initial meeting with six Native American professionals in Columbus, Ohio. A Temporary Executive Committee of 18 members was formed; of these, at least 10 had attended an off-reservation school (Hertzberg, 1971, p. 36). The SAI was the first Native American-run civil rights organisation in the United States; broadly, it campaigned for improved educational opportunities, living conditions and civil rights.

The SAI's first national conference was held in October 1911; over 50 Native American delegates attended. By 1913, the group had grown to include over 200 Native American members. It was at the 1913 conference that the SAI articulated its main platform, calling for (among other things) citizenship, reforms to the school system, and the opening of the Court of Claims to Native Americans. Membership subsequently declined as the SAI was affected by internal disagreements on policy and the onset of World War I. While the SAI disbanded in the early 1920s, many of the demands put forward in its 1913 platform were implemented over the next two decades (Hertzberg, 1971, p. 117). In addition, the SAI arguably created a platform for other Native American civil rights organisations (such as the National Congress of American Indians) that emerged later in the twentieth century.

The SAI's founders and early leadership came from a variety of cultural backgrounds, representing at least 10 tribes. Some had White ancestry and a familiarity with Western

customs. Others were "full blood" individuals that were among the first from their tribes to make the transition to Western society. Perhaps the most common feature among them was exposure to off-reservation schools (Hertzberg, 1971, pp. 38-49). Based on this observation, scholars have posited that off-reservation schools may have served the unintended purpose of strengthening Native American identity and associated activism (Nagel, 1996, p. 116).

3 Data

3.1 Data Sources

Annual Reports of the Commissioner of Indian Affairs: My main sources of historical administrative data are annual reports from the Indian Office—Annual Reports of the Commissioner of Indian Affairs (Annual Reports)—from 1879 to 1906. These reports—which were submitted to the Secretary of the Interior—provided an overview of the Indian Office's activities in the prior year, and included reports by local officials tasked with the supervision of reservations (Indian agents), as well as "superintendents" in charge of reservation and off-reservation schools. In addition, Annual Reports generally include schedules listing extant reservations and the tribes occupying these reservation in the year of the report.

Off-reservation School Attendance Records: Attendance records for off-reservation schools, provided they were kept, are stored in various National Archives facilities across the United States. These records have not been systematically digitised. However, I have obtained attendance records from four off-reservation schools: Carlisle (from the Carlisle Indian School Digital Resource Center), Hampton (from Brudvig, 1994), Chemawa, and Haskell (both digitised from images on FamilySearch.org). These data cover the universe of attendees at these schools from their opening dates until their closures (Carlisle and Hampton) or 1900 (Chemawa and Haskell). These four schools accounted for over 30 per

cent of attendance at off-reservation schools in 1900 (Office of Indian Affairs, 1900, p. 16). Records typically contain the names, ages (at entry), tribes and home agencies of attendees.

Historical Censuses: I draw on the publicly-available and restricted full count U.S. censuses for the years 1910, 1920, and 1930 from IPUMS (Ruggles et al., 2022, 2021). In principle, the censuses cover the entire population of the United States. The number of individuals classified as "Indian" in the censuses range from 280,000 (in 1910) to 328,000 (in 1930).[17] The publicly available censuses include information on an individual's location at the time of enumeration, as well as demographic and socioeconomic information (e.g., age, state of birth, marital status, and occupation). In addition to the variables in the publicly-available censuses, the restricted censuses include first names and surnames.

I also use new person-level information on Native Americans, collected as part of the 1910 census on the so-called "Indian schedules". In addition to the standard census variables described above, the Indian schedules contain Native American-specific information, including a person's tribe, degree of Native American ancestry (e.g., "mixed blood" or "full blood"), educational attainment, and housing conditions ("civilized" or "aboriginal"). I digitise this information and link the constructed variables to IPUMS person-level identifiers; the resulting dataset covers over 90 per cent of the Native American population in 1910.[18]

Finally, I collect the names, locations, and ages of all males appearing in local "Indian censuses" taken by Indian agents in over the period 1930 to 1934.[19] In principle, Indian censuses during this period included all individuals under the jurisdiction of the agency,

[17] I identify Native Americans using the RACE variable from IPUMS.

[18] Enumerators were instructed to use the Indian schedules "for the enumeration of Indians living on reservations or in tribal relations, and also... in certain counties containing a considerable number of Indians." (U.S. Census Bureau, 2002). Therefore, while the vast majority of Native Americans were enumerated using the Indian schedules, coverage is not complete.

[19] The collection of Indian censuses is known as the Indian Census Rolls. The Rolls can be accessed through FamilySearch.org and Ancestry.com.

including those living on allotments or surrounding settlements (National Archives and Records Administration, 2014). Importantly, from 1930 onwards Indian agents were specifically instructed to remove the names of individuals whose whereabouts had been unknown for some time, meaning that appearance in an Indian census was indicative of an ongoing attachment to or presence at the agency. Further information on the Indian censuses is available in Appendix A.

Geocoded Historical Locations: In order to identify locations recorded in historical censuses, I use a database of geolocated places published by the Census Place Project (Berkes, Karger and Nencka, 2023). This dataset consists of crosswalks between individual identifiers in IPUMS' public releases of full count U.S. censuses and geographic coordinates (latitudes and longitudes), constructed by underlying string variables in historical censuses, and matching them to (already geocoded) historical places from IPUMS' NHGIS (Manson et al., 2023) or the Geographic Names Information System (United States Geological Survey, 2021). On average, the Census Place Project crosswalks successfully match 94 per cent of individuals to a sub-county location for the 1850 to 1940 census waves.

Membership of the Society of American Indians: I obtain a list of members of the SAI from two sources. First, I draw on a list of around 100 members in 1911 compiled by Clark (2004). Second, I digitise membership lists from three volumes of the Quarterly Journal of the Society of American Indians, for the years 1913, 1914 and 1915. There are roughly 400 unique members across the years 1911 to 1915.

3.2 Data Construction

Matching Native Americans to Reservations and Settlements in 1910: Around 260,000 Native Americans were counted in continental U.S. states in the 1910 census—the first census in which the majority of the first generation to be exposed to off-reservation schools

had reached adulthood.[20] I use newly-digitised information from the 1910 Indian schedules, as well as geocoded locations from the 1910 census, to match over 90 per cent of 1910 Native American population to a unique reservation or non-reservation settlement.

I primarily match on the basis of tribe (as reported in the Indian schedules) and state of residence in 1910, using a tribe-to-reservation correspondence from the 1909 Annual Report. This method allows me to account for cases where (different bands of) the same tribe occupied reservations in different states, as was often the case (e.g., Cheyenne on the Cheyenne and Arapaho reservation in Oklahoma and Cheyenne on the Tongue River reservation in Montana). To increase the rigour of the matching procedure, I match Native Americans to (geocoded) locations from the Census Place Project, and overlay these points on digitised 1910 reservation boundaries. I then require that individuals matched to a reservation on the basis of their tribe resided within 50 kilometres of that reservation.[21] In some cases, individual sheets of the Indian schedules record the name of a reservation, which allows me to match all individuals on that sheet to the reservation. Combining tribe-based and reservation-based matches, I am able to match around 190,000 individuals (73 per cent of the Native American population) to a unique reservation.

I complement the method above by matching on the basis of 1910 location alone, as in Dippel and Frye (2020). This method is necessary in cases where information on tribe is not available (either because this information is not legible, or because the Indian schedule was not used at all). As above, I match Native Americans to (geocoded) locations from the Census Place Project, and overlay these points on digitised 1910 reservation boundaries. All individuals residing in a place that can be matched to a unique reservation within a distance of 20 kilometres are assigned to that reservation. I am able to match around 194,000 of individuals to a unique reservation on the basis of their location.

[20] This figure excludes roughly 20,000 Alaska Natives.

[21] Without this restriction, Shoshoni living in the vicinity of Elko, Nevada (which is today a reservation) would be erroneously matched to Duck Valley reservation some 100 kilometres to the north.

To account for Native American populations that did not fall under the jurisdiction of the Indian Office in 1910, I use information on tribe and location to identify (non-reservation) settlements. Firstly, I identify locations from the Census Place Project with at least 25 Native Americans of the same tribe who are not within 50 kilometres of a historical reservation. I then identify connected components (i.e., settlements that can be reached through a chain of overlapping buffers).[22] Using this method, I assign around 30,000 Native Americans to over 100 non-reservation settlements. These settlements include communities that were not formal reservations at the time, but are today, such as the Penobscot residing near Indian Island (now the Penobscot Reservation), and the Choctaw residing in Neshoba County, Mississippi (now the Mississippi Choctaw Reservation).

Finally, I combined tribe-based matches, location-based matches, and non-reservation settlement assignments. If an individual is matched using one method, but not the other, I keep the unique match. If an individual is matched to different reservations using the two methods, I prioritise the tribe-based match, which more accurately reflects the individual's "home" reservation. In total, I am able to match 92 per cent of Native American population in 1910 to a unique reservation or non-reservation settlement.

Off-reservation School Treatment Years: To construct a dataset on reservation-level exposure to the off-reservation school system, I begin by reading all off-reservation school reports contained in Annual Reports from 1879 to 1906. I use information on tribes, reservations, or agencies contained in school reports, along with tribe-to-reservation schedule from 1909, to identify the years (if any) that each reservation in my dataset sent students to any off-reservation school. In cases where the information contained in school reports does not identify a unique reservation, I triangulate between school reports, agency reports, and attendance data for Carlisle, Chemawa, Hampton, and Haskell.

[22] I use a buffer of 25 kilometres, or a 5 to 10 hour walking distance.

I take the first year that any off-reservation school visited a reservation as that reservation's treatment year. I am able to identify around 130 reservations that were treated on or before 1910 (out of around 150 reservations in my dataset). Figure 3 shows the number of reservations treated for the first time in a given year, and the cumulative number of reservations that were treated up to and including that year. Following a initial concentration of treatment activity in the early 1800s—corresponding to the opening of Carlisle and Chemawa, and the introduction of Hampton's Indian programme—treatment years are fairly evenly spread over 1885 to 1900, which is consistent with the gradual rollout of off-reservation schools across the country over this period.

Combining and Linking Datasets: In order to construct my main dataset, I combine individual-to-reservation matches in 1910 with reservation treatment years, assigning a treatment year to each individual on the basis of their reservation. I then link individuals from this dataset to 1930s Indian censuses using a standard algorithm in the economic history literature developed by (Abramitzky, Boustan and Eriksson, 2019) (ABE-JW).The ABE-JW algorithm requires exact matches on state of residence, allows for minor spelling differences in names, and for birth years to differ by +/- 2 years between records.[23]

In order to examine naming practices, I classify the roughly 55,000 unique first names of Native Americans in the 1910 census as being of "Western" origin (e.g., "Mary", "Mary Jane"), "non-Western" origin (e.g., "Naun Gaun Way We"), or "Mixed" origin (e.g., "Ha-Gah-Kah John"). My classification procedure combines algorithmic string-matching techniques with large language model-based evaluation to determine whether names are of European origin. Further details on the procedure, along with the full large language model prompt, are available in Appendix B.

Finally, I hand-link individuals from the 1911 to 1915 SAI membership lists to IPUMS'

[23] I use publicly available codes from the Census Linking Project (https://censuslinkingproject.org).

restricted release of the 1910 census. I generally match on the basis of name and state of residence in 1910, but also use additional information (e.g., age, occupation) to distinguish between multiple potential matches. I am able to find unique matches for just over 100 male members (out of the roughly 200 unique male members).

4 Empirics

4.1 Identification

My goal is to estimate the effects of exposure to the off-reservation school system during the late nineteenth century on the educational, economic, and cultural trajectories of the first generation of Native American children exposed to the programme.

Simple comparisons between treated and non-treated reservations, or younger and older cohorts, would be problematic in the presence of unobservable reservation characteristics or cohort trends that are correlated with the outcomes of interest. In order to estimate the causal effects of exposure to the off-reservation school system, I adopt a cohort-based event study design. This design exploits two sources of variation: firstly, that reservations were exposed to the off-reservation school system at different points in time, or not at all (cross-sectional variation), and secondly, that individuals already past schooling age when an off-reservation school first came to their reservation were less likely to attend (cohort variation).

With respect to cross-sectional variation, I identify the first year that each reservation or settlement in my dataset sent students to an off-reservation school (Figure 3) to identify "ever treated" and "never treated" reservations. Reservations are assigned to the "ever treated" group if they appear in my reservation treatment year dataset. Reservations or settlements are assigned to the "never treated" group if they do not appear in the reservation treatment year dataset, and there are no individuals on the reservation or settlement that graduated from an off-reservation school.

With respect to cohort variation, I identify cohorts that, on the basis of their age, were unlikely to be recruited when an off-reservation school first came to their reservation. Communications from the Office of Indian Affairs (e.g., Office of Indian Affairs, 1902) indicate that—by the early twentieth century—individuals over the age of 20 were deemed to be "too old" to attend an Indian school. However, Figure 4—constructed using observed ages at admission in attendance records from Carlisle, Chemawa, and Haskell—shows that these limits were not strictly applied, with a non-negligible number of admissions of individuals up to age 25. This was especially true in the early years of the programme, when age guidelines had not yet been formulated. Guided by Figure 4, I assume that individuals aged over 26 (0.5 per cent of the sample) constituted the first cohort that was "too old" to be recruited to an off-reservation school.

4.2 Specification

My main specification takes the following form:

$$y_{i,r,c} = \sum_{j;\, j \neq 28} \beta_j \, age\ at\ exposure_{j(r,c)} + \alpha_r + \alpha_c + X_i' \gamma + \varepsilon_{i,r,c} \tag{1}$$

Here, $y_{i,r,c}$ is the outcome (measured in 1910, 1920, or 1930) of individual i from reservation r of birth cohort c. $age\ at\ exposure_{j(r,c)}$ are event time indicators for the age of cohort c from reservation r when an off-reservation school first recruited from the reservation. In order to increase precision, I group these indicators into 2-year bins (e.g., $j = 28$ captures individuals that were 27 or 28 when an off-reservation school first recruited from their reservation). In line with the discussion in the previous section, I set $j = 28$ as the reference group.

Since $age\ at\ exposure_{j(r,c)}$ varies at the reservation by cohort level, I am able to include

fixed effects for reservation (α_r) and cohort (α_c), with the latter grouped into 2-year bins. I also include tribe fixed effects (from the Indian schedules) in X_i. Tribe fixed effects capture time-invariant differences across distinct tribes occupying the same reservation (e.g., Northern Arapaho and Shoshone on the Wind River Reservation), as well as tribe-specific patterns across different reservations occupied by the same tribe.[24] I cluster standard errors at the reservation-level (i.e., the level of treatment).

The coefficients of interest are β_j, which represent the effect being exposed to the off-reservation school system at event time j. These coefficients provide intent-to-treat estimates; I am not estimating the effect of actually attending an off-reservation school (since I do not observe attendance for the majority of schools), but instead inferring exposure on the basis of an individual's reservation and cohort. An advantage of this approach is that intent-to-treat estimates capture both the effect of attendance, as well as community-level effects on non-attendees (i.e., spillovers). I show in section 5 that this measure of exposure predicts self-reported attendance at off-reservation schools (collected on the 1910 Indian schedules).

4.3 Outcomes and Estimation Sample

Estimation Sample: My main sample consists of male Native American household heads living in non-group quarters that were born between 1845 and 1890 (i.e., aged between 20 and 65 in 1910). I restrict attention to household heads in non-group quarters because some outcomes (e.g., related to marriage) are only relevant for households. I focus on males between the ages of 20 and 65 (i.e., roughly working age) because I am interested in measuring labour market outcomes. I exclude individuals that were matched to reservations under the jurisdiction of the Union Agency (i.e., the Cherokee, Chickasaw, Choctaw, Creek, and Seminole reservations), since Native Americans on these

[24] Around 80 per cent of reservations in my sample were occupied by multiple tribes. For single-tribe reservations, tribe fixed effects are absorbed by reservation fixed effects.

reservations were not targeted by the off-reservation school system, and were significantly more culturally and economically assimilated than the Native American population as a whole.[25] I also exclude individuals assigned to a non-reservation settlement in the vicinity of Pembroke, North Carolina, which corresponds to what is today the Lumbee Tribe of North Carolina. I do so because the Lumbee were not comparable—in terms of cultural and ethnic characteristics—to other Native American groups at the time (Pritzker, 1998, p. 552). Further information on excluded reservations an non-reservation settlements is available in Appendix C.

An underlying assumption of my empirical strategy is that an individual lived on the same reservation in 1910 as they did during the rollout of the off-reservation school system. For this reason, I identify individuals who—on the basis of their tribe—were not residing on their "home" reservation in 1910. I assume that individuals matched to a reservation were not on their home reservation if both (a) there were fewer than 100 individuals of the same tribe on the reservation, and (b) the tribe accounted for less than 5 per cent of the total reservation population.[26] All such individuals (roughly 4 per cent) are excluded from my main sample.

Given that reservations were treated from 1879 to 1910 and that my population of interest consists of individuals of working age in 1910, my estimation sample is necessarily unbalanced. For example, if a reservation was treated in 1880, I am able to observe outcomes for individuals born up to 35 years prior to the year of treatment (born in 1845, and aged 65 in 1910), and individuals born up to 10 years after the year of treatment (born in 1890, and aged 20 in 1910). On the other hand, if a reservation was treated in 1905, I can only observe outcomes for individuals born between 60 and 15 years prior to

[25] I also use person-level information on tribe to drop Cherokee, Chickasaw, Choctaw, Creek and Seminole individuals living in Oklahoma and adjacent states.

[26] I use both restrictions since—in a few cases—tribes on their own reservation accounted for less than 5 per cent of the population (e.g., Ute residing on the Navajo Reservation).

the year of treatment. Consistent with empirical approaches used in recent studies that face similar issues (e.g., Truffa and Wong, 2025), I restrict attention to reservations with "balanced" cohort representation around the time of treatment—relative to the "cutoff" age of 26—to ensure that my results are not driven by compositional changes. Specifically, I only keep "ever treated" reservations that have full cohort representation 4 periods (8 years) before treatment (i.e., aged 34 in the treatment year) and 6 periods (12 years) after treatment (i.e., aged 15 in the treatment year). For the purposes of calculating means for reservation-by-cohort bins, I further require that each reservation-by-cohort bin contains at least three observations. Finally, to reduce noise and ensure comparability with the treated sample, I restrict "never treated" reservations to those with at least 20 inhabitants in 1910. After applying these restrictions, I am left with a sample of around 15,000 individuals from 58 reservations (38 treated, 20 never treated).

Outcomes: I primarily measure outcomes in the 1910 census. Several outcomes (whether an individual appears in SAI membership lists, and whether they appear in an Indian census circa 1930) are obtained by matching records in the relevant datasets to the 1910 census. I examine four sets of outcomes: measures of educational attainment, measures of cultural assimilation, measures of economic integration, and measures of ethnic identity. Educational outcomes consist of indicators for: having graduated from an off-reservation school and being able to read and write (both self-reported). Cultural outcomes consist of indicators for: being able to speak English, reporting a "Western" first name, being coded as "mixed blood", and having a "mixed blood" spouse, conditional on marriage (all measured in 1910). Intermarriage and naming practices are standard measures of assimilation in the economic history literature (e.g., Fouka, 2019; Abramitzky, Boustan and Erikkson, 2020). On the other hand, perceived ancestry has not—to the best of my knowledge—been used as a measure of cultural assimilation in this literature. In my context, this outcome is arguably a

good proxy for unobservable dimensions of cultural assimilation, such as a person's accent and style of dress. Economic outcomes consist of indicators for: being in the labour force, being recorded as a "ration Indian" (i.e., without an occupation, and wholly dependent on the Government for support), having a white collar, skilled, or unskilled occupation (with the latter including farm labourers), or being a farmer or farm manager (excluding farm labourers).[27] Finally, I measure ethnic identity using (a) an indicator for being linked from the 1910 census to SAI membership lists, and (b) an indicator for being linked from the 1910 census to an Indian census between 1930 and 1934. More details on the outcomes used in my analysis are available in Appendix C. Table 1 shows descriptive statistics.

4.4 Threats to Identification

There are several threats to identification. First, my empirical strategy relies on the "parallel trends" assumption. In my context, this means that outcomes on treated reservations would have evolved in a similar way in the absence of treatment. Since recruitment decisions are unlikely to have been random, a possible threat is that off-reservation schools targeted reservations where educational, socioeconomic and assimilation outcomes were already on an "upwards" trajectory. I cannot categorically rule out this possibility, but I am able to provide evidence against it. In the results that follow, I document the absence of pre-trends for all outcomes.

Second, other policies that occurred at the same time as the rollout of the off-reservation school system may act as confounders. The only other major policy targeting Native Americans during this period was the "allotment" of reservations. This policy began in 1888, and sought to promote individual property rights by dividing and assigning reservation lands to individual Native Americans. It is unlikely that allotment had a meaningful effect on educational, socioeconomic, and assimilation outcomes during the

[27] Broad occupational categories are taken from Long and Ferrie (2013).

period of my study, since allotted lands were held in trust (i.e., could not be transferred) until the early-1900s (Dippel, Frye and Leonard, 2023). Furthermore, using person-level data on allotment status from the 1910 Indian schedules, I find that exposure to the off-reservation school system did not affect the probability of having an allotment.

Third, the recent econometrics literature has highlighted the potential for biases in staggered DiD estimates in the presence of heterogeneous treatment effects. In my setting, these biases could occur if the profile of effects on reservations treated earlier in the rollout (e.g., in Dakota Territory) differed from that on reservations treated towards the end of the rollout (e.g., in the Southwest). To address this concern, I also report estimates using a method proposed by Sun and Abraham (2021), which is robust to heterogeneous treatment effects in settings with staggered adoption.

5 Results

In this section, I set out my results. I first examine the effects of exposure to the off-reservation school system (based on an individual's reservation and cohort) on educational outcomes. I then document the effects of exposure on measures of cultural assimilation and economic integration. I conclude with results on ethnic identification and community attachment.

5.1 Educational Outcomes

I first examine the effects of exposure to the off-reservation school system (on the basis of reservation and cohort) on two educational outcomes, both measured in 1910: (a) the probability of graduating from an off-reservation school and (b) the probability of being able to read and write. Results are shown in panel A (probability of graduation from an off-reservation school) and panel B (probability of being able to read and write) of Figure 5. Reassuringly, there are no visible pre-trends in any of these outcomes. Panel A

serves as a validation of my empirical strategy. There is no effect on the probability of graduation for individuals that were aged over 30 when an off-reservation school first started recruiting from their reservation. However, the estimated effect begins to increase for cohorts aged in their late twenties (though statistically insignificant for older cohorts), and stabilises at around 1.5 percentage points, on average, for cohorts under the age of 20 (typical recruitment ages, based on Figure 4). The profile of estimated effects is similar with respect to the probability of being able to read and write (panel B).

Using an approach akin to Truffa and Wong (2025), I summarise magnitudes by presenting simple averages of (post-treatment) event time effects in Table 2. Motivated by Figure 4, I present average effects for three age groups based on the likelihood that they were recruited to an off-reservation school when their reservation was first treated: individuals aged between 34 and 29 (very unlikely—well past schooling age), individuals aged between 26 and 21 (unlikely—past typical non-Indian schooling age, but a non-negligible share of admitted students in my attendance data), and individuals aged between 20 and 15 (likely—typical ages of recruitment to off-reservation schools). While the effects on educational outcomes are small in absolute magnitude, they are both economically significant. For the youngest cohorts, the average effect on the probability of graduation (1.5 percentage points) represents a 50 per cent increase over the mean graduation rate in my sample, while the average effect on the probability of being able to read and write (7.5 percentage points) represents a 20 per cent increase over the corresponding sample mean.

5.2 Cultural Outcomes

Next, I examine the effects of exposure to the off-reservation school system on cultural outcomes. Results are shown in panels C to F of Figure 5. Panel C indicates that exposure was accompanied by a suggestive increase in the probability of speaking English, though the average effect size is small, and all coefficients are statistically indistinguishable from zero.

Panel D, on the other hand, shows that treated cohorts were substantially more likely to report a "Western" first name, with particularly strong effects among the youngest cohorts. Panel E presents results with respect to blood quantum reporting, showing an immediate and persistent increase in the likelihood of being recorded as "mixed blood" among treated cohorts. The abrupt increase among treated cohorts, along with a lack of visible pre-trends, makes demographic explanations (e.g., intermarriage) unlikely. Furthermore, my results with respect to identity and cultural attachment (discussed below) suggest that these effects reflected the perceptions of census enumerators, rather than deliberate changes in self-identification among Native Americans. Being perceived as mixed-race is likely to have reflected a variety of assimilation-relevant factors (e.g., accent and dress), and therefore provides strong evidence that exposure to the off-reservation school system promoted cultural assimilation. Finally, panel F indicates that more assimilated Native American household heads tended to have spouses who were themselves more assimilated, as measured by the probability of being recorded as "mixed blood".

As before, Table 2 presents simple averages of post-treatment event time effects for cultural outcomes. Average effects among the youngest cohorts on the probability of reporting a Western name, the probability of being recorded as "mixed blood", and the probability of having a "mixed blood" spouse are statistically significant at the 5 per cent level, and represent increases of between 15 and 35 per cent over the corresponding sample means.

5.3 Economic Outcomes

I now examine the effects of exposure to the off-reservation school system on economic (labour market) outcomes. Results are shown in panels A to F of Figure 6. Broadly, my results are suggestive of integration into the (Western) labour market, but only among younger treated cohorts. These cohorts were more likely to be in the labour force (panel A),

and less likely to be recorded as a "Ration Indian" (i.e., without an occupation, and wholly dependent on the Government for support). Interestingly, there appears to have been a divergence in labour market trajectories: based on broad occupational categorisations from Long and Ferrie (2013), treated cohorts moved into both "unskilled" occupations (e.g., farm labourers) and "white collar" ones (e.g., managers). Patterns with respect to other occupational categories ("skilled" occupations and farm owners or managers) are less pronounced. As above, Table 3 summarises the magnitudes of event time effects with simple averages. Even among the youngest cohorts, most effects are not statistically significant at conventional levels. The one exception is the probability of having an "unskilled" occupation (column 3).

5.4 Identity and Attachment

Finally, I examine the effects of exposure to the off-reservation school system on ethnic identification and cultural attachment. Did off-reservation schools succeed in displacing Indigenous identities? On the one hand, the effects I find with respect to cultural assimilation imply such an outcome. On the other hand, the historical literature—and accounts of students themselves—suggest an alternative possibility: that exposure to off-reservation schools provoked a resistance to cultural loss, and potentially strengthened the identities that the programme sought to displace.

To investigate the effects of exposure to off-reservation schools on identity and attachment, I re-estimate 1 on (a) an indicator for appearing in SAI membership lists between 1911 and 1915, and (b) an indicator for being linked to a local Indian census between 1930 and 1934. Table 4 presents results. Column 1 shows that treated cohorts were more likely to be members of the SAI (significant at the 5 per cent level). Columns 2 to 4 present results with respect to the probability of appearing in an Indian census in the early 1930s. Column 2 shows that—on average—treated cohorts were no more likely to be linked between the

1910 and 1930 population censuses using the Census Tree (Price et al., 2023). Column 3, however, reveals that the same cohorts were more likely to be linked from the 1910 census to a 1930s Indian census—after restricting attention to reservations where Indian censuses were taken in the early 1930s. The average effect—significant at the 5 per cent level—represents a roughly 30 per cent increase over the sample mean. As noted in section 3, Indian agents were specifically instructed to remove the names of individuals whose whereabouts were unknown from local census rolls in 1930. This means that an individual's appearance in an Indian census around this time is indicative of an ongoing connection to or presence at their home community. To account for mortality—and address the fact that the youngest cohorts were marginally more likely to be linked between the 1910 and 1930 censuses—column 4 presents results from the same specification when limiting attention to the subsample of individuals that could be linked from the 1910 census to the 1930 census using the Census Tree. The estimated effect with this additional restriction remains statistically significant and of a similar (relative) magnitude to that in the full sample. Taken together, these results suggest that exposure to the off-reservation school system—despite the policy's goals—did *not* lead to the displacement of Indigenous identities.

Finally, I examine whether—as posited in Native American sociological literature—off-reservation schools inadvertently strengthened ethnic identity among Native Americans. The first possibility—suggested by (Nagel, 1996, p. 116)—is that Indian boarding schools strengthened (pan-)Indian identity by acting "as ethnic melting pots, producing an alloy of tribal, regional, and supratribal identifications". The second possibility—noted by Frederick (1993) in interviews with mixed-race Native Americans—is that such schools "often served the ironic purpose of strengthening the Indian identity of bi-racial Indians". To investigate these possibilities, I split my main sample into (a) reservations with high (above median) or low (below median) ethnic diversity and (b) reservations with high (above median)

or low (below median) "full blood" population shares.[28,29] I then estimate equation 1 on each subsample, with an indicator for appearing in an Indian census as the outcome. Table 4 presents results. Column 1 shows results from the full sample (i.e., column 2 of Table 4). Columns 2 and 3 present estimates across reservations with high and low ethnic diversity, and columns 4 and 5 present equivalent estimates across reservations with high and low "full blood" population shares. The results—which demonstrate that effects were stronger on reservations that were (a) ethnically diverse or (b) had a relatively high mixed-race population—are consistent with the claim that off-reservation schools inadvertently *strengthened* ethnic identity among Native Americans.

5.5 Robustness

Robust TWFE Estimation: The applied econometrics literature has highlighted potential biases in two-way fixed effects models in settings with staggered treatment adoption. With respect to event study models such as equation 1, the main concern is that coefficient estimates are based on comparisons between not-yet-treated and already-treated units. In the presence of treatment effect heterogeneity, the estimated coefficients of a given lead or lag may be contaminated by the effects from other relative periods (Sun and Abraham, 2021). To address this concern, I assess the robustness of my results on educational, cultural, and economic outcomes to the estimation strategy proposed by Sun and Abraham (2021). Appendix Figure D.1 and Appendix Figure D.2 presents these results. In almost all cases, the pattern and magnitude of coefficient estimates are similar to those generated by the standard TWFE model.

[28] I measure ethnic (tribal) diversity at the reservation level using the ethnic fractionalisation index of Alesina et al. (2003). For a given reservation or settlement, this index measures the probability that two randomly selected individuals have different tribal affiliations.

[29] I calculate baseline "full blood" population shares using only cohorts born in 1860 or earlier—well before exposure to the off-reservation school system. This ensures that classification reflects pre-existing characteristics rather than treatment-induced changes in perceived or self-reported blood quantum.

Native Americans Counted as "White" in 1910: As described in section 3, the construction of my main sample relies on racial classification in the 1910 census. While it is reasonable to expect that few individuals switched from "Indian" to "White" between 1900 and 1910 (given the Indian schedules were used in both years), those that *did* switch would not be included in my main estimates. In practice, sample selection due to changes in racial classification prior to 1910 is unlikely to affect my main conclusions; using the Census Tree to link Native American males from 1900 to 1910, I find that a negligible share (around 1 per cent) switched from "Indian" in 1900 to "White" in 1910.

6 Conclusion

Throughout history, governments have sought to promote cultural assimilation and economic integration with varying levels of coercion. While these efforts have typically targeted vulnerable, disenfranchised, or otherwise marginalised groups, among the most coercive policies have been reserved for Indigenous populations.

While prior work in economics has studied the long-term effects of Indigenous assimilation policies, this paper provides the first comprehensive, quantitative analysis of the off-reservation school system—arguably the most coercive assimilation effort in U.S. history—in its historical context. As a first step to doing so, I digitise new data on Native Americans (covering over 90 per cent of the 1910 Native population), match Native Americans in the 1910 census to reservations and settlements, and produce the first reservation-level dataset on off-reservation school recruitment patterns during their rollout. It is hoped that these datasets facilitate previously unfeasible research on Native Americans in economic history.

Using these data, I show that exposure to the off-reservation school system led to cultural assimilation and—to a lesser extent—labour market integration. However—as emphasised in the broader historical literature—my findings indicate that Native Americans were not

passive actors in this process, and cultural assimilation did not completely displace ethnic identity and cultural connection. Indeed, my results with respect to appearance in Indian censuses some 20 years later suggest that exposure to the off-reservation school system may have inadvertently strengthened ethnic identity and community attachment—with implications for tribal membership in the past and today. Taken together, my results highlight the agency and resilience of Native American individuals and communities in response to the off-reservation school system, contributing to a growing literature in economics that studies the historical experiences of Indigenous populations (e.g., Feir, 2025; Carlos, Feir and Redish, 2022).

References

ABC News. 2008. "Thousands Greet Stolen Generations Apology." *https://www.abc.net.au/news/2008-02-13/thousands-greet-stolen-generations-apology/1041322*, Accessed 1 December 2022.

Abdelgadir, Aala, and Vasiliki Fouka. 2020. "Political Secularism and Muslim Integration in the West: Assessing the Effects of the French Headscarf Ban." *American Political Science Review*, 114(3): 707–723.

Abramitzky, Ran, and Leah Boustan. 2024. "Immigrants and Their Children Assimilate Into US Society and the US Economy, Both in the Past and Today." *PNAS Nexus*, 3(10): pgae344.

Abramitzky, Ran, Jacob Conway, Roy Mill, and Luke Stein. 2023. "The Gendered Impacts of Perceived Skin Tone: Evidence from African-American Siblings in 1870–1940." National Bureau of Economic Research Working Paper 31016.

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. 2019. "To the New World and Back Again: Return Migrants in the Age of Mass Migration." *ILR Review*, 72(2): 300–322.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, Santiago Pérez, and Myera Rashid. 2025. "Census Linking Project: Version 3.0." *https://censuslinkingproject.org*, [dataset].

Abramitzky, Ran, Leah Platt Boustan, and Katherine Erikkson. 2020. "Do Immigrants Assimilate More Slowly Today Than in the Past?" *American Economic Review: Insights*.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy*, 122(3): 467–506.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2016. "Cultural Assimilation during the Age of Mass Migration." National Bureau of Economic Research Working Paper 22381.

Adams, David Wallace. 2020. *Education for Extinction: American Indians and the Boarding School Experience, 1875-1928.* ERIC.

Adukia, Anjali, Richard Hornbeck, Daniel Keniston, and Benjamin Lualdi. 2025. "The Social Construction of Race During Reconstruction." University of Chicago Working Paper.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *The Quarterly Journal of Economics*, 115(3): 715–753.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth*, 8(2): 155–194.

Australian Human Rights Commission. 1997. "Bringing Them Home Report."

BBC News. 2019. "China Muslims: Xinjiang Schools Used to Separate Children from Families." https://www.bbc.com/news/world-asia-china-48825090, Accessed 2 October 2025.

BBC News. 2025. "Parades, Flags and Songs: The Campaign to Force Ukrainian Children to Love Russia." https://www.bbc.com/news/articles/cz6g20dwy0wo, Accessed 2 October 2025.

Berkes, Enrico, Ezra Karger, and Peter Nencka. 2023. "The Census Place Project: A Method for Geolocating Unstructured Place Names." *Explorations in Economic History*, 87: 101477. Methodological Advances in the Extraction and Analysis of Historical Data.

Board of Indian Commissioners. 1869. "Report of the Board of Indian Commissioners, 1869."

Brudvig, Jon. 1994. "Hampton Normal and Agricultural Institute: American Indian Students: 1878-1923."

Bureau of Land Management. 2024. "PLSS Townships." Accessed 2024.

Cantoni, Davide, Yuyu Chen, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang. 2017. "Curriculum and Ideology." *Journal of Political Economy*, 125(2): 338–392.

Carlos, A. M., D. L. Feir, and A. Redish. 2022. "Indigenous Nations and the Development of the U.S. Economy: Land, Resources, and Dispossession." *The Journal of Economic History*, 82(2): 516–555.

CBC News. 2008. "PM Cites 'Sad Chapter' in Apology for Residential Schools." https://www.cbc.ca/news/canada/pm-cites-sad-chapter-in-apology-for-residential-schools-1.699389, Accessed 1 December 2022.

Child, Brenda Joyce. 1993. "A Bitter Lesson: Native Americans and the Government Boarding School Experience, 1890-1940." PhD diss. The University of Iowa.

Clark, David Anthony Tyeeme. 2004. "Representing Indians: Indigenous Fugitives and the Society of American Indians in the Making of Common Culture." PhD diss. The University of Kansas.

Clots-Figueras, Irma, and Paolo Masella. 2013. "Education, Language and Identity." *The Economic Journal*, 123(570): F332–F357.

Dahis, Ricardo, Emily Nix, and Nancy Qian. 2020. "Choosing Racial Identity in the United States, 1880-1940." National Bureau of Economic Research Working Paper 26465.

Dippel, Christian. 2014. "Forced Coexistence and Economic Development: Evidence From Native American Reservations." *Econometrica*, 82(6): 2131–2165.

Dippel, Christian, and Dustin Frye. 2020. "Native Americans in the Historical Census: New Data and Applications." Working paper.

Dippel, Christian, Dustin Frye, and Bryan Leonard. 2023. "Property Rights Without Transfer Rights: A Study of Indian Land Allotment." *Working paper*, 2131–2165.

Feir, Donn L. 2016. "The Long-term Effects of Forcible Assimilation Policy: The Case of Indian Boarding Schools." *Canadian Journal of Economics/Revue canadienne d'économique*, 49(2): 433–480.

Feir, Donn L. 2025. "Economic History and Indigenous Peoples in North America." *Annual Review of Economics*, 17(1): 479–509.

Feir, Donn L, Rob Gillezeau, and Maggie E C Jones. 2023. "The Slaughter of the Bison and Reversal of Fortunes on the Great Plains." *The Review of Economic Studies*, rdad060.

Fouka, Vasiliki. 2019. "Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I." *The Review of Economic Studies*, 87(1): 204–239.

Fouka, Vasiliki. 2024. "State Policy and Immigrant Integration." *Annual Review of Political Science*, 27: 25–46.

Frederick, Paula A. 1993. "(Bi)racial Constraints or Ethnic Options? Identity Among a Biracial Population of American Indians." Miami, FL.

Frye, Dustin, and Dominic P. Parker. 2021. "Indigenous Self-Governance and Development on American Indian Reservations." *AEA Papers and Proceedings*, 111: 233–37.

Gregg, Matthew T. 2018. "The Long-term Effects of American Indian Boarding Schools." *Journal of Development Economics*, 130: 17–32.

Hertzberg, Hazel. 1971. *The Search for an American Indian Identity: Modern Pan-Indian Movements.* Syracuse University Press.

Institute for Government Research. 1928. *The Problem of Indian Administration.* Baltimore: Johns Hopkins University Press.

Jones, Maggie E. C. 2024. "The Intergenerational Legacy of Indian Residential Schools." *Demography*, 61(6): 1871–1895.

Lagakos, David, Stelios Michalopoulos, and Hans-Joachim Voth. 2025. "American Life Histories." National Bureau of Economic Research NBER Working Paper 33373, Cambridge, MA.

Lomawaima, K. Tsianina. 1994. *They Called it Prairie Light: The Story of Chilocco Indian School.* University of Nebraska Press.

Long, Jason, and Joseph Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States since 1850." *American Economic Review*, 103(4): 1109–1137.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2023. "IPUMS National Historical Geographic Information System: Version 18.0 [dataset]."

Miller, M. C. 2024. "Assimilation and Economic Development: The Case of Federal Indian Policy." *Public Choice*, 199: 387–404.

Nagel, Joane. 1996. *American Indian Ethnic Renewal: Red Power and the Resurgence of Identity and Culture.* Oxford University Press.

National Archives and Records Administration. 2014. "Indian Census Rolls, 1885-1940." https://www.archives.gov/research/census/native-americans/1885-1940.html, Accessed October 23, 2025.

Office of Indian Affairs. 1878. "Annual Report of the Commissioner of Indian Affairs, For the Year 1878." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1879. "Annual Report of the Commissioner of Indian Affairs, For the Year 1879." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1887. "Annual Report of the Commissioner of Indian Affairs, For the Year 1887." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1890. "Annual Report of the Commissioner of Indian Affairs, For the Year 1890." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1891. "Annual Report of the Commissioner of Indian Affairs, For the Year 1891." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1896. "Annual Report of the Commissioner of Indian Affairs, For the Year 1896." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1897. "Annual Report of the Commissioner of Indian Affairs, For the Year 1896." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1898. "Annual Report of the Commissioner of Indian Affairs, For the Year 1898." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1900. "Annual Report of the Commissioner of Indian Affairs, For the Year 1900." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1902. "Education Circular No. 85."

Office of Indian Affairs. 1909. "Annual Report of the Commissioner of Indian Affairs, For the Year 1909." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1910. "Annual Report of the Commissioner of Indian Affairs, For the Year 1910." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1926. "Annual Report of the Commissioner of Indian Affairs, For the Year 1926." Washington, D.C.: G.P.O.

Office of Indian Affairs. 1930*a*. "Circular 2653." U.S. Department of the Interior Circular.

Office of Indian Affairs. 1930*b*. "Circular 2676." U.S. Department of the Interior Circular.

OpenAI. 2025. "Prompt Engineering Guide." Accessed: 2025-10-19.

Powell, J. W. 1899. "Eighteenth Annual Report of the Bureau of American Ethnology to the Secretary of the Smithsonian Institution 1896-97."

Pratt, Richard Henry. 1912. "One of General Sheridan's Ways With Indians and What Came of It."

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. 2023. "The Census Tree, 1910-1930."

Pritzker, Barry M. 1998. *Native Americans: An Encyclopedia of History, Culture, and Peoples.* ABC-CLIO, Inc.

Prucha, Francis Paul. 1984. *The Great Father: The United States Government and the American Indians.* University of Nebraska Press.

Rogers, John. 1974. *Red World and White: Memories of a Chippewa Boyhood.* Norman: University of Oklahoma Press.

Ruggles, Steven, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. 2021. "IPUMS Ancestry Full Count Data: Version 3.0 [dataset]."

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek. 2022. "IPUMS USA: Version 12.0 [dataset]."

Smith, Andrea. 2009. "Indigenous Peoples and Boarding Schools: A Comparative Study." United Nations, Permanent Forum on Indigenous Issues Technical Report.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225(2): 175–199.

Truffa, Francesca, and Ashley Wong. 2025. "Undergraduate Gender Diversity and the Direction of Scientific Research." *American Economic Review*, 115(7): 2414–2448.

United States Geological Survey. 2021. "Geographic Names Information System." Updated August 27.

U.S. Census Bureau. 2002. "Measuring America: The Decennial Censuses from 1790 to 2000." U.S. Census Bureau POL/02-MA(RV).

U.S. Department of the Interior. 2024. "Federal Indian Boarding School Initiative." *https://www.doi.gov/priorities/strengthening-indian-country/federal-indian-boarding-school-initiative*, Accessed 12 April 2024.

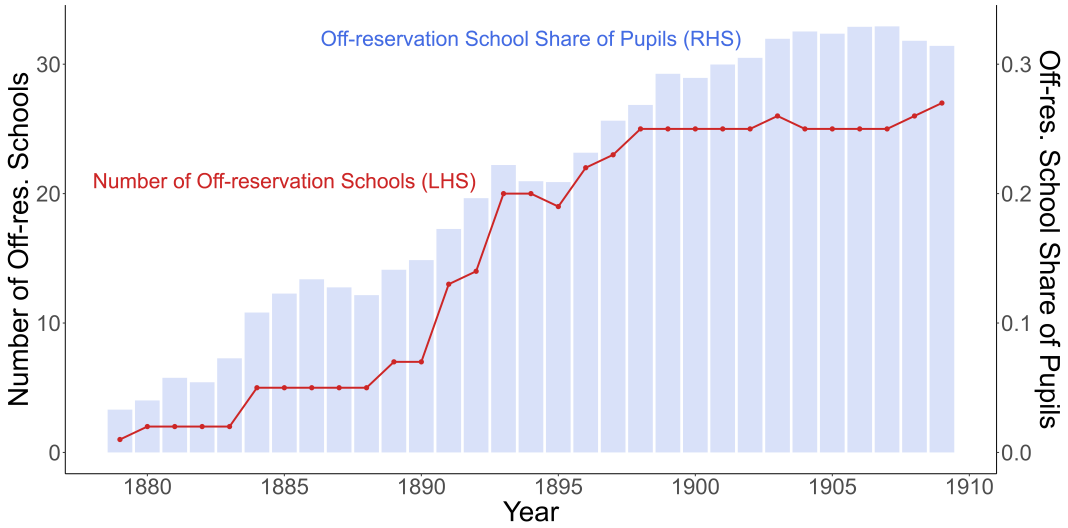
U.S. Geological Survey. 2023. "National Hydrography Dataset."

Vuckovic, Myriam. 2008. *Voices from Haskell.* University Press of Kansas.

Weinberg, Marjorie. 2004. *The Real Rosebud: The Triumph of a Lakota Woman. Foreword by Luke Yellow Robe.* Lincoln: University of Nebraska Press.

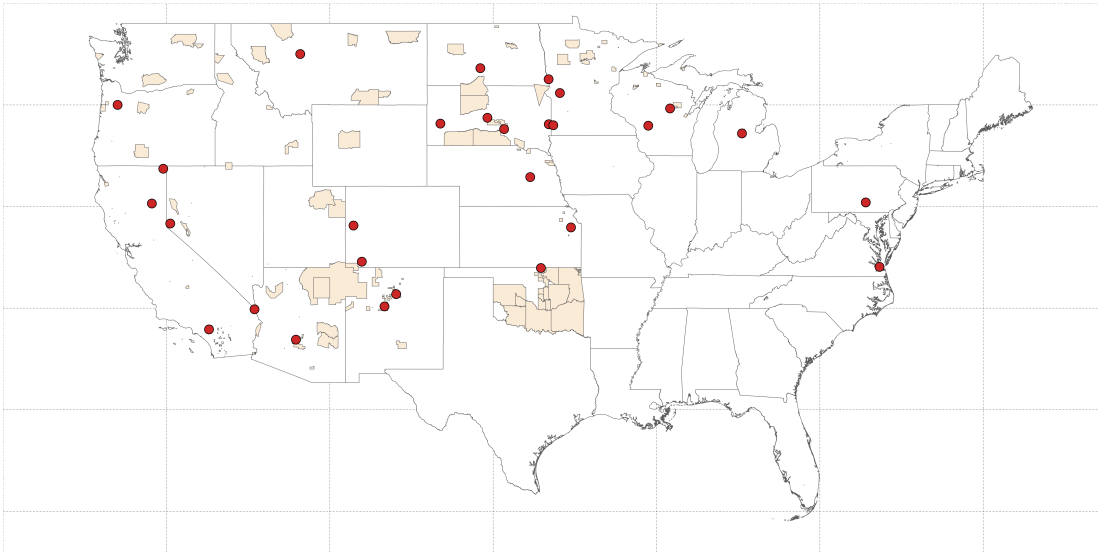
Figures and Tables

Figure 1. The Off-reservation School System, 1879 - 1909



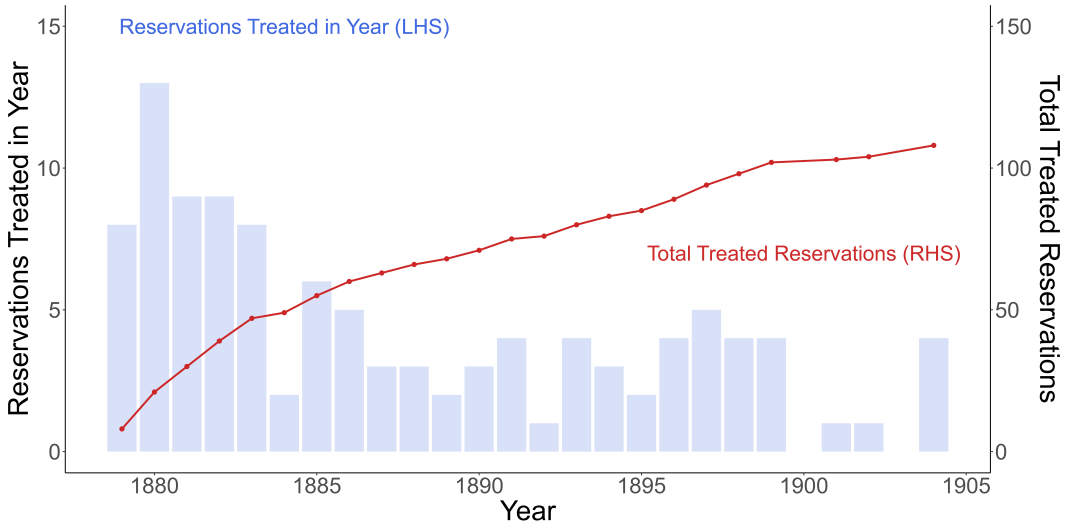
Notes: This figure shows the cumulative number of off-reservation schools operating between 1879 and 1909 (left axis) and the share of pupils attending off-reservation schools (right axis). Share of pupils attending off-reservation schools is calculated as the average attendance at off-reservation schools over the average attendance at all schools (federal government boarding and day, non-government boarding and day, and state public schools) in a given year.
Source: Own calculations using data from Annual Reports of the Commissioner of Indian Affairs, 1880 - 1909.

Figure 2. Distribution of Off-reservation Schools, 1910



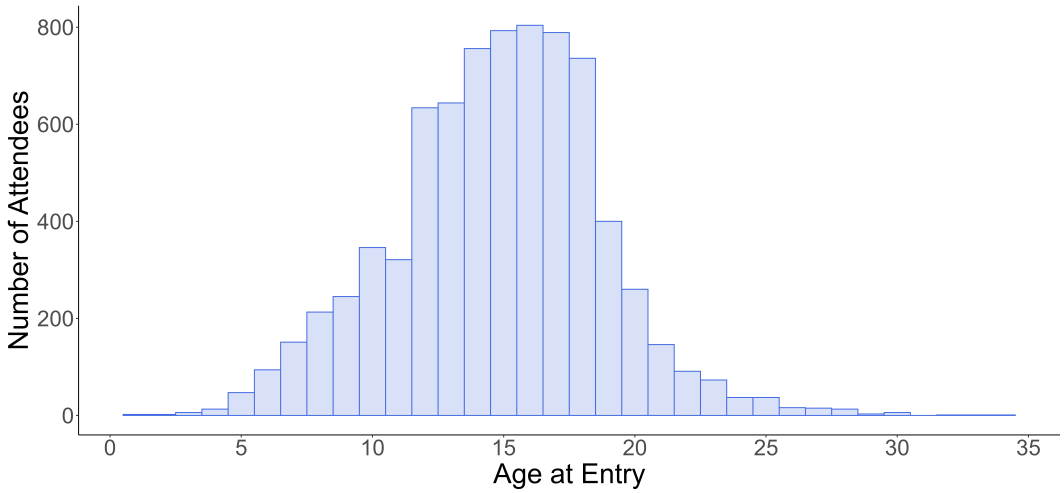
Notes: This figure shows the locations of off-reservation schools in operation in 1910. Red points represent off-reservation schools, and polygons represent reservation boundaries as at 1910. Source: Own work using school locations from Adams (2020).

Figure 3. Off-reservation School Treatment Years, 1879-1905



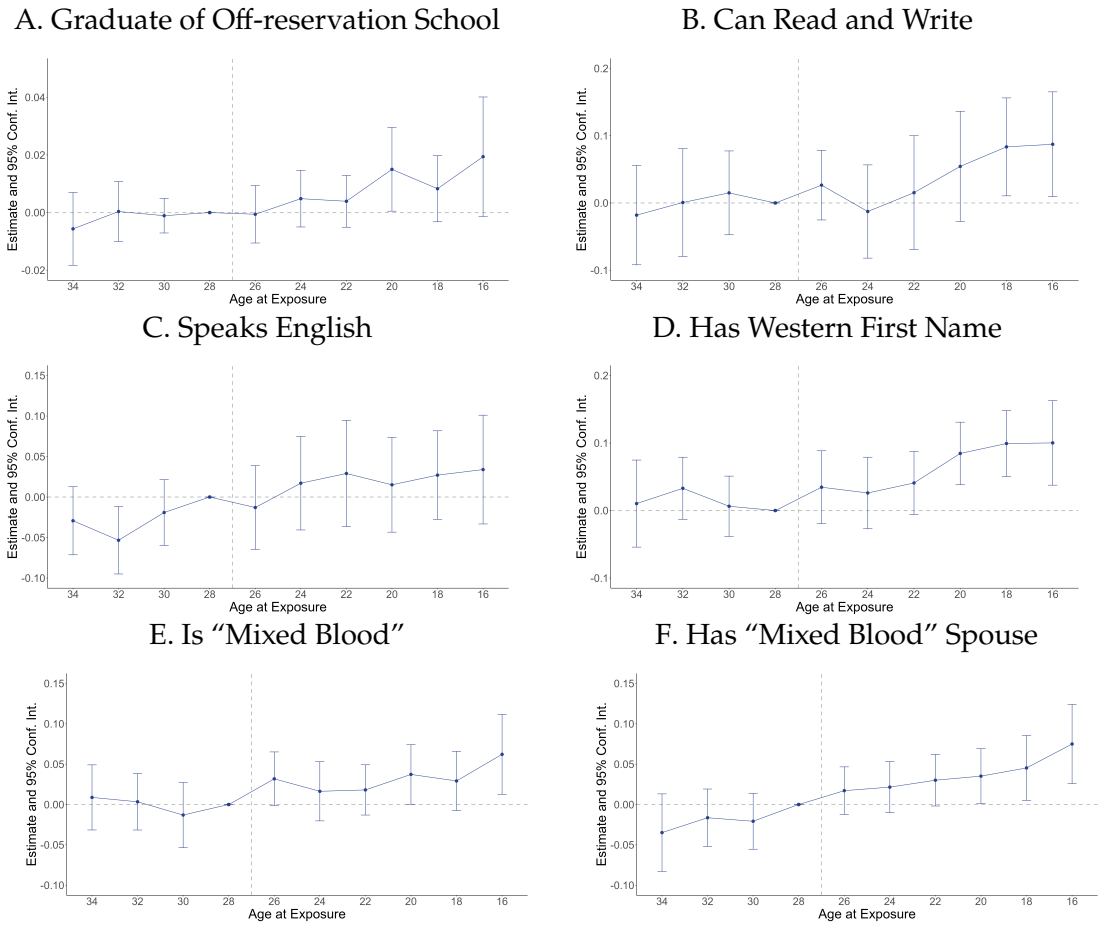
Notes: This figure shows the number of reservations that were treated by an off-reservation school for the first time in a given year (left axis) and the total number of reservations treated by an off-reservation school up to and including that year (right axis).
Source: Own calculations using data from Annual Reports of the Commissioner of Indian Affairs, 1880-1906, and attendance records for Carlisle, Chemawa, Hampton, and Haskell.

Figure 4. Ages of Entry to Off-reservation Schools, 1879 - 1900



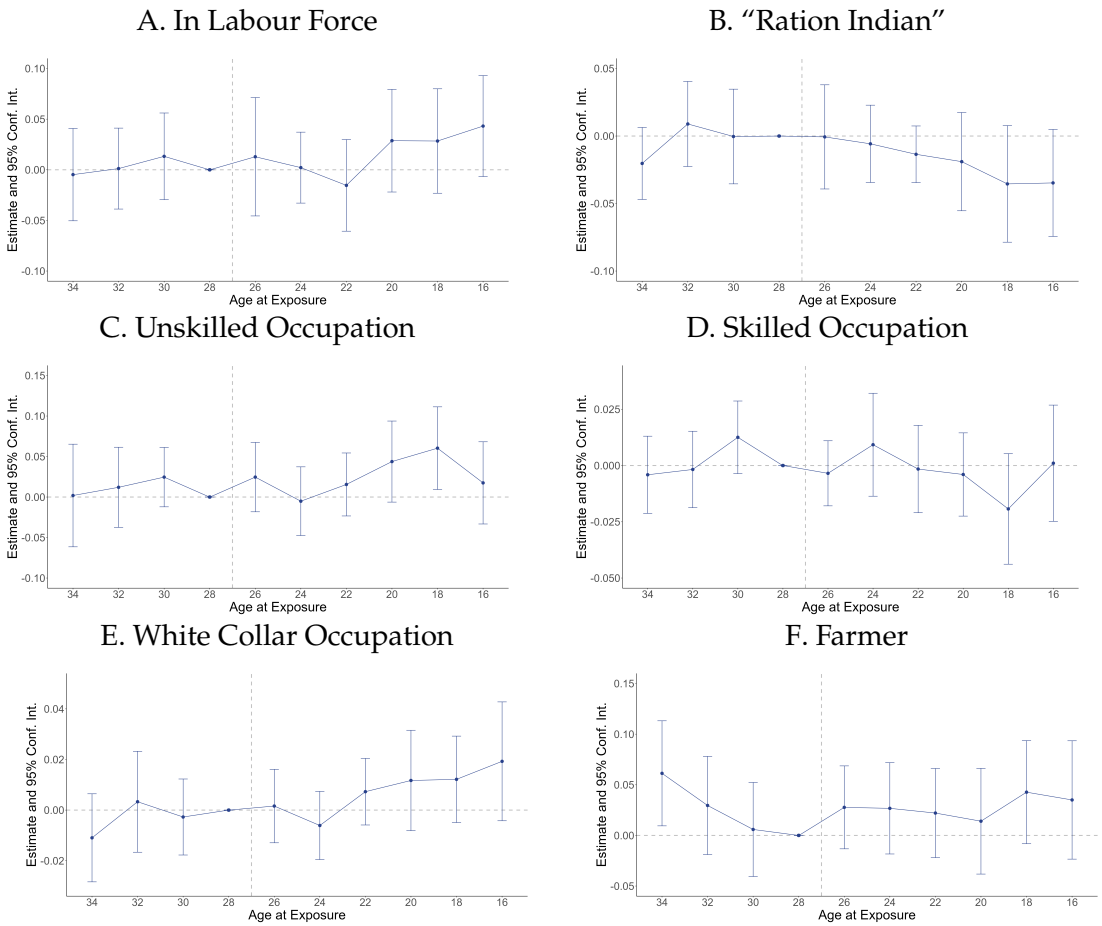
Notes: This figure shows the ages of students when admitted into one of Carlisle, Chemawa, or Haskell.
Source: Own calculations using attendance records for Carlisle, Chemawa, and Haskell.

Figure 5. Educational and Cultural Outcomes



Notes: This figure shows estimates from equation 1 for the listed outcomes. The sample consists of male, Native American household heads aged between 20 and 65 in 1910. The sample excludes all individuals matched to reservations under the Union Agency, all Cherokee, Chickasaw, Choctaw, Creek and Seminole in Oklahoma and adjacent states, and all Native Americans at the non-reservation settlement in the vicinity of Pembroke, North Carolina. All outcomes are measured in 1910. All regressions include reservation fixed effects and cohort fixed effects (2-year bins). Standard errors are clustered at the reservation-level.

Figure 6. Economic Outcomes



Notes: This figure shows estimates from equation 1 for the listed outcomes. The sample consists of male, Native American household heads aged between 20 and 65 in 1910. The sample excludes all individuals matched to reservations under the Union Agency, all Cherokee, Chickasaw, Choctaw, Creek and Seminole in Oklahoma and adjacent states, and all Native Americans at the non-reservation settlement in the vicinity of Pembroke, North Carolina. All outcomes are measured in 1910. "Ration Indian" is an indicator equal to 1 if an individual's occupation string was "ration" or similar. "Unskilled Occupation", "Skilled Occupation", "White Collar Occupation" and "Farmer" are based on the categorisation from Long and Ferrie (2013) and the IPUMS variable OCC1950. All regressions include reservation fixed effects and cohort fixed effects (2-year bins). Standard errors are clustered at the reservation-level.

Table 1. Descriptive Statistics

	Year	Mean	SD	N
Demographic				
Age	1910	41.096	11.829	16,302
Is "Mixed Blood"	1910	0.162	0.369	15,981
Educational				
Graduate of Off-reservation School	1910	0.022	0.148	16,302
Can Read and Write	1910	0.345	0.475	16,302
Speaks English	1910	0.602	0.489	16,302
Cultural				
Has Western First Name	1910	0.652	0.476	16,183
Has "Mixed Blood" Spouse	1910	0.140	0.347	13,733
Economic				
In Labour Force	1910	0.829	0.376	16,302
"Ration Indian"	1910	0.059	0.236	14,816
Unskilled Occupation	1910	0.347	0.476	14,816
Skilled Occupation	1910	0.036	0.186	14,816
White Collar Occupation	1910	0.023	0.150	14,816
Farmer	1910	0.397	0.489	14,816
Identity and Attachment				
SAI Member	1911-1914	0.002	0.044	16,302
Linked to 1930s Indian Census	1930-1934	0.184	0.388	16,302

Notes: This table shows descriptive statistics. The sample consists of male, Native American household heads and aged between 20 and 65 in 1910. The sample excludes all individuals matched to reservations under the Union Agency, all Cherokee, Chickasaw, Choctaw, Creek and Seminole in Oklahoma and adjacent states, and all Native Americans at the non-reservation settlement in the vicinity of Pembroke, North Carolina. "Ration Indian" is an indicator equal to 1 if an individual's occupation string was "ration" or similar. "Unskilled Occupation", "Skilled Occupation", "White Collar Occupation" and "Farmer" are based on the categorisation from citelong2013intergenerational and the IPUMS variable OCC1950.

Table 2. Educational and Cultural Outcomes

	Graduate of Off-reservation School	Can Read and Write	Speaks English	Has Western First Name	Is "Mixed Blood"	Has "Mixed Blood" Spouse
	(1)	(2)	(3)	(4)	(5)	(6)
Age 34 to 29	-0.002	-0.001	-0.034	0.017	0.000	-0.024
	(0.004)	(0.032)	(0.016)	(0.022)	(0.017)	(0.018)
Age 26 to 21	0.003	0.010	0.011	0.034	0.022	0.023
	(0.004)	(0.027)	(0.025)	(0.021)	(0.013)	(0.014)
Age 20 to 15	0.014	0.075	0.025	0.095	0.043	0.052
	(0.007)	(0.035)	(0.027)	(0.024)	(0.018)	(0.018)
Year	1910	1910	1910	1910	1910	1910
Mean dep. var	0.022	0.345	0.602	0.652	0.162	0.140
R2	0.076	0.298	0.355	0.369	0.431	0.436
No. reservations	58	58	58	58	58	58
No. cohorts	24	24	24	24	24	24
Obs.	16,302	16,302	16,302	16,183	15,981	13,733

Notes: This table shows results from estimating equation 1. The sample consists of male, Native American household heads aged between 20 and 65 in 1910. The sample excludes all individuals matched to reservations under the Union Agency, all Cherokee, Chickasaw, Choctaw, Creek and Seminole in Oklahoma and adjacent states, and all Native Americans at the non-reservation settlement in the vicinity of Pembroke, North Carolina.

Table 3. Economic Outcomes

	In Labour Force	"Ration Indian"	Unskilled Occupation	Skilled Occupation	White Collar Occupation	Farmer
	(1)	(2)	(3)	(4)	(5)	(6)
Age 34 to 29	0.003	-0.004	0.013	0.002	-0.003	0.032
	(0.017)	(0.013)	(0.020)	(0.006)	(0.007)	(0.020)
Age 26 to 21	0.000	-0.007	0.012	0.001	0.001	0.026
	(0.020)	(0.013)	(0.017)	(0.008)	(0.005)	(0.019)
Age 20 to 15	0.033	-0.030	0.041	-0.007	0.014	0.031
	(0.022)	(0.019)	(0.023)	(0.010)	(0.009)	(0.024)
Year	1910	1910	1910	1910	1910	1910
Mean dep. var	0.829	0.059	0.347	0.036	0.023	0.397
R2	0.269	0.262	0.252	0.110	0.030	0.277
No. reservations	58	58	58	58	58	58
No. cohorts	24	24	24	24	24	24
Obs.	16,302	14,816	14,816	14,816	14,816	14,816

Notes: This table shows results from estimating equation 1. The sample consists of male, Native American household heads aged between 20 and 65 in 1910. The sample excludes all individuals matched to reservations under the Union Agency, all Cherokee, Chickasaw, Choctaw, Creek and Seminole in Oklahoma and adjacent states, and all Native Americans at the non-reservation settlement in the vicinity of Pembroke, North Carolina. All outcomes are measured in 1910. "Ration Indian" is an indicator equal to 1 if an individual's occupation string was "ration" or similar. "Unskilled Occupation", "Skilled Occupation", "White Collar Occupation" and "Farmer" are based on the categorisation from citelong2013intergenerational and the IPUMS variable OCC1950. All regressions include reservation fixed effects and cohort fixed effects (2-year bins). Standard errors are clustered at the reservation-level.

Table 4. Identity and Community Attachment

	SAI Member	Linked to 1930 Census	Linked to 1930s Indian Census	
	(1)	(2)	(3)	(4)
Age 34 to 29	0.004	0.011	0.003	0.027
	(0.002)	(0.013)	(0.015)	(0.058)
Age 26 to 21	0.003	0.008	0.045	0.127
	(0.001)	(0.019)	(0.017)	(0.053)
Age 20 to 15	0.005	0.047	0.066	0.161
	(0.002)	(0.025)	(0.021)	(0.048)
Year	1910	1930	1930	1930
Mean dep. var	0.002	0.233	0.195	0.509
R2	0.016	0.175	0.143	0.153
No. reservations	58	45	45	45
No. cohorts	24	24	24	24
Obs.	16,302	15,418	15,418	3,592

Notes: This table shows results from estimating equation 1. The sample consists of male, Native American household heads aged between 20 and 65 in 1910. The sample excludes all individuals matched to reservations under the Union Agency, all Cherokee, Chickasaw, Choctaw, Creek and Seminole in Oklahoma and adjacent states, and all Native Americans at the non-reservation settlement in the vicinity of Pembroke, North Carolina. "SAI Member" is an indicator equal to 1 if the individual was successfully linked from SAI membership lists between 1911 and 1915 to the 1910 census. "Linked to 1930 Census" is an indicator equal to 1 if the individual could be linked to the 1930 Census using the Census Tree. "Linked to 1930s Indian Census" is an indicator equal to 1 if the individual was linked from the 1910 census to any Indian census between 1930 and 1934 using the ABE-JW algorithm. The sample in columns 2 and 3 is restricted to reservations where at least one individual was linked to a 1930s Indian census. The sample in column 4 is further restricted to individuals linked from the 1910 census to the 1930 census using the Census Tree. All regressions include reservation fixed effects and cohort fixed effects (2-year bins). Standard errors are clustered at the reservation-level.

Table 5. Identity and Community Attachment, Heterogeneity

	Linked to 1930s Indian Census				
	Full Sample	Low Diversity	High Diversity	Low "Mixed Blood"	High "Mixed Blood"
	(1)	(2)	(3)	(4)	(5)
Age 34 to 29	0.003	0.002	0.011	0.023	-0.003
	(0.015)	(0.021)	(0.027)	(0.022)	(0.026)
Age 26 to 21	0.045	0.015	0.073	0.025	0.077
	(0.017)	(0.029)	(0.024)	(0.025)	(0.031)
Age 20 to 15	0.066	0.044	0.090	0.004	0.157
	(0.021)	(0.026)	(0.035)	(0.030)	(0.030)
Year	1930	1930	1930	1930	1930
Mean dep. var	0.195	0.146	0.235	0.169	0.239
R2	0.143	0.180	0.113	0.156	0.130
No. reservations	45	15	30	24	21
No. cohorts	24	24	24	24	24
Obs.	15,418	6,903	8,515	9,693	5,725

Notes: This table shows results from estimating equation 1. The sample consists of male, Native American household heads aged between 20 and 65 in 1910. The sample excludes all individuals matched to reservations under the Union Agency, all Cherokee, Chickasaw, Choctaw, Creek and Seminole in Oklahoma and adjacent states, and all Native Americans at the non-reservation settlement in the vicinity of Pembroke, North Carolina. "Linked to 1930s Indian Census" is an indicator equal to 1 if the individual was linked from the 1910 census to any Indian census between 1930 and 1934 using the ABE-JW algorithm. The sample is restricted to reservations where at least one individual was linked to a 1930s Indian census. All regressions include reservation fixed effects and cohort fixed effects (2-year bins). Standard errors are clustered at the reservation-level.

Supplementary Appendix

A Context

A.1 Additional Information on Off-reservation Schools

Rollout of Off-reservation Schools: The first off-reservation boarding schools, the Carlisle Indian School, was opened in 1879. Carlisle represented the culmination of efforts by an Army Officer, Richard Henry Pratt, to develop a new model of education for Native Americans. Pratt's interest in Native American education policy stemmed from his experience supervising prisoners of war at Fort Marion, Florida. He subsequently arranged for the transfer of some of these prisoners of war to be educated at the Hampton Institute, Virginia. Based on the success of this "experiment", the decision was made of open Carlisle the following year (Office of Indian Affairs, 1879, p. VIII). This model gained popularity with policymakers, and five off-reservation schools were opened over the next five years. These schools, in Kansas, Nebraska, New Mexico, Oklahoma and Oregon, were located closer to the communities from which they recruited, but were nonetheless located outside reservations. Schools continued to be opened over the next 20 years. Apart from Carlisle, Hampton and Haskell, all of these schools tended to recruit locally (within state or from adjacent states). Table A.6 shows the the names, locations, and opening years of off-reservation schools.

Curriculum at Off-reservation Schools: The 1890 Annual Report of the Commissioner of Indian Affairs includes an Appendix titled "Rules for Indian Schools" (Office of Indian Affairs, 1890). This document sets out the curriculum of Native American students, known as the "course of study". This curriculum was designed for reservation boarding schools, but was 'to be followed as far as practicable in day schools' (Office of Indian Affairs, 1890, p. CLVI).

The course of study consisted of two four-year grades: primary grade and advanced

grade. The first year of primary grade mainly consisted of English language instruction, basic reading and writing, and numbers from 1 to 10 (p. CLVI). The second year of primary grade built on the previous year's study of English, and added orthography, form and colour, penmanship and drawing, and geography (e.g., of the reservation or county). The third and fourth years of primary grade continued with the study of these topics, with arithmetic added in the fourth year (p. CLVIII).

The advanced grade covered similar content at a more advanced level. In terms of reading, each year was assigned a "Reader" with increasing complexity. The list of Readers, at least in principle, common across schools.[30] Content on US history, physiology and hygiene, and civil government were added in the final years of advanced grade (p. CLIX). In addition to academic work, students were provided with "industrial training". This covered topics such as farming, instruction in trades (for boys), housekeeping (for girls). At least half of the school day was to be devoted to industrial work (p. CLII).

The level of education at off-reservation schools, at least until the 1920s, was generally at the primary level. While the largest schools, such as Carlisle and Haskell did offer commercial and normal (teacher training) courses beyond the 8-year programme, they never aspired to provide a high-school education (Pratt, 1912, p. 13). In fact, even these post-primary courses were scaled back when, in the early-1900s, school adminstrators were explicitly instructed *not* to provide instruction above the level of the eighth grade (Vuckovic, 2008, p. 94). This reflected a change in Indian education policy in the early-1900s, which led to a shift from academic training to vocational work (Adams, 2020, p. 172). It was not until the-late 1920s that the first off-reservation schools were authorised to teach senior high-school grades (i.e., grade 10 to grade 12) (Office of Indian Affairs, 1926, p. 7).

[30]The course of study was accompanied with a book list, set out on pp. CLXI-CLXII.

A.2 Native Americans in Historical Censuses

Population Censuses: The availability of information on Native Americans in historical decennial censuses (i.e., 1790 to 1940) varies from census to census. No attempt was made to count Native Americans until the 1860 census, and this enumeration only included individuals that had left their tribes and lived in White communities. Native Americans, both on reservations and in White communities, were in principle fully counted from the 1890 census onwards.[31]

The information collected on Native Americans also varied from census to census. In 1900 and 1910, special schedules was used to enumerate Native Americans. These schedules included information on an individual's tribe, degree of "Indian blood" (e.g., "full", "half"), whether they were living in a polygamous relationship, and their dwelling type (e.g., "aboriginal" or "civilized"). These "Indian schedules" were not used in later censuses. However, census enumerators were given special instructions to collect information on tribal affiliation and "Indian blood" in the normal schedules for the 1930 census.

Indian Censuses: The Indian Appropriations Act of 1884 stated that "each Indian agent be required, in his annual report, submit a census of the Indians at his agency or upon the reservation under his charge." (23 Stat. 76, 98). The following year, the Commissioner of Indian Affairs sent further instructions to Indian agents, including a sample form (National Archives and Records Administration, 2014). In principle, these censuses—henceforth "Indian censuses"—were to be conducted annually. However, in practice, Indian censuses were not conducted in some years for various reasons (e.g., lack of agency staff)—especially during the nineteenth and early twentieth centuries. The same information—name, sex, age, and household relationship—was generally collected until the late 1920s, when the

[31] Individual records from the 1890 census are no longer available today. For more information on the fate of these records, see: https://www.archives.gov/publications/prologue/1996/spring/1890-census.

forms were expanded to include allotment numbers, tribe, blood quantum, marital status, and whether the individual was living at the jurisdiction where they were enrolled (i.e., their "home" agency).

In the early 1930s, the Indian Office took steps to improve the accuracy of Indian censuses. Agents were instructed to "show only Indians at your jurisdiction... Names of Indians removed from the rolls since the last census, because of death or otherwise, must be entirely omitted" (Office of Indian Affairs, 1930*b*, cited in National Archives and Records Administration, 2014). The same applied to "Indians whose whereabouts have been unknown for a considerable number of years", who were also to be "dropped from the rolls" (Office of Indian Affairs, 1930*a*, cited in National Archives and Records Administration, 2014). Individuals that had received allotments—and technically, not living under the jurisdiction of the agency—were nevertheless to be included in the censuses. These changes mean that—post-1930—an individual's appearance in Indian censuses was informative about their ongoing attachment to or presence at the agency.

Importantly, appearance in an Indian census had consequences for tribal membership at the time, and implications for membership today. In particular, tribal constitutions that were written following the Indian Reorganzation Act of 1934 regularly used Indian censuses to form "base rolls", allowing descendants of individuals appear on such rolls to subsequently claim membership. For example, Article II of the Constitution of the Oglala Sioux Tribe provides that "Membership of the Oglala Sioux Tribe shall be automatic when: (a) The person's name appears on the official roll of the Oglala Sioux Tribe of the Pine Ridge Reservation as of April 1, 1935... (b) A child is born to any member of the Oglala Sioux Tribe."

Table A.6. Off-reservation School Opening Years

School	Location	Year opened
Hampton	*VA*	*1868*
Carlisle	PA	1879
Chemawa	OR	1880
Chilocco	OK	1884
Genoa	NB	1884
Albuquerque	NM	1884
Haskell	KS	1884
Grand Junction	CO	1886
Santa Fe	NM	1890
Fort Mojave	AZ	1890
Carson	NV	1890
Fort Stevenson	ND	1891
Pierre	SD	1891
Phoenix	AZ	1891
Fort Lewis	CO	1892
Fort Shaw	MT	1892
Perris	CA	1893
Flandreau	SD	1893
Pipestone	MN	1893
Mount Pleasant	MI	1893
Tomah	WI	1895
Wittenberg	WI	1895
Greenville	CA	1897
Morris	MN	1898
Chamberlain	SD	1898
Fort Bidwell	CA	1898
Rapid City	SD	1898
Riverside	CA	1902
Wahpeton	ND	1908
Bismarck	ND	1908
Cushman	WA	1912

Notes: This table shows the names, locations and opening years of off-reservation schools. The Hampton Institute (VA) – while not formally part of the off-reservation school system – served a comparable role.
Sources: Adams (2020) and Gregg (2018).

B Data

B.1 Data Sources

Annual Reports of the Commissioner of Indian Affairs: My main archival sources of data are annual reports by the Indian Office, known as the Annual Reports of the Commissioner of Indian Affairs. I focus on Annual Reports during the roll-out of the off-reservation school system, from 1879 to 1906. The content of Annual Reports varies from year to year, but usually includes individual reports by off-reservation school superintendents and Indian agents. School reports contain information on the general affairs of the school (e.g., buildings, staff, curriculum), as well as (but not always) information on the tribes, reservations, or agencies from which schools recruited students. Though not their focus, agent reports sometimes include information on transfers of children made to off-reservation schools.

The Annual Reports contain statistical tables at the agency-level with information on educational and assimilation-related outcomes. The variables in the statistical tables include: population, literate population, English-speaking population, population in "citizens" (western) dress, the presence of schools (day or reservation boarding) and church-going population. In order to obtain information on the pre-treatment characteristics of agencies, I digitise the available statistical tables for the years 1876 (the earliest available) to 1879.

In addition, each Annual Report contains a schedule listing all agencies and reservations in that year, along with the tribes occupying those reservations. Figure B.1 shows an excerpt of from the 1909 Annual Report. As no 1910 schedule was published, I use the 1909 schedule as the basis of my analysis (e.g., tribes mentioned in pre-1909 school reports are mapped to 1909 reservations when determining reservation treatment years).

1910 Indian Schedules: For the 1910 population census, specific sheets were used for collecting information on Native Americans: the so-called "Indian schedules". In addition to information collected on regular schedules (e.g., name, age, occupation, and birthplace), the Indian schedules included information specific to Native Americans, such as tribe, degree of Indian ancestry (blood quantum), and educational attainment. Figure B.2 provides an example of one such sheet.

Unlike the regular schedules, the Indian schedules have not been systematically digitised. The currently available 1.4 per cent oversample from IPUMS does not include all variables and, due to its small sample size, is not amenable to cross-census linking. As part of this project, I digitise information on all Native Americans enumerated using the 1910 Indian schedules in the contiguous 48 states (over 90 percent of the population), and link this information to IPUMS *histid* identifiers.[32] The resulting dataset includes person-level information on tribe, blood quantum, educational attainment, allotment status, and housing conditions.

Indian Censuses: From the mid-1880s onwards, Indian agents conducted local enumerations of the population under their jurisdiction. While these "Indian censuses" were conducted annually in principle, they were, in practice, less regular. Indian censuses were less detailed than the population census, but nevertheless included valuable demographic information on individuals enumerated, including information on tribe, blood quantum, allotment status, and, from 1930 onwards, whether the individual was residing on the reservation. In response to concerns about the accuracy of Indian censuses, agents were instructed to remove the names of individuals whose whereabouts was unknown in 1930 (National Archives and Records Administration, 2014). The collection of censuses — known as the Indian Census Rolls — has been digitised and made available through genealogical

[32]The Indian schedules were not used in Alaska.

websites. I access the Indian Census Rolls using *Ancestry.com*, and extract the names, ages, and locations (i.e., state and/or reservation) of all males who appeared in a local Indian census between 1930 and 1934. The censuses contain between 125,000 and 135,000 records. This figure is somewhat lower than the male Native American population in 1930 (169,000), as censuses were not conducted at the Union Agency.[33]

Historical Reservations and Settlements: I compile a list of reservations and Native American settlements under federal supervision as at 1910, using schedules from the Indian Office's 1909 and 1910 Annual Reports (Office of Indian Affairs, 1909, 1910). I identify most reservations using a reservation schedule from the 1909 Annual Report, and use a population table from the 1909 and 1910 Annual Reports to identify a number of settlements.[34]

B.2 Data Construction

Historical Reservations and Settlements: I compile spatial data on historical reservations from the early twentieth century using a variety of sources. I primarily rely on state-level maps of reservations published by the Bureau of American Ethnology Powell (1899), which I georeference.[35] Then, for each state, I overlay current reservation boundaries from IPUMS (Manson et al., 2023) on the georeferenced map. I use the following process to determine historical reservation boundaries:

- When current reservation boundaries closely match those on the corresponding historical map, I use current boundaries (e.g., the Pine Ridge Reservation, South

[33]While population statistics are not available in 1930, the Union Agency accounted for around 25 per cent of the Native American population in 1910.

[34]The additional settlements are: Camp Verde (Arizona), Fort Mojave (Arizona), Fallon (Nevada), Fort McDermitt (Nevada), Lovelocks (Nevada), Roseburg (Oregon), Flandreau (South Dakota), and Wittenberg (Wisconsin). I also add a settlement for the Cornplanter Reservation (Pennsylvania), which was listed as a place in the 1910 census.

[35]I georeference maps using latitude and longitude lines marked on the map. I use a third order transformation to convert the raster dataset to map coordinates.

Dakota).

- When current reservation boundaries do not match those on the corresponding historical map, I add or remove sections based on the description of the reservation in Powell (1899) (e.g., the Rosebud Reservation, South Dakota).

Additions and removals are usually related to administrative boundaries (e.g., county, state), natural features (e.g., rivers), Public Land Survey System (PLSS) coordinates. To maximise accuracy, I harmonise these additions to existing, high-quality spatial data. Specifically, I draw on datasets from Manson et al. (2023) (for historical state and county boundaries), U.S. Geological Survey (2023) (for rivers and streams), and Bureau of Land Management (2024) (for PLSS boundaries). Figure B.3 shows an example of the output, with current boundaries in orange and historical boundaries in blue. On the left, the current boundaries of the Pine Ridge Reservation closely match those on the corresponding map from Powell (1899). On the right, the current boundaries of the Rosebud Reservation are smaller than those of the historical one, so I reconstruct historical boundaries based using PLSS townships (eastern and southern boundary) and the White River (northern boundary).

Off-reservation School Treatment Years: Using primary historical sources, I construct the first dataset on reservation-level exposure to off-reservation schools during the nineteenth and early twentieth centuries in three steps.

First, I identify reports by off-reservation school superintendents between 1880 and 1906 from Annual Reports of the Commissioner of Indian Affairs.[36] I supplement this with reports by Indian agents over the same period. I identify any references to tribes, reservations, or agencies contained in school reports, and any references to off-reservation

[36]Three off-reservation schools were established after 1906. Two schools (Bismarck and Wahpeton) opened in North Dakota in 1908, and another (Cushman) opened in Washington in 1912.

schools contained in agency reports.

Second, for each year, I match references to tribes, reservations, or agencies in school reports to 1910 reservations using my schedule of historical reservations and settlements. I do the same with references to off-reservation schools in agency reports. As part of this process, I map pre-1910 reservations and agencies to their 1910 equivalents. This provides an initial set of treatment years.

Third, I use attendance records from Carlisle, Chemawa, Hampton, and Haskell in parallel to identify the reservations treated by these schools from their opening years until 1900. Since attendance records include information on tribes *and* home agencies/reservations, they often enable me to uniquely identify reservations that could not be identified on the basis of school reports alone.[37]

I combine matches from all the methods above, and identify the first year that a reservation sent students to an off-reservation school as that reservation's treatment year.

Western and Non-Western First Names: I classify the 55,000 unique names of Native Americans in the 1910 census as being of "Western" or "non-Western" origin in five steps. First, I extract the first names of all Native Americans from the restricted version of IPUMS' release of the 1910 full count census (Ruggles et al., 2021). I leave single-word names as is, and split multi-word names (e.g., "Mary Jane", "Ha-Gah-Kah John", "Naun Gaun Way We") into single-word tokens.

Second, I construct a list of "Western" first names and nicknames. This list is based on a set of around 6,000 first names (including misspellings and diminutives) used in publicly available codes from the Census Linking Project (Abramitzky et al., 2025). I supplement this list with around 3,600 Saint names compiled by Abramitzky, Boustan and Eriksson (2016), and around 1,300 Spanish names from *behindthename.com* (given that Spanish names

[37] For example, references to "Sioux" students in school reports cannot be matched to a unique reservation, as in 1910 various bands of Sioux occupied 13 reservations across five states.

were prevalent among Native Americans in parts of the Southwest).[38]

Third, I match single-word tokens to my list of Western first names, using exact matches and (if no exact match is found) fuzzy matches. For fuzzy matches, I require that both the target and candidate names are at least 4 characters in length, start with the same letter, have a Levenshtein edit distance of less than 25 per cent of target name (e.g., if the target name is 4 characters, then no more than 1 insertion, deletion or substitution is permitted) and have a Jaro-Winkler distance of less than 0.05.

Fourth, I use GPT-4o-mini to classify all tokens as "Western" or "non-Western", setting the model's temperature parameter to 0 to prioritise stability.[39] In line with best practices, my prompt includes explicit instructions and examples (OpenAI, 2025), and requires that the model provides a short justification for its answer (Lagakos, Michalopoulos and Voth, 2025). Figure B.4 shows the prompt in full.

Fifth, I combine directly matched and GPT-labelled classifications into a list of Western-origin tokens, merge with the original, tokenised names, and classify names as being of "Western", "Mixed", or "Non-Western" origin as follows:[40]

- If all tokens in a name (excluding single-character tokens) are matched to my list of Western-origin tokens, the name is coded as "Western". Based on this rule, "Mary", "Mary Jane" and "A Mary Jane" would be coded as "Western".

- If some but not all tokens in a name (excluding single-character tokens) are matched to my list of Western-origin tokens, the name is coded as "Mixed". Based on this rule, "Ha-Gah-Kah John" and "A Ha-Gah-Kah John" would be coded as "Mixed".

- If no tokens in a name (excluding single-character tokens) are matched to my list of

[38] All Spanish-origin names as at April 2025 were extracted from *behindthename.com*.
[39] The temperature parameter cannot be set with GPT-5 series models.
[40] I also manually code titles (e.g., "chief"), animal names (e.g., "bull", "eagle") and descriptors (e.g., "american", "man", "red", "white") as "non-Western" at this stage.

Western-origin tokens, the name is coded as “Non-Western”. Based on this rule, “Naun Gaun Way We” and “A Naun Gaun Way We” would be coded as “Non-Western”.

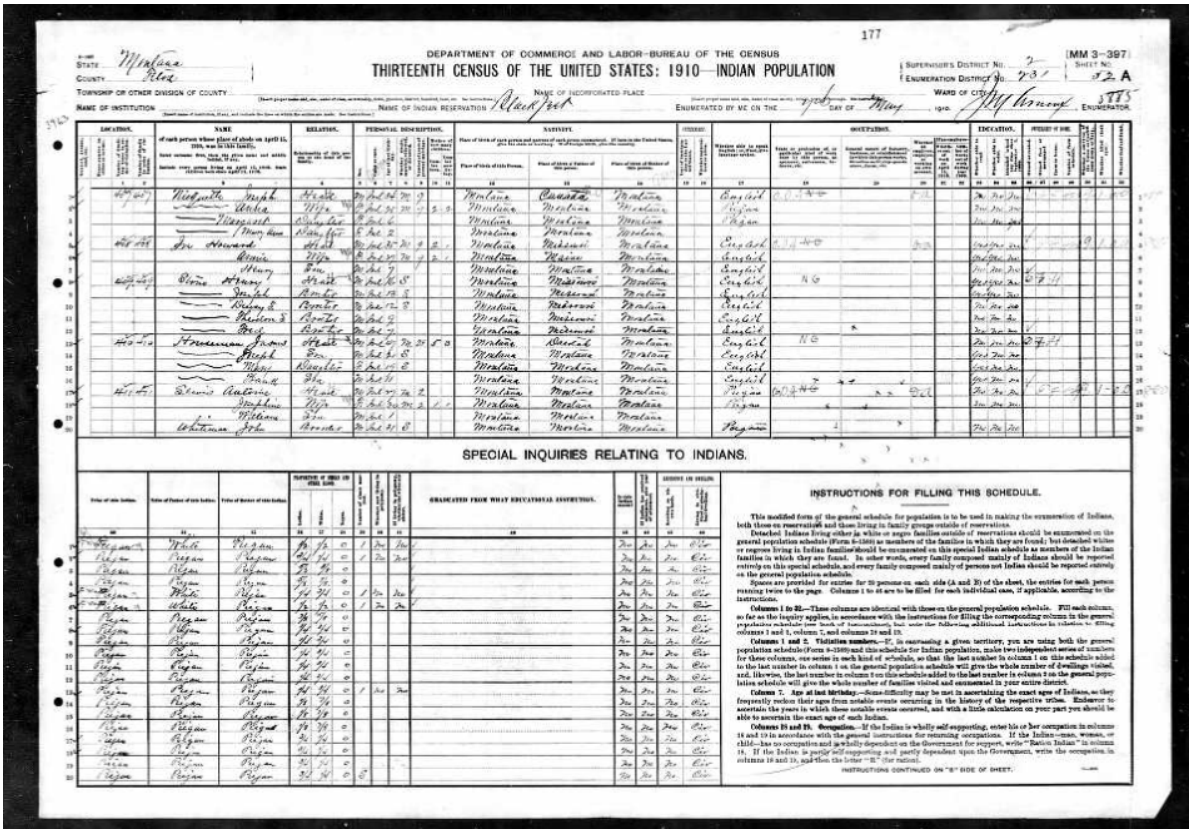
Figure B.1. Extract of Reservation Schedule, 1909

TABLE 47.—*Schedule showing each Indian reservation, under what agency or school, tribes occupying or belonging to it, area not allotted or specially reserved, and authority for its establishment.*

Name of reservation and tribe.	Area (unallotted).	Date of treaty, law, or other authority establishing reserve.
ARIZONA.	*Acres.*	
Camp McDowell..... (Under Phoenix School.) Tribe: Mohave Apache.	24,971	Executive order, Sept. 15, 1903; act of Apr. 21, 1904, vol. 33, p. 211. (See Ann. Rept. 1905, p. 98.)
Colorado River[a]..... (Under Colorado River School.) Tribes: Chemehuevi, Kawia, Cocopa,[d] Mohave.	[b] [c] 240,640	Act of Mar. 3, 1865, vol. 13, p. 559; executive orders, Nov. 22, 1873, Nov. 16, 1874, and May 15, 1876. (See sec. 25, Indian appropriation act, approved Apr. 21, 1904, vol. 33, p. 224.)
Fort Apache..... (Under Fort Apache School.) Tribes: Chilion, Chiricahua, Coyotero, Mimbreño, and Mogollon Apache.	[b] 1,681,920	Executive orders, Nov. 9, 1871, July 21, 1874, Apr. 27, 1876, Jan. 26 and Mar. 31, 1877; act of Feb. 20, 1893, vol. 27, p. 469; agreement made Feb. 25, 1896, approved by act o. June 10, 1896, vol. 29, p. 358. (See act of June 7, 1897, volf 30, p. 64.)
Gila Bend..... (Under Pima School.) Tribe: Papago.	[c] 22,391	Executive order, Dec. 12, 1882. (See 4106, 36409–9.)
Gila River..... (Under Pima School.) Tribes: Maricopa and Pima.	357,120	Act of Feb. 28, 1859, vol. 11, p. 401; executive orders, Aug. 31, 1876, Jan. 10, 1879, June 14, 1879, May 5, 1882, and Nov. 15, 1883.
Havasupai (Supai)..... (Under Havasupai School.) Tribe: Havasupai.	[b] 518	Executive orders, June 8 and Nov. 23, 1880, and Mar. 31, 1882.
Hopi (Moqui)..... (Under Moqui School.) Tribe: Hopi (Moqui).	2,472,320	Executive order, Dec. 16, 1882. Allotments being made under act of Mar. 1, 1907 (34 Stat. L., 1015–1021).
Navaho[e]..... (Under Leupp, Moqui, Navaho, Western Navaho, and San Juan schools.) Tribe: Navaho.	12,115,283	Treaty of June 1, 1868, vol. 15, p. 667, and executive orders Oct. 29, 1878, Jan. 6, 1880, two of May 17, 1884, and Nov. 19, 1892. 1,769,600 acres in Arizona and 967,680 acres in Utah were added to this reservation by executive order of May 17, 1884, and 46,080 acres in New Mexico restored to public domain, but again reserved by executive orders, Apr. 24, 1886, Jan. 8, 1900, and Nov. 14, 1901. Executive orders of Mar. 10, 1905, and May 15, 1905, 61,523 acres added to reservation, and by executive order of Nov. 9, 1907, as amended by executive order of Jan. 28, 1908, 2,972,160 acres were added. 470 Indians have been allotted 74.715 acres under the act of Feb. 8, 1887 (24 Stats., 388), as amended by the act of Feb. 28, 1891 (26 Stats., 794), and by executive order of Dec. 30, 1908, the surplus lands, approximately 506,000 acres, in that part of the extension in New Mexico east of the first guide meridian west were restored to the public

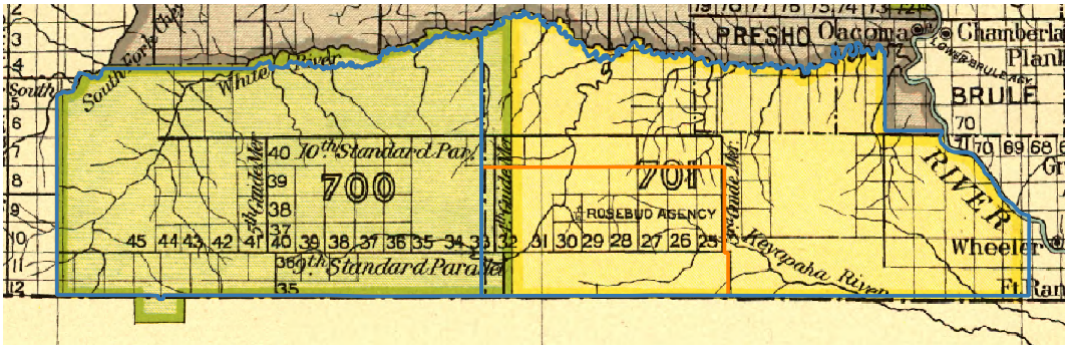
Notes: This figure shows an extract of the 1909 schedule of reservations.
Source: Office of Indian Affairs (1909)

Figure B.2. Indian Schedule from the 1910 Population Census



Notes: This figure shows an example sheet from the 1910 "Indian schedules". The top half includes standard information (e.g., name, age, birthplace, and occupation). The bottom half — or "Special Inquiries" — includes information specific to Native Americans (e.g., tribe, degree of Indian ancestry, and allotment status).
Source: Ancestry.com

Figure B.3. Construction of Historical Reservation Boundaries



Notes: This figure shows the polygons representing the Pine Ridge Reservation (left) and Rosebud Reservation (right), overlaid on a historical map from (Powell, 1899). Historical boundaries are shown in blue, and current boundaries are shown in orange.
Source: Own work, with current reservations from Manson et al. (2023).

Figure B.4. Prompt for Classification of Native American Names

You will be provided with the first name of a person from the 1910 U.S. census.

Your task is to identify whether the provided name is a "Western" name. For the purposes of this task, "Western" names include first names and surnames of Northern, Western, Southern, or Eastern European origin. English, German, Spanish, Portuguese, Italian, and French first names and surnames, including likely misspellings or phonetic spellings, should always be treated as Western names.

Please base your answer on the following rules:

- Answer "0" if the provided name is clearly not any of the following: a Western first name or surname, a diminutive form of a Western first name, or a nickname derived from a Western name (e.g., "Ee-Koo-Nah").
- Answer "0" if the provided name is short (two or three letters) and could plausibly be an Indigenous or non-Western name (e.g., "Ba", "De", "Ke", "Gah", "Na"), unless it is a common Western name with this exact spelling (e.g., "Ed", "Jim", "Tim").
- Answer "1" if the provided name is clearly one of the following: a Western first name or surname (e.g., "James", "Jose", "Smith", "Johnson"), a diminutive form of a Western first name (e.g., "Jim", "Maggie"), a nickname derived from a Western name (e.g., "Jimmy"), or is likely to be a misspelling or phonetic spelling of one of these (e.g., "Betsee" as a phonetic spelling of "Betsy").
- Answer "1" if the provided name is derived from Western literature or history (e.g., "Sherlock", "Ulysses", "Napoleon"), even if it was not a common name at the time.

Please use these examples and explanations as a guide:

- Example 1: "john" = "1", as "John" is an English first name.
- Example 2: "jose" = "1", as "Jose" is a Spanish first name.
- Example 3: "cooper" = "1", as "Cooper" is an English surname.
- Example 4: "betsee" = "1", as this is a likely phonetic spelling of "Betsy", a diminutive form of the English first name "Elizabeth".
- Example 5: "antuine" = "1", as this is a likely misspelling of "Antoine", a French first name.
- Example 6: "gaudlupe" = "1", as this is a likely misspelling of "Guadalupe", a Spanish first name.
- Example 7: "ba" = "0", as this is a short name that is plausibly Indigenous or non-Western.
- Example 8: "gah" = "0", as this is a short name that is plausibly Indigenous or non-Western.
- Example 9: "little" = "0", as this is a descriptor, not a Western name.
- Example 10: "walks" = "0", as this is a verb, not a Western name.

In addition, you must provide:

- A confidence score for your classification, on a scale from 1 to 10, where 1 indicates very low confidence and 10 indicates very high confidence.
- A justification, in no more than one sentence, which explains the reasoning behind your classification.

Please adhere to the following guidelines:

- Base your classification on the linguistic and cultural characteristics of the name itself. Do not speculate about the person's ethnicity.
- Return only a valid JSON on a single line with fields "label", "confidence", and "justification".
- "label" must be one of: "1", "0".
- "confidence" must be an integer from 1 to 10.
- Do not include explanations or any extra text.

C Empirics

C.1 Estimation Sample

Exclusion of Certain Reservations and Settlements: I exclude five reservations and one non-reservation settlement from my estimation sample on the basis of institutional factors and pre-treatment characteristics. First, I exclude individuals assigned to a non-reservation settlement in the vicinity of Pembroke, North Carolina, which corresponds to what is today the Lumbee Tribe of North Carolina. I do so because the Lumbee were not comparable—in terms of cultural and ethnic characteristics—to other Native American groups at the time. In particular, "full blood" individuals accounted around 6 per cent of the settlement population, compared to over 80 per cent among other never treated reservations—reflecting the settlement's historically mixed-race status (Pritzker, 1998, p. 552). Second, I exclude five reservations under jurisdiction of the then Union Agency (the Cherokee, Chickasaw, Choctaw, Creek, and Seminole reservations in eastern Oklahoma).[41] I do so because because these tribes (a) operated their own schools and (b) were not formally targeted by the off-reservation school system until the early 1900s.[42] While not as extreme as the Lumbee, these tribes were also substantially more assimilated at baseline that other never treated groups.

C.2 Outcomes

In this section, I provide definitions of the main outcomes used in my analysis. Where relevant, I also describe how I construct outcomes from underlying IPUMS variables. Variable names in uppercase letters refer to IPUMS variables; variable names in bold are

[41] I also exclude all individuals belonging to the corresponding tribes in adjacent states).

[42] The 1902 *Rules for the Collection of Pupils for Nonreservation Schools*, for example, states that "children of... the Five Civilized Tribes (Cherokee, Choctaw, Creek, Chickasaw, and Seminole, including Delaware and Shawnee incorporated with the Cherokee) cannot be enrolled in the government Indian schools" (Office of Indian Affairs, 1902).

those used in the paper.

Graduate of Off-reservation School: This outcome is based on Indian schedule entry: "Graduated From What Educational Institution". I code the outcome as 1 if any school from Table A.6 was reported.

Can Read and Write: This outcome is based on the IPUMS variable LIT. I code the outcome as 1 if an individual could read and write (LIT = 4), and 0 if they could not read and write (LIT = 1).

Speaks English: This outcome in based on the IPUMS variable SPEAKENG. I code the outcome as 1 if an individual could speak English (SPEAKENG = 2), and 0 if they could not speak English (SPEAKENG = 1).

Has Western Name: This outcome is based on the (restricted) IPUMS variable NAMEFRST. I code the outcome as 1 if an individual has a "Western" or "Mixed" first name, and 0 otherwise. Appendix B provides further information on the process used for classifying names as "Western", "Mixed", or "non-Western".

Is "Mixed Blood": This outcome is based on the Indian schedule entry: "Proportion of Indian and Other Blood: Indian". I code the outcome as 0 if the entry plausibly corresponded to "full" Indian ancestry (e.g., "full", "f", "all", "100", "1"), and 1 otherwise.

Has "Mixed Blood" Spouse: This outcome is based on the Indian schedule entry: "Proportion of Indian and Other Blood: Indian". I code the outcome as 0 if the entry plausibly corresponded to "full" Indian ancestry (e.g., "full", "f", "all", "100", "1"), and 1 otherwise. I use the IPUMS variable RELATE to identify spouses. I exclude non-married individuals from estimation.

Links to 1930s Indian Census: I code the outcome as 1 if an individual was linked, using the ABE-Jaro Winkler algorithm, from the 1910 census to any Indian census conducted between 1930 and 1934, and 0 otherwise.

In Labour Force: This outcome is based on the IPUMS variable LABFORCE. I code the outcome as 1 if an individual was in the labour force (LABFORCE = 2), and as 0 if they were not in the labour force (LABFORCE = 1).

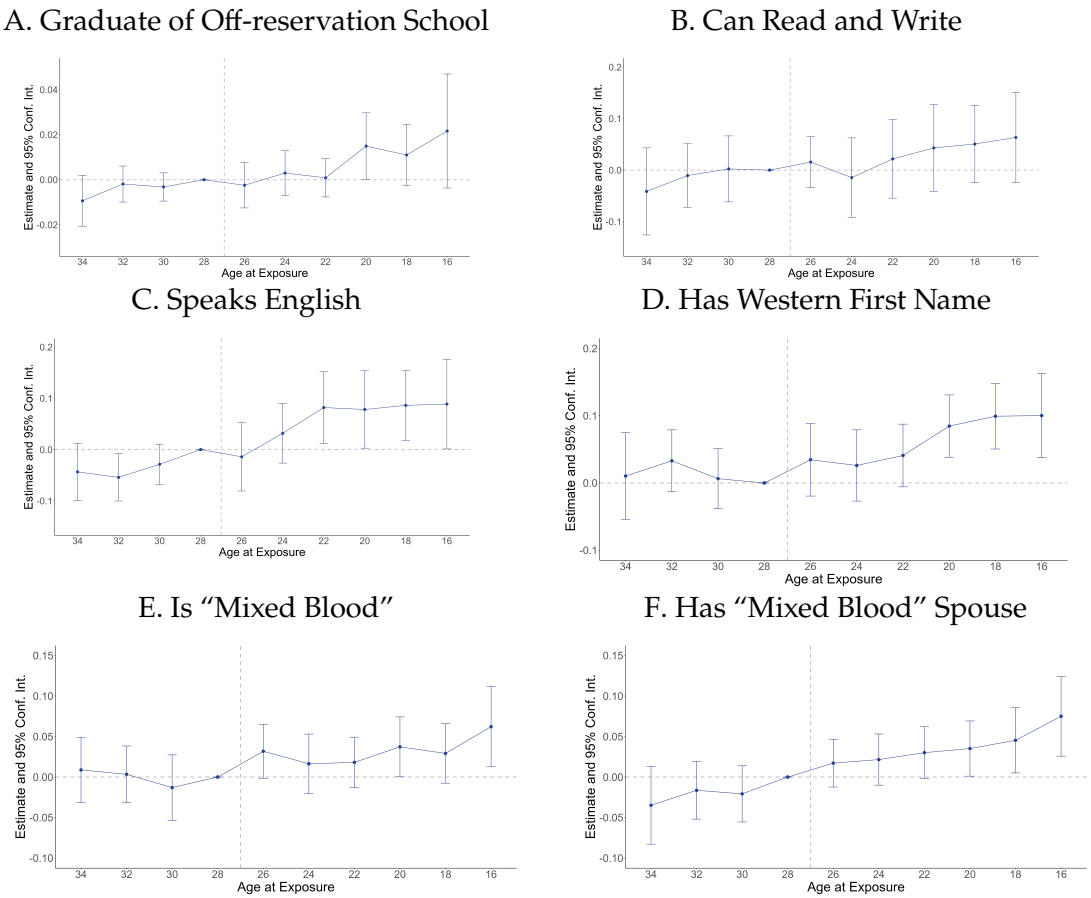
Broad Occupational Categories: This outcome is based on the IPUMS variable OCC1950. I group occupations into 4 broad categories following Long and Ferrie (2013). These are: "White Collar", "Farmer" (excluding farm labourers), "Skilled", and "Unskilled".

Ration Indian: This outcome is based on the raw occupation string in IPUMS' restricted release of the 1910 census. I code the outcome as 1 if the entry plausibly corresponded to "Ration Indian" (e.g., "ration", "rati"), and 0 otherwise. I exclude individuals with missing/unknown occupations (OCC1950 = 997 and 999) from estimation.

D Robustness

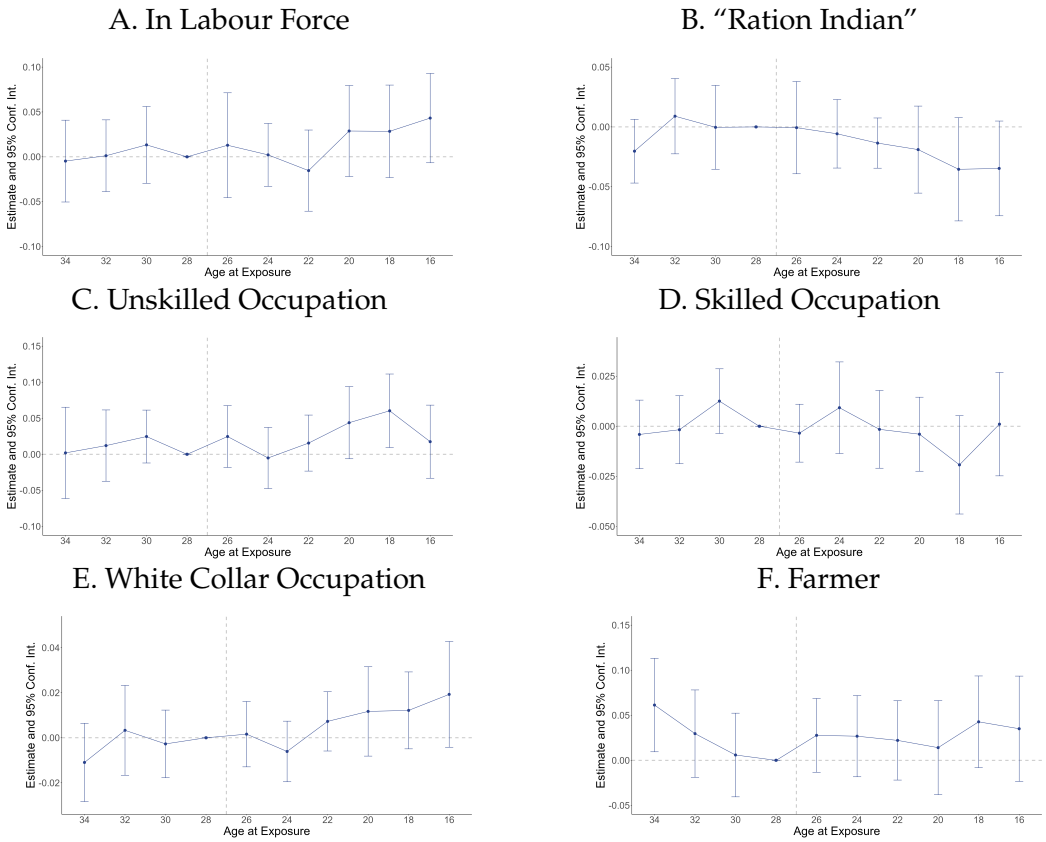
D.1 Sun and Abraham (2021) Estimates

Figure D.1. Educational and Cultural Outcomes, Sun and Abraham



Notes: This figure shows estimates from equation 1 for the listed outcomes using the estimator proposed by Sun and Abraham (2021). The sample consists of male, Native American household heads and aged between 20 and 65 in 1910. The sample excludes all individuals matched to reservations under the Union Agency, all Cherokee, Chickasaw, Choctaw, Creek and Seminole in Oklahoma and adjacent states, and all Native Americans at the non-reservation settlement in the vicinity of Pembroke, North Carolina. All outcomes are measured in 1910. All regressions include reservation fixed effects and cohort fixed effects (2-year bins). Standard errors are clustered at the reservation-level.

Figure D.2. Economic Outcomes, Sun and Abraham



Notes: This figure shows estimates from equation 1 for the listed outcomes using the estimator proposed by Sun and Abraham (2021). The sample consists of male, Native American household heads and aged between 20 and 65 in 1910. The sample excludes all individuals matched to reservations under the Union Agency, all Cherokee, Chickasaw, Choctaw, Creek and Seminole in Oklahoma and adjacent states, and all Native Americans at the non-reservation settlement in the vicinity of Pembroke, North Carolina. All outcomes are measured in 1910. "Ration Indian" is an indicator equal to 1 if an individual's occupation string was "ration" or similar. "Unskilled Occupation", "Skilled Occupation", "White Collar Occupation" and "Farmer" are based on the categorisation from Long and Ferrie (2013) and the IPUMS variable OCC1950. All regressions include reservation fixed effects and cohort fixed effects (2-year bins). Standard errors are clustered at the reservation-level.