SSAI Sociological Studies of American Indians

Working Paper 2021-04

Literacy Among American Indians: Levels and Trends from 1900 to 1930 and Across Birth Cohorts from 1830 to 1920

By
Arland Thornton
and
Linda Young-DeMarco

Institute for Social Research
University of Michigan
Ann Arbor, Michigan

April 2, 2021

The Sociological Studies of American Indians: ssai.isr.umich.edu
Institute for Social Research
The University of Michigan

This paper is part of a working paper series concerning the Sociological Studies of American Indians. These papers are not finished products but are “works in progress” and are made available for two purposes. The first purpose is to disseminate the information contained in the papers to people interested in the topics covered. The second purpose is to solicit comments, criticisms, and suggestions concerning the material presented, with the goal being to improve the presentation of results and the validity and usefulness of the interpretations and explanations provided. We invite readers to email such feedback to the author(s) of this working paper.

Literacy Among American Indians: Levels and Trends from 1900 to 1930 and Across Birth Cohorts from 1830 to 1920

by

Arland Thornton[1]

and

Linda Young-DeMarco[2]

University of Michigan

Ann Arbor, MI 48109

April 2, 2021

[1] Department of Sociology, Population Studies Center, Survey Research Center, and Native American Studies, arlandt@umich.edu.
[2] Survey Research Center, lyoungdm@umich.edu.

Abstract

We investigate levels and trends in literacy--being able to read and write--among American Indians in the United States. Using 1900-1930 decennial census data, we document levels and trends in reading and writing for the 1900 through 1930 period and for birth cohorts from 1830 through 1920. We thus provide for American Indians a large-scale picture of the history of literacy. We document the pace and extent of American Indian literacy from very low for the birth cohorts of the early 1800s to fairly universal for the cohorts of the early 1900s. We also demonstrate that the increases in Native literacy were closely related to birth cohort, with successive new birth cohorts having higher levels of literacy. We found little evidence that increases in reading and writing from 1900 to 1930 happened because adults increased their literacy after the school years and as they matured across the adult life course. We also document important gender differences in American Indian literacy, with the proportion literate being lower for women than for men, but with the gender gap decreasing in later birth cohorts. There were also substantial literacy inequalities across geographical regions of the country—ranging from 19 to 74 percent literate across regions in 1900. The trajectories of literacy attainment also varied across regions in interesting ways. We also document that American Indian literacy was higher among those living in urban areas, those more integrated into the Euro-American community, and those with Euro-American ancestry.

Introduction[3]

This paper investigates levels and trends in literacy--being able to read and write--among American Indians. We accomplish this using data from the 1900 through 1930 federal decennial censuses, documenting levels and trends in being able to read and write for people 10 years of age and older at the time of the 1900-1930 censuses. Because each of these censuses includes information about people ten years of age and over, we are also able to document levels and trends of literacy across birth cohorts from the early 1800s through 1920. Our results thus provide a large-scale perspective on the history of American Indian literacy across both census years and birth cohorts.

We also investigate how changes in American Indian literacy unfolded across time. One hypothesis is that such changes in reading and writing were the results of changes occurring across cohorts of individuals—that is, with each new cohort reaching adulthood having different experiences during childhood and the transition to adulthood. That is, it is likely that successive new cohorts of American Indians increasingly became more literate as they received more schooling in Euro-American institutions, both through compulsory enrollment in schools and through the voluntary efforts of their parents. Another hypothesis is that changes in literacy would have occurred within cohorts after

[3] We have received valuable input and guidance from many people concerning numerous aspects of American Indian history and culture and many dimensions of census data that have been very valuable in formulating and writing this paper. These knowledgeable and helpful individuals include: Lillian Ackerman, Trent Alexander, George Alter, Laurie Arnold, Dennis Baird, Larry Cebula, Harold Crook, Phil Deloria, Norm DeWeaver, Greg Dowd, Beth Erdey, Bonnie Ewing, Steve Evans, Rodney Frey, Joe Gone; Kayla Gonyon, David Hacker, Eric Hemenway, Fred Hoxie, Nicholas Jones, Susan Leonard, Carolyn Liebler, Kevin Lyons, Scott Lyons, Diane Mallickan, Rachel Marks, Kendra Maroney, Alan Marshall, Ruth McConville, Chris Miller, Melissa Parkhurst, Allen Pinkham, Josiah Pinkham, Alphonse Pitawanakwat, Nancy Shoemaker, Lindsey Willow Smith, Matthew Snipp, Jason Sprague, Zoe Higheagle Strong, Rebecca Thornton, Connie Walker, and Patsy Whitefoot. We also appreciate the helpful comments and suggestions provided by the participants in a seminar at the University of Michigan Population Studies Center. Assistance in preparation of the paper was provided by Lindsey Willow Smith and Jeffrey Swindle. At the same time that we appreciate the input and assistance of these individuals, we retain responsibility for any errors in the paper.

childhood, with literacy increases occurring among adults after the school years and as they matured through the adult life course. This would have occurred because the opportunities and needs for adults to gain literacy increased across time as they interacted with Euro-American institutions and with their own school-going and literate children. These between-cohort and within-cohort mechanisms are not mutually exclusive, and we hypothesize that the changes occurred through both mechanisms.

As just suggested, we began our research with the expectation that literacy is related to school attendance and the accumulation of school experience. This understanding motivated us to examine literacy in relationship to years of school completed. Among the issues we address in this paper is the question: to what extent was school attainment translated into literacy within the American Indian population?

While we examine the relationship of Euro-American school attendance to Native reading and writing, we note that Native parents and communities have been teaching their children since the beginning of time. Like all societies, American Indians taught their children how they believed the world worked and how people should live in that world, providing knowledge and skills about numerous aspects of Native society, economics, and culture. The Euro-Americans emphasized a school system that included a focus on reading and writing, the topics of our research.

In addition to documenting national levels and trends of American Indian literacy and examining their relationship to school attainment, we document differences in literacy levels and trends across other important social categories. Central among these potential social influences are gender and place of residence in the country; we examine differences between women and men and across nine different census regions. We also investigate differences between those residing in rural and urban areas, between those more and less integrated into the Euro-American community, and between those with

and without Euro-American ancestors. We document important differences in literacy across all of these social categories, providing insights into how literacy levels and changes varied across groups.

To our knowledge, this is the first investigation of national levels and trends in literacy among American Indians over a long time period. To be sure, there is a large and impressive literature documenting Native schooling, with a substantial focus on the experiences of American Indian children in schools. Parts of this literature consider the broad sweep of American Indian experience with schools (for example, Huff 1997; Reyner and Edor 2017; Coleman 1993; Fear-Segal 2007; Churchill 2004; Berkhofer 1972; Trafzer, Keller, and Sisquoc 2006; Adams 1995; DeJong 2007). Other parts take a local or case study approach, focusing on individual schools or tribes—or groups of schools or tribes (Baird, Mallickan, and Swagerty 2015; Lomawaima 1994; Gram 2015; Parkhurst 2014; Vučković 2008; Blackbird 1887; Karamanski 2012; Littlefield 1989; Widder 1999; Ehle 1988; Mehesuah 1998; Josephy 1965). Although this literature provides much understanding concerning the nature of American Indian schooling, it provides little information about Native literacy and how it varied across time and across important social categories—a contribution that we make in this paper.

Because of the newness of our research on American Indian literacy, we primarily take a descriptive approach and leave the task of interpreting and explaining trends and differentials to subsequent discussions. At the same time, we note that our research sets the stage for future examination into American Indian literacy, including how the levels and trends of literacy were affected by other key social and demographic factors such as tribal affiliation, experience with military conflict with the United States, economic interactions with Euro-Americans, such as through the fur trade, the timing and extent of land loss to Euro-Americans, and the presence of Christian missionaries.

Our research is significant on several levels. The focus of our project is on the descendants of the original inhabitants of what is now called the United States--people who have experienced centuries

of encounters with Euro-American diseases, military power, thirst for land, and social and economic organizations. During the period of our study, American Indians also experienced substantial isolation, discrimination, and disadvantage. Although much of the epidemiological, military, and land and population loss history has been narrated, many aspects of the social and demographic history of American Indians remain to be told. For example, what were the trajectories of school attainment and literacy across time? And, what were the forces that produced the trends and differential experiences and attributes? These are important issues that greatly deserve study.

The literature on Native schooling indicates large differentials, controversies, and disagreements concerning many aspects of school experience among American Indians. There is an extensive literature documenting that school attendance was forced on many American Indian families and their children through various mechanisms (for example, Coleman 1993; Lomawaima 1994; Fear-Segal 2007; Churchill 2004; Hoxie 2001; Vučković 2008; Trafzer, Keller, and Sisquoc 2006; Child 1998). This coercive element was likely associated with the many reports of parents and children resisting, sometimes vigorously, the requirement to attend school (for example, Coleman 1993; Lomawaima 1994; Fear-Segal 2007; Gram 2015; Churchill 2004; Berkhofer 1972; Vučković 2008; Karamanski 2012; Widder 1999; Ehle 1988; Trafzer, Keller, and Sisquoc 2006; Child 1998; Adams 1995). In contrast, there are also many accounts of American Indians being motivated by a variety of reasons to desire Euro-American schooling and literacy skills for themselves or for their children (for example, Josephy 1965; Baird, Mallickan, and Swagerty 2015; Huff 1997; Reyner and Edor 2017; Lomawaima 1994; Fear-Segal 2007; Gram 2015; Hoxie 2001; Berkhofer 1972; Cleland 1992; Vučković 2008; Blackbird 1887; Karamanski 2012; Littlefield 1989; Widder 1999; Ehle 1988; Trafzer, Keller, and Sisquoc 2006; Child 1998; Mehesuah 1998).

There are also disagreements and controversies about the experiences of American Indian children in schools and the effects of those experiences on adult life. There are numerous accounts of

neglect, abuse, and high levels of illness and mortality among Native school children (for example, Reyner and Edor 2017; Coleman 1993; Lomawaima 1994; Fear-Segal 2007; Churchill 2004; Trafzer, Keller, and Sisquoc 2006; Child 1998; Adams 1995; DeJong 2007), but there are also many accounts of Native children reporting positive experiences in these schools (for example, Reyner and Edor 2017; Lomawaima 1994; Vučković 2008; Littlefield 1989; Trafzer, Keller, and Sisquoc 2006). In addition, many American Indians learned few skills of value for adult life or suffered in adulthood because of their school experiences (for example, Coleman 1993; Lomawaima 1994; Churchill 2004; Vučković 2008; Trafzer, Keller, and Sisquoc 2006), but many others obtained skills in school, including reading, writing, and arithmetic (for example, Coleman 1993; Lomawaima 1994; Berkhofer 1972; Vučković 2008; Karamanski 2012; Littlefield 1989; Ehle 1988; Trafzer, Keller, and Sisquoc 2006; Adams 1995; Mehesuah 1998). Despite the many disagreements and controversies, there is widespread agreement that many Euro-American schools taught American Indian children Euro-American ways while vigorously working to extinguish Native culture, language, religion, and lifestyles (for example, Coleman 1993; Fear-Segal 2007; Gram 2015; Parkhurst 2014; Churchill 2004; Vučković 2008; Trafzer, Keller, and Sisquoc 2006; Child 1998; Adams 1995).

It is likely that the different perspectives and controversies over the nature of American Indian schools reflect differences across time and place in the experiences of individuals and groups—and differences in the emphases of the writers. While we recognize the importance of understanding these issues in the lives of American Indians—and evaluating the different experiences and perspectives--such investigations are outside the scope of our research. Our goal is not to adjudicate whether the general processes and overall outcomes associated with the expansion of literacy and Euro-American schooling were good or bad or a mixture of the two, but to document and understand the extent of literacy and the growth of literacy during this important period of time. We approach our project from the disciplines of sociology and demography, with the goals of documenting levels and trends of literacy and

how those levels and trends were experienced across a wide array of social and demographic groups. We believe that the documentation of such levels, trends, and differentials have relevance for the research literatures outlined above, but we leave those implications for future research and conjecture.

We recognize as we address our aims that the concepts of "race" and "American Indian" have long been central in American society and to the decennial census, and both are problematic terms. Despite the long tradition of racial categorization in American society and in the census, race has become a suspect category, with no scientific basis for distinguishing among groups identified as separate races. Nevertheless, race continues to be a social category widely used in society and the census (Hirschman 2004; American Anthropological Association 1998; Jacobson 2002; Sollors 2002; Bentley 2003; Office of Management and Budget 1997).

The category of "American Indian" was not one originally invented by Native peoples of what is now known as the Western Hemisphere. In the year 1491, there was, of course, no "America", and India was in Asia. The Natives of the now-called Western Hemisphere thought of themselves by more specific names such as Niimiipuu, Haudenosanee, Ojibwe, and Lakota (Lyons 2010, pages 66-69; Thornton 1989, pages 186-189; Nagel 1996, pages 3-4). The American Indian label was created by Europeans as part of their colonization process to refer in general to the original peoples of the hemisphere and their descendants. The problematic nature of this language has led to the creation of such terms as "Native Americans" and "First Nations". These terms, however, have their own difficulties; for example, in 1491, there was no existing America where one could be considered a "native." In addition, the term "native American" is sometimes used to mean someone who was himself or herself born in what is now known as America rather than someone who is a descendent of someone living in what is now America in the year 1491.

With these considerations, we interchangeably use the three terms mentioned above, but rely primarily on the category of "American Indian" for two reasons. First, American Indian is the category used by the Census Bureau in the data collections we discuss. And, second, many descendants of people resident in the Western Hemisphere in the year 1491 refer to themselves and their groups as American Indians (Lyons 2010, pages 66-69)[4]. Nevertheless, as we use the language of "American Indian", "Native American", and "First People", we recognize the colonial and problematic nature of the categories. We also use the term "Native" in this paper as a shorthand term for "American Indian".

We now turn to our primary data source—the 1900-1940 federal decennial censuses—documenting both the nature of the censuses and explicating some of their weaknesses for the study of American Indians. As part of this discussion, we explain the data files that we use in our analyses and document the measurement of literacy and school attainment. We then turn to our results investigating the substantive issues outlined above.

Data Sources

The 1900-1940 Decennial Census Program

The primary data for this project are the 1900, 1910, 1920, 1930, and 1940 decennial censuses conducted by the U.S. Census Bureau. These five decennial censuses were designed to enumerate the entire population of the United States, including American Indians, from 1900 to 1940. With minor exceptions discussed below, the same enumeration schedules used for the general population were also used for American Indians. The 1900-1940 body of census data contains comparable data on several

[4] This observation can be illustrated by data from the 2000 census that asked people: "What is your ancestry or ethnic origin?". People could provide one or more ancestries. Of those who indicated "Native American" or "American Indian" as their first ancestry, 73 percent indicated "American Indian" (U.S. Census Bureau 2007, Table 1). The number reporting themselves as "Native American" was only slightly higher than the number reporting "Indian" without any modifier, and it is very likely that many of those saying "Indian" without a modifier would have thought of themselves as "American Indian".

important attributes such as geographical location, age, sex, and literacy. For this paper, we focus entirely on the data for individuals enumerated as American Indians residing in the coterminous U.S. and exclude residents of Alaska and Hawaii.

In addition to the questions asked of all residents in the U.S. irrespective of race, both the 1900 and 1910 censuses asked a large part of the American Indian population additional questions (Thornton and Young-DeMarco 2021). For 1900 the Census Office reported that American Indians living on reservations, in "tribal relations", or in "counties containing any considerable number of Indians" were asked the additional questions (Census Office 1900, page 9). In 1910, the additional questions asked in the census "were used in all the districts containing Indians on reservations and throughout counties where as many as 20 Indians were returned at the census of 1900" (U.S. Bureau of the Census 1915, page 9). The questions included in these special American Indian supplements that are relevant for our paper concerns the amount of White ancestry in the person's lineage.

The 1930 census took a different approach for obtaining additional information about Amercan Indians. Information about all American Indians, without respect to geographical location, was ascertained using the same form as used for the general population. The Census Bureau reported that "In the enumeration of Indians, however, instead of asking for the State or country of birth of the person's mother and father, the enumerator was instructed to ask whether the Indian was of full or mixed blood and for the tribe to which he or she belonged" (U.S. Bureau of the Census 1937, page 1). The added questions on tribe and ancestry were written in the blank columns vacated by the two questions about the place of birth of parents.

Our choice of the 1900-1940 decennial censuses for our study is motivated by the fact that individual level decennial census data for all American Indians are first available for empirical analyses beginning in 1900. American Indians who were living away from communities of high Native

concentration were included in the censuses from 1790 through 1880, but American Indians were not included in these censuses if they were living on reservations or other largely Native communities. Beginning in 1890, the censuses were designed to include all American Indians. However, almost all the 1890 individual-level manuscript census records were destroyed in fires, making them unavailable for analysis. As we discuss below, the 1900-1940 census data are publicly available in individual-level files, but the 1950 and onward data are not publicly available in such files.

Individual Data Files for Analysis.

For our empirical analyses of these censuses, we rely on the complete count and sample data files disseminated as part of the University of Minnesota's Integrated Public Use Microdata Series (IPUMS) (Ruggles, Flood, Goeken, Grover, Meyer, Pacas, and Sobek no date). Our first source of IPUMS data is the IPUMs complete count data files for 1900-1940 that were created and disseminated by Ancestry.com, Family Search, and the University of Minnesota to include 100 percent of all residents of the U.S. (IPUMS USA, No Date(a)). One limitation of these complete-count census files is that, to our knowledge, each of the 1900, 1910, and 1930 files publicly available from IPUMS does not include the information obtained in the special American Indian census supplements those years.

Our second source of IPUMS data is the independently-created sample data files for 1900 through 1940 for all residents of the U.S. (IPUMS USA No Date (b)). The 1900 and 1910 IPUMS samples for American Indians actually combine two different subsamples. The first subsample for each of these years includes the records of American Indians enumerated with the regular enumeration schedule, but does not include the records of American Indians enumerated with both the regular enumeration schedule and the special supplementary American Indian schedules used in those years. The second subsample for both 1900 and 1910 is a special twenty percent sample of American Indians enumerated with both the regular enumeration schedule and the special supplementary enumeration schedule

(Minnesota Population Center 2005; Hacker and Haines 2005, 2011). Together, these two samples, when properly weighted, permit making population estimates about the total American Indian population included in the 1900 and 1910 censuses.

Race in the Decennial Censuses

For each decennial census from 1900 to 1940, the Census Bureau identified racial categories for grouping people. However, both the number and content of racial categories used by the Census Bureau changed from 1900 to 1940, with both additions and deletions of categories between censuses (Bennett 2000; Snipp 2003; Thornton and Young-DeMarco 2021; U.S. Bureau of the Census No Date). The racial categories used in 1900 were American Indian, White, Black, Chinese, and Japanese. In 1910 Mulatto was added to this list, and in 1920 the 1910 list was expanded to include Filipino, Hindu, Korean, and "other". In 1930, the Mulatto category was dropped and a category for Mexican was added. The Mexican category was then dropped in 1940. The category "American Indian" remained constant across the five censuses.

As Jobe (2004, page 75) has stated, "Through the Census of 1950, the race of an individual was determined by the enumerator". A 1950 special census report about non-whites indicated that information about race was not ordinarily based on a question to the respondent but on the observation and judgement of the enumerator (U.S Bureau of Census 1953, page 3B-4). As discussed by Thornton and Young-DeMarco (2021), the instructions provided by the Census Bureau to enumerators to decide the race of individuals were very brief and varied across time—and were not clear and precise in specifying exactly who was an American Indian.

Completeness of Coverage and Reporting of American Indians

Our interpretation of the census data is affected by our understanding that many people who were descendants of residents of North America in the year 1491 were not enumerated in the 1900-1940 censuses. Undercounts are common in censuses and are particularly common among minorities and people with fewer resources (Hacker 2013; Hoy 2015; Snipp 1997, 2003; Lujan 2014; Meister 1980; Kesslen 2019). The conditions producing undercounts among Native peoples were likely exacerbated in the early 20th century by differences in language and culture and by a distrust by American Indians of Euro-American institutions that led to a desire to be invisible to Euro-American authorities, including census enumerators (Hoy 2015; Lujan 2014; Liebler 2018; Kesslen 2019; Jobe 2004; Thornton, Young-DeMarco, and Smith 2021).

There are also substantial reasons to believe that numerous descendants of residents of North America in the year 1491 who were enumerated in the 1900-1940 censuses were not recorded as American Indian but as another race (Thornton, Young-DeMarco, and Smith 2021). This understanding is supported by the view of a prominent American Indian—Arthur Parker—who wrote regarding the 1910 census that "It is to be seriously doubted that the census enumerators obtained anywhere near the exact number of persons of Indian blood who are merged in the general population. The writer knows of many who refuse for various reasons to be regarded as Indians….." (Parker 1915, pages 205-207). A similar assessment was provided by the Census Bureau concerning the 1930 census results: "if all persons having even a trace of Indian blood were returned as Indians, the number would far exceed even the total returned at the census of 1930" (U.S. Bureau of the Census 1937, page 2). A more recent assessment by Jobe (2004, page 75) came to a similar conclusion: "In the late nineteenth and early 20th centuries, many Indians were probably overlooked by the census because they did not appear to be Indian, did not live on Indian reservations, were not recognized by the community as Indian, or chose to hide Indian ancestry from enumerators".

Three bodies of evidence from decennial censuses between 1960 and 2000 also support the idea that previous censuses enumerated and categorized as American Indian substantially fewer individuals than the number of people with ancestors present in North America in the year 1491. First, the censuses after 1950 switched from enumerator determination of race to self-identification of race and discovered that many people who would have been identified as non-Native by enumerators identified themselves as Native (Passel 1976, 1996; Snipp 1989, 2003; Jobe 2004; Eschbach 1993; Thornton, Young-DeMarco, and Smith 2021). Second, beginning in 1980 the censuses asked people to identify both their race and their ancestry and found that many more people identified themselves as having Native ancestry than identified their race as American Indian (Snipp 1989, 2003; Liebler 2018; Passel 1996). Third, beginning in 2000, people were allowed to record more than one racial identity and many reported that they were both American Indian and some other race (Bennett 2000; Snipp 2003; Liebler and Ortyl 2014; Perlmann and Waters 2002)—many of whom would have reported themselves as non-Native if they were asked to report only one race (Bentley et al 2003). Although we cannot know for sure how these post-1940 results apply to the early 1900s, it is likely that just as large fractions of the people of Native ancestry in the later period were enumerated as non-Native, it is likely that large fractions of people of Native ancestry during the early 20th century were also enumerated as having a race other than American Indian.

It is also likely that the recording of Native people as American Indian was substantially more common among those living on or near reservations and other Native communities than among those who were "merged in the general population" (Parker 1915, pages 205-207; also U.S. Bureau of the Census 1937). As we document later in our results section, literacy rates tend to be higher among those who are more integrated into the general Euro-American population, suggesting that the lower representation of the less-integrated people in the census would skew the results toward lower levels of literacy. This expectation is also consistent with research from post-1940 censuses showing that people

located outside areas of high concentrations of American Indians and people who identify less closely as Native have, on average, higher levels of socioeconomic attainment, including more years of school attainment (Snipp 1989; Eschbach 1993; Eschbach et al. 1998; Liebler and Ortyl 2014; Passel 1996; DeWeaver 2003).

These considerations lead to the expectation that our results from the 1900-1940 censuses will only reflect the literacy of a subset of the population descended from the inhabitants of North America in the year 1491. Furthermore, this subset is drawn disproportionately from those Natives who were most integrated into American Indian communities and least integrated into Euro-American communities and are likely to have lower levels of literacy than the total population of descendants of people living in North America in the year 1491. However, our analyses of trends will be little affected if the types of people included in the censuses remained constant across time. In this regard, our confidence is bolstered by the Census Bureau's observation that a general study of the 1910 and 1930 censuses "leads to the conclusion that both censuses were reasonably accurate and comparable with each other as regards those Indians" it covered, which as we have noted were those most closely associated with Native communities (U.S bureau of the Census 1937, page 1).

We note, however, that a systematic evaluation of the comparative coverage of the 1900-1940 censuses revealed that there were likely some differences in the level of coverage (Thornton and Young-DeMarco 2021). That study suggests that the 1910 and 1930 censuses likely had better coverage and/or more identification of people as American Indian than did the 1900 and 1920 censuses. We will return to these observations when we discuss the levels of literacy reported.

Measurement and Coding of Literacy and School Attainment

Literacy was ascertained in the 1900 through 1930 censuses, but not in 1940 and later. This requires us to restrict our analyses of literacy itself to the 1900-1930 censuses. School attainment, on

the other hand, was ascertained only in 1940 and not in earlier censuses. This prevents conducting analyses directly comparing literacy and school attainment at the individual level. However, by moving our analyses to the cohort level, we can compare literacy and school attainment across different birth cohorts—an analysis that we describe in more detail later.

Measures of literacy were collected only for individuals ages 10 and older. A person was not required to be able to read and/or write in English in order to be considered literate; being able to read and/or write in any language counted. For the 1900-1920 censuses the enumerator collected the information using two questions – one for reading and one for writing while in 1930, one question obtained both pieces of information.

Beginning in 1910, a clarification concerning the literacy of blind individuals was added to the enumerator instructions. Enumerators were told to record literacy in the affirmative if the person could read and/or write prior to becoming blind, or had been taught how to read and/or write if blind at birth. The 1930 census year was the only time enumerators received any instruction relative to determining whether any individual was *not* literate, i.e. "Do not return any person as able to read and write simply because he can write his own name."

It is also important to note that the measurement of literacy is, of necessity, crude. There were no tests given to determine whether a person could read and/or write. We only have the judgement of the person providing the information to the enumerator—or perhaps in some cases, the enumerator's own judgement. Such judgements could affect the results through both random measurement error and by biases that might be held by the enumerator. In our analyses, we categorized a person as literate if the census recorded the person as able to read and/or write and "0" if the person could do neither.

School attainment was ascertained for all persons in 1940 (but not the earlier censuses) with a question asking about the highest grade of school completed. Enumerators were instructed to record

the highest full grade of school completed rather than half years or uncompleted grades. Children not yet of school age and individuals who had never attended school were recorded as zero, Training at vocational schools or colleges were not to be included unless they were part of a regular school system. Since substantial amounts of American Indian experience in schools involved vocational training, this procedure likely led to an underestimation of actual schooling. We recoded the detailed IPUMS variable for years of school completed so that it ranged from 0-17+ and analyzed it in two ways: by estimating mean years completed and the percentage who attended school for one or more years.

Results

National Literacy Levels and Trends from 1900 to 1930

In Table 1, we list the reported percentages literate and illiterate for each of the four census years from 1900 to 1930 from three different data sources: the IPUMS full count data; the IPUMS sample data; and the official government publications for the respective census years. A comparison of the numbers in Table 1 indicates high levels of consistency across the three data sources for 1910, 1920, and 1930; the largest difference among the three estimates is 2.6 percentage points. However, for 1900, the difference between the government report and the IPUMS full count data is 6.2 percentage points, whereas the difference between the government report and the IPUMS sample is smaller.

The close fit between the government and IPUMS complete count literacy data for 1910, 1920, and 1930, along with the complete population of people in these files, led us to use the complete count files as our primary individual-level data sources for these years. The 6.2 percentage point discrepancy between the government and IPUMS complete count data for 1900, along with the closer fit between the government and IPUMS samples, led us to use the IPUMS sample for our main individual-level data source for that year. We also rely on the sample data for 1910 and 1930 for analyses of American Indian supplemental data only available on the sample files.

We begin our discussion of substantive results by examining the reported levels and trends of literacy from 1900 to 1930 documented in Table 1 using the published government census reports. These data indicate that approximately 44 percent of people identified as American Indian in the 1900 census were literate. Also recall that we estimate that this percentage is likely to be an underestimate for all descendants of Native Americans, as census under-enumeration likely produced results that are weighted more toward those who were less integrated into Euro-American society and towards those with less school attendance and literacy.

Table 1 also documents that the 1900 to 1930 period was a time of substantial increases in the percentage reported to be literate. According to the published census reports, the increases in literacy are about 10 percentage points per decade across the three time periods, reaching 74 percent in 1930. These are substantial increases, especially when recognizing that these differences reflect percentage changes in the entire Native population and not for successive cohorts of young people—a comparison we examine later. As indicated earlier, the levels and trends reflected in the IPUMS samples and complete-count data vary relatively slightly from the published census reports.

Comparisons of American Indians with Whites and Blacks

In order to put these levels and trends in literacy for American Indians in context, we provide similar data in Table 2 for Whites and Blacks for the four census years. These cross-group comparisons indicate that as early as 1900 reported literacy was nearly universal for Whites—being 94 percent. The percent literate in 1900 was much lower for Blacks and American Indians. Furthermore, the reported 55 percent literacy rate for Blacks was more than ten percentage points higher than the 44 percent literacy rate for American Indians. The percentage point increase from 1900 to 1930 was similar for Blacks and American Indians, with the gap in 1940 being about ten percentage points. Both Blacks and American Indians narrowed the literacy gap with Whites, but even in 1930 reported literacy levels remained

substantially higher for Whites. We noted earlier that there are likely biases in the American Indian literacy data, and there may also be biases in the data for Blacks and for Whites.

National Levels and Trends of American Indian Literacy by Age and Birth Cohort

Table 3 documents literacy levels by year and age, showing very large and important age differences. Literacy levels for 1900 start at 3-6 percent for the people ages 85 and older. The percentages literate increase to around 10 percent for those ages 60-84 in 1900 and to around 25 percent for those 40-49 in 1900. The percentage literate increases even faster for yet younger people, reaching nearly one-half for those 25-29 and nearly two-thirds for those 15-19 in 1900. Similar very large age patterns exist for the other census years, although the particular numbers and trends vary across the years. The percentage literate reaches 87 percent in 1930--or close to universal—for both those ages 10-14 and 15-19. Also, as we document below, there is a tendency for literacy to increase as children mature from 10-14 into adulthood, which means that the 87 percent literacy rate for those 10-14 in 1930 is likely to reach somewhat higher levels as these children reach adulthood.

We can also consider the age differences in Table 3 as birth cohort differences. This is true because age and birth cohort are perfectly related, with birth cohort simply indicating the census year minus age. We adopt this birth cohort approach in Table 4 where we combine the data for the four census years into one data file and estimate the percentage literate for those born in 1830 or earlier and for each subsequent five-year birth cohort through 1915-1920. We collapse the earliest birth cohorts into one group born in 1830 or earlier, because of the small population in the five-year birth cohorts before 1830, because of age reporting difficulties at older ages, and because of the inconsistency of estimates for the early birth cohorts.

The data in Table 4 document much of the history of literacy growth among American Indians from the early part of the 19th century through the early part of the 20th century, with very few being

literate among the earliest birth cohorts to nearly universal literacy among the most recent birth cohorts. More precisely, literacy increased from around 10 percent for those born by 1830 to around one-third for those born between 1861 and 1865 and to around two-thirds for the birth cohorts of 1881-1885 and 1886-1890. And, by the time of the birth cohorts of 1911-1915 and 1916-1920, the percent literate had increased to 87 percent. The percent literate for the birth cohort of 1916-1920 (age 10-14 in 1930) is likely to increase to somewhat higher levels as it ages from childhood into adulthood. These results thus suggest that literacy increased, at least in part, because successive birth cohorts of individuals reached adulthood with higher levels of literacy. We will return later to the importance of birth cohort as an explanation for the increases in literacy observed over census years.

Although we have interpreted these birth cohort differences as indications of increasing levels of literacy across successively more recent birth cohorts, we must recognize that the cohort patterns displayed in Table 4 could also be affected by differential mortality. Importantly, the data in Table 4 represent literacy levels for people in a birth cohort who have survived to a particular census year and not all people in the birth cohort—and for the earlier birth cohorts the percentage who survived to the 1900-1930 censuses would have been quite small. This produces the possibility that mortality disproportionately removes either the less literate or the more literate from the surviving population, leaving in the population a group that is either more or less literate than would have been the case at younger ages. That is, if mortality differentially removed the most literate from the population, it would lead to lower levels of literacy among the older people at the time of the censuses who were also of the earliest birth cohorts. And, if mortality differentially removed the least literate from the population, it would lead to higher levels of literacy among the older people at the time of the censuses who were also of the earliest birth cohorts. This potential source of bias could also affect all comparisons across cohorts--with the differential bias potentially varying across gender, region, and other population characteristics, with potential influence on those comparisons.

We do not know the direction and magnitude of differential mortality among American Indians in the 19th and early 20th centuries. However, it is generally understood that mortality differences tend to remove those with less schooling and the less wealthy—and presumably the less literate—in a population. If this were true among American Indians in the 19th and early 20th centuries, it would lead to the opposite birth cohort pattern observed in Table 4. In fact, if mortality differentials removed the less literate people more, it would suggest that the lower literacy of the earlier cohorts occurred despite mortality differentials leading in the opposite direction. On the other hand, if mortality differentially removed the more literate, it would contribute to the pattern observed in Table 4—lower levels of literacy among the earlier birth cohorts of people who were also older at the censuses. While it is possible that differential mortality removing the most literate could have contributed to the birth cohort differentials observed in Table 4, it is unlikely that such differential mortality could have produced the entire large differentials observed across birth cohorts. The differences across the later birth cohorts appear to be too large to be explained entirely by differential mortality removing the more literate. Mortality would have to have been very high and very strongly biased to removing the more literate to produce the very large differences observed, especially across the later birth cohorts. And, as previously observed, such mortality differentials selecting the more literate would be counter to the differential mortality patterns observed for other populations.

Comparison of American Indian Birth Cohort Trends with Whites and Black

Earlier, we observed that American Indian literacy in 1900 was substantially lower than the nearly universal level of White literacy and somewhat less than Black literacy—and that between 1900 and 1930 Black and American Indian literacy increased similarly and substantially narrowed the very large earlier gap with Whites. In Figure 1, we take another approach to this question and examine levels and trends across birth cohorts of American Indians, Blacks, and Whites. This figure demonstrates that

the near universality of literacy that we observed for Whites in 1900 was already largely in place by the birth cohorts of the early 1800s—and became even more universal across subsequent birth cohorts. In sharp contrast, Black literacy for the earliest birth cohorts was just less than one-fifth, but still somewhat higher than the one-tenth figure for American Indians. Literacy for both American Indians and Blacks increased dramatically over the next nearly hundred years of birth cohorts. Interestingly, the Native-Black difference appears to have widened across cohorts and then subsequently narrowed, virtually disappearing for the latest birth cohorts.

Trends for Birth Cohorts across Censuses and Ages

Table 3 contains data that can provide insights into whether actual individuals in various cohorts became more literate as they aged across censuses. That is, did literacy increase among cohorts of adults as they aged from one census to another? That question can be answered by following the data for particular cohorts across time in a downward step-wise staircase fashion in Table 3. We make such patterns more explicit in Table 5 where we present interdecadal literacy transition ratios by dividing the percent literate for an age (or birth cohort) group at one census by the percent literate for the same group at the previous census. We do this for age groups—first for 5-year age groups and then for one larger age group.

Table 5 demonstrates a pattern of interdecadal literacy transition ratios that vary considerably across the three interdecadal periods. We see that the literacy transition ratios are almost entirely above one for the 1900-1910 and 1920-1930 intervals and almost entirely below one for the 1910-1920 interval. In fact, our summary measure that focuses on the early and middle-adult years—transitioning from ages 20-49 to ages 30-59—are respectively 1.10, .93, and 1.05 for the three consecutive intervals. Taken at face value—and ignoring such things as the possibility of differential mortality and census coverage—these results suggest that the 1900-1910 and 1920-1930 periods witnessed significant

increases in literacy during adulthood while the 1910-1920 period was one of significant decline in adult literacy. Although this is possible, we find the pattern to be unlikely because it implies that the pattern of literacy change across adulthood varied greatly by census intervals—being positive for the first and last intervals and negative for the middle interval. We have no plausible explanation for such a result.

We believe that the pattern of interdecadal literacy transition ratios presented in Table 5 can be better understood by focusing on differences in census coverage across the four censuses. Recall that in our earlier discussion of differential census coverage, we indicated that coverage was greater in 1910 and 1930 than in 1900 and 1920. If greater census coverage brings in proportionately higher numbers of literate people into the American Indian census population, as we suggested earlier, then we would expect that observed literacy would increase for the same birth cohorts of people between 1900 and 1910 (when coverage increased) and for the same birth cohorts of people between 1920 and 1930 (when coverage again increased), which is exactly what the patterns in Table 5 suggest. In addition, we would expect that observed literacy would decline between 1910 and 1920 (when census coverage declined)—a pattern that is also observed in Table 5. This suggests that the interdecadal transition ratios reported in Table 5 may better reflect completeness of census coverage and a positive relationship between coverage and observed literacy levels than increasing or decreasing levels of literacy across birth cohorts as they aged ten years.

Note, however, that there is an important exception to this pattern of increasing-declining-increasing interdecadal literacy transition ratios across the three census intervals. That is for the people transitioning from ages 10-14 to ages 20-24--who have ratios above one for each of the interdecadal periods. This suggests that for this age range there are increases in literacy as people age. However, this period begins when people are still of school age—with many likely in school--and ends in young

adulthood. Thus, this increase in literacy is likely the result of childhood experience in school rather than increased literacy in adulthood after schooling is completed.

In order to control for substantial differences in census coverage, we calculated another set of interdecadal literacy transition ratios—but this time across twenty-year periods instead of across ten-year periods. That is, we calculated twenty-year interdecadal literacy transition ratios from 1900 to 1920, which were relatively low coverage censuses and from 1910 to 1930, which were relatively high coverage censuses. In calculating these twenty-year transition ratios, we also recognize that coverage may not have been exactly the same in 1900 and 1920 and in 1910 and 1930, but comparing these years should help minimize differences in census coverage.

The twenty-year transition ratios are reported in Table 6. They show that for both two-decade periods those who aged from 10-14 to 30-34 had transition ratios above one. This is consistent with the one-decade transition rates reported earlier being over one for those aging from 10-14 to 20-24 reported in Table 5. It thus again offers support for the idea that literacy continues to increase as children attend school after ages 10-14.

It is important, however, to place the literacy increases from ages 10-14 to 30-34 in larger perspective. We saw in Table 3 that in each of the 1900-1930 censuses literacy was already quite widespread by ages 10-14—being 63 percent in 1900 and 87 percent in 1930. These high percentages literate by ages 10-14 combined with the very small increases in literacy between ages 10-14 and ages 30-34 indicate that most of the American Indians who were to become literate by ages 30-34 had done so by ages 10-14.

Table 6 results also suggest that there is very little, if any, increase in literacy as people age across the adult years. In fact, for both 1900-1920 and 1910-1930, the estimated transition ratios are below one for the youngest adult years. These low transition ratios are particularly compelling as we

expect that any literacy gained in adulthood would be greatest for the youngest adults—a pattern that is contradicted by the data displayed in Table 6.

For the middle and later years of adulthood, the two-decade transition ratios show a fluctuating pattern—ranging from below one to above one. To smooth out these fluctuations we calculated two-decade transition ratios for those ages 20-49 who transitioned to ages 40-69 across the two-decade periods. We did not include those 50 years and older at the initial period to minimize the effects of differential mortality at the older ages. These two-decade transition ratios were 1.02 for the 1900-1920 period and 1.01 for the 1910-1930 period. These summary results indicate that if there was any gaining of literacy for adults during the 1900 to 1930 period, it was very slight. We turn later to additional analysis of this issue.

It should be noted that while the 1900-1930 transition ratios suggest that there is no or very little increase in literacy as cohorts of individuals age across the adult years, we only have evidence of this for the 1900 to 1930 period. This relationship may also hold for the period before 1900, but our evidence does not confirm this possibility, as we have no data about within cohort trends before 1900. However, there are reports from the 19th century of literacy being attained during adulthood[5]. Other data will be needed to test how widespread the acquisition of literacy was during adulthood before 1900.

We mentioned earlier that differential mortality by literacy could affect literacy patterns observed across birth cohorts. Here we add that differential mortality could also influence the results just discussed that compare literacy levels for the same birth cohorts across their life courses. If mortality was higher for the literate, it would bias the interdecadal literacy transition ratios downward.

[5] One example of this appears in an 1830 report from the Superintendent of Indian Affairs (1830, page 168) stating that in one tribe "in addition to the scholars in the regular schools, 250 adults and young people have been taught to read their native language"

Also, if mortality was lower for the literate, it would bias the literacy transition ratios upward. We do not currently understand either the direction or magnitude of such bias, if any.

Relating Birth Cohort Trends in Literacy with Trends in School Attainment.

We now shift our attention to the school attainment data contained in the 1940 census. Unfortunately school attainment was not ascertained in the 1900-1930 censuses and literacy was not ascertained in 1940—making it impossible to see how literacy and school attainment were related at the individual level. However, it is possible to examine their relationship at the birth cohort level by estimating school attainment at the cohort level and comparing that attainment with literacy at the cohort level. Unfortunately, the time differences in coverage of school attainment and literacy make it impossible to make these comparisons across all birth cohorts that we have literacy information for, but we can make such comparisons for a significant number of cohorts.

In Figure 2 we plot the trends in literacy and average number of years of schooling completed by birth cohort. A comparison of the two trend lines indicates a close correlation between average years of schooling and literacy levels. While the different metrics of average years of school attendance and literacy make it impossible to compare levels, we see that the time trends are very parallel.

Figure 3 provides another look at the relationship between school attendance and literacy by plotting both literacy levels and the percentage who had completed one or more years of schooling across birth cohorts. In this figure we see that the trends in literacy and having completed any schooling are not only parallel but very closely overlap in levels. That is, the percentage literate and the percentage having completed any schooling match very closely across the cohorts observed.

Figure 3, thus, adds the additional insight that at least in the reporting of literacy and school attainment in censuses, literacy and any school attainment are very closely connected. This suggests

that at least in census reporting people closely equate any school attendance with literacy. It further suggests that the level of literacy required to report being literate on censuses in this period is quite low—perhaps even as low as the first grade level. This also raises issues of the functionality of "literacy" at this low grade level, as it suggests that a very substantial number of those with one year of schooling but classified as "literate" in the census probably could not read and write basic documents. Of course, many of those classified as literate in the census completed substantially more than one year of schooling and were likely able to read and write basic documents.

Covariates with Literacy Levels and Trends

We now turn our attention from national levels and trends in American Indian literacy to an examination of differentials across gender, geography, integration with the Euro-American community, and having non-Native ancestors. We first examine bivariate associations and later turn to multivariate analyses that take into account the relationships among the socioeconomic factors we analyze.

Levels and Trends by Gender. Table 7 shows that literacy was lower for females in each of the four censuses. For the 1900 data, the percentage literate for women was a full 8 percentage points lower than the percentage for men. Literacy increased across the 1900-1930 period for both genders, but for women more than for men. This means that the gender gap in literacy between women and men narrowed slowly but steadily between 1900 and 1930. It had fallen to about 7 percent in 1910, 5 percent in 1920, and finally to about 3 percent in 1930.

Figure 4 provides a birth cohort perspective on levels and trends by gender. We see there that literacy increased substantially for both women and men across the birth cohorts examined, but the time trends were not the same for the two genders. Whereas the gender differential began relatively small with the earliest birth cohorts (between 3 and 7 percentage points for the first two birth cohorts), it increased fairly consistently to a 14 percentage point gap for the 1856-1860 birth cohort. Thus, during

these years, literacy was increasing at a faster rate among men than among women. However, this trend reversed direction after the birth cohort of 1866-1870 when the gap declined from about 14 percentage points to about 7 percentage points for the birth cohort of 1881-1885, and then to about 1 percentage point by the birth cohort of 1901-1905. By the most recent birth cohorts the differential was zero or had even been reversed. Thus, the last third of the 19th century and the first part of the 20th century was a period of equalization of literacy among American Indian men and women.

Levels and Trends by Region. In Table 8 we document literacy levels and trends across census years by region, and in Figure 5 we show these regional differences by birth cohorts. In this analysis, we use the same nine census regions defined by the U.S. Bureau of the Census. The nine census regions and the states contained in them during the 1900-1930 period are specified below, with the percentage distributions of American Indians by states within regions noted in parentheses[6].

New England: Connecticut (7.2); Maine (45.7); Massachusetts (35.0); New Hampshire (1.4); Rhode Island (9.8); Vermont (0.8).

Middle Atlantic: New Jersey (3.1); New York (80.2); Pennsylvania (16.7).

East North Central: Illinois (1.3); Indiana (1.2); Michigan (38.3); Ohio (1.1); Wisconsin (58.2).

West North Central: Iowa (1.2); Kansas (5.6); Minnesota (21.0); Missouri (0.8); Nebraska (8.3); North Dakota (16.5); South Dakota (46.6).

South Atlantic: Delaware (0.1); Florida (2.7); Georgia (0.6); Maryland (0.2); North Carolina (87.4); South Carolina (3.2) Virginia (5.3); West Virginia (0.2); District of Columbia (0.3).

East South Central: Alabama (20.4); Kentucky (4.3); Mississippi (70.7); Tennessee (4.6).

West South Central: Arkansas (0.3); Louisiana (1.4); Oklahoma (96.8); Texas (1.6).

Mountain: Arizona (41.3); Colorado (1.8); Idaho (4.7); Montana (14.4); Nevada (7.0); New Mexico (25.4); Utah (3.5); Wyoming (2.0).

Pacific: California (54.4); Oregon (15.2); Washington (30.4).

[6] These percentages were calculated from the combined data for American Indians from the complete count data for 1910, 1920, and 1930 and from the weighted oversample for 1900.

Before discussing the differences in literacy across regions, it is important to note that geographical mobility across regional boundaries is a particularly difficult, complex, and relevant topic. The reason is that very large numbers of American Indians were forcefully removed from their homelands and moved to other parts of the country—with many of them residing in what is now known as Oklahoma. Sorting out the forces—including migration and return migration—that produced the regional patterns will be a substantial research project of its own.

As documented in Table 8, the 1900 data show substantial differences across regions, with the lowest percentage literate being in the Mountain region at 19 percent, and the largest being over 70 percent literate in the New England and Middle Atlantic regions. Also relatively high in literacy in 1900 was the West South Central region at 63 percent, while the Pacific and East South Central regions were similar to the Mountain region in being relatively low on literacy (between 26 and 31 percent literate). In 1900, the two North Central regions and the South Atlantic region had middle levels of literacy (around 47-48 percent).

Importantly, the trend towards more literacy observed at the national level holds across regions—as there were increases in literacy levels for all nine regions between 1900 and 1930. Despite the increases being universal across regions, their magnitudes varied by region. The observed increases were particularly large for the East North Central, West North Central, and Pacific regions. Despite the changes across the 1900-1930 period, the differentials in literacy levels remained high—ranging from 45 percent in 1930 for the East South Central region to about 94 percent in the New England region.

Figure 5 displays birth cohort trends for the same nine regions and shows very divergent patterns of literacy trajectories across the regions. Unfortunately, with nine regions (and nine lines) on the same graph, we can only discuss the most significant regional differentials. We begin with the three regions identified earlier as having the highest literacy rates among American Indians in the year 1900:

New England; the Middle Atlantic: and the West South Central. We see in Figure 5 that literacy generally increased in each of these three regions across the entire time period—reaching levels of 95 percent (or nearly 95 percent) by the birth cohorts of 1891-1895 in New England, 1901-1905 in the Middle Atlantic, and 1911-1915 in the West South Central.

Both the East North Central and the West North Central started with lower levels of literacy than those just discussed but increased rapidly across the middle of the 19th century, so rapidly, in fact, that literacy levels in both of these regions surpassed those in the West South Central by the birth cohort of 1881-1885. Literacy in both north central regions also crossed the 95 percent line early in the 20th century. The trajectory for the Pacific Region was very similar to that for the two north central regions, but even more dramatic as the Pacific Region began from a lower level of literacy and equaled the two north central regions by the birth cohort of 1911-1915.

The trajectories for the two southeastern regions—the South Atlantic and East South Central—display substantially different literacy trends than those described above. The trends for the South Atlantic and East South Central regions generally followed those for the regions discussed above through the birth cohorts of about 1866-1870, then substantial differences emerged. For the South Atlantic Region, the upward trend in literacy slowed considerably after the 1866-1870 birth cohort, and the trend for the East South Central Region may have even been downward for a few birth cohorts, before increasing later to earlier levels. These data, thus, suggest that there may have been something associated with the Civil War that put a substantial brake on the expansion of American Indian literacy in these two southeastern regions during subsequent years. More research will be needed to understand this phenomenon.

Literacy in the Mountain Region also followed a different trajectory from the other regions; literacy both started very low in this region, and substantial time passed before substantial literacy

increases emerged. More precisely, except for the earliest birth cohort, literacy was below 5 percent until the birth cohort of 1851-1855, and stayed below 10 percent until the birth cohort of 1866-1870. Following the 1866-1870 birth cohort, literacy increased rapidly in the Mountain Region, eventually surpassing both the East South Central and South Atlantic regions. Because of its low and slow start, literacy levels in the Mountain Region were still just over 70 percent for the most recent birth cohorts.

Rural-Urban Residence. We now shift our geographic lens from region to rural-urban residence. We use the standard 1930-1940 Census Bureau distinction of urban populations being individuals living in cities and incorporated places of 2500 or more people, and for some states located mainly in New England and the Middle Atlantic, townships and other non-incorporated political subdivisions of 10,000 and greater with a population density of 1,000 or more per square mile; rural populations are defined as those residing in less populous communities (IPUMS USA, No Date(c)). We do not break the urban population down further by size of place because in 1900 only 2 percent of the American Indian population was urban and this increased only to 10 percent in 1930.

Table 9 shows substantial differences in literacy between the few urban American Indians observed and the much more numerous rural group. At 78 percent in 1900, urban American Indian literacy nearly doubled the 41 percent literacy of rural Natives. Literacy rates increased rapidly for both groups, but importantly for our current purposes, rural-urban differences remained very high in 1930. While urban literacy reached 95 percent in 1930, it had only risen to 71 percent among the rural residents.

Integration into the Euro-American Community. We now turn to another dimension of American Indian life—connections with the Euro-American population. For this dimension, we take advantage of the census design discussed above for American Indians in 1900 and 1910 that divided the Native population into two groups: those who were more integrated into the Euro-American population (living

in areas outside of high Native concentration) and those who were less integrated into the Euro-American population (living in high Native concentration areas). We compare literacy levels of these two groups.

As shown in Table 10, those individuals who were the most integrated geographically into Euro-American society had substantially higher levels of literacy than those who were less integrated. In fact, the differential was 17 percentage points in 1900 and 21 percentage points in 1910. Unfortunately, we do not know if this differential was the product of the more integrated receiving more school experience and literacy or the attainment of literacy leading to people leaving areas of lower Euro-American presence to areas populated with more Euro-Americans.

White Ancestry. American Indian society has long been characterized by significant numbers of children being born to multi-racial parents (Snipp 1997, 2002; Thornton 1987). By the 1910 census, only 56 percent of those identified as American Indian in that census were reported as definitely having no non-Native ancestry (U.S. Bureau of the Census 1915, page 31). Several scholars have suggested that commitment to establishing schools and school attendance itself was more common among Natives who had multiple ancestries than among those who did not (Berkhofer 1972; Ehle 1988; Littlefield 1989: Mihesuah 1993; Widder 1999). To the extent this was true, it would likely be reflected in higher levels of literacy among multi-ancestry Natives. We now address this question and ask if the census data support the idea that people with more White ancestry had higher levels of literacy than others.

The extent of non-Native ancestry is often referred to as "blood quantum", and the 1900 and 1910 censuses asked those American Indians receiving the special Native supplement to quantify their amount of "white blood." This information was obtained in 1900 by instructing the enumerator to ask: "Has this Indian any white blood; (if so, how much)?" Those with no white blood were given a value of "0" and "if he or she has white blood" the enumerators were told to "write ½, ¼, ⅛ whichever fraction is

nearest the truth." Ancestry information was collected somewhat differently in the 1910 special American Indian supplement by having enumerators fill out in three separate columns the individual's fraction of "American Indian", "White" and "Negro" blood. If the American Indian had no mixed blood, then "full" was to be written in the "American Indian" column and the remaining two columns left blank. Otherwise, all three columns were to be filled out "and the sum, in each case, should equal 1…" Our analyses of blood quantum for 1900 and 1910 uses the special twenty percent sample of American Indians enumerated with both the regular and Native supplement schedules. Since no unique spot on the 1930 census form was set aside for supplemental American Indian questions, enumerators were instructed to use the column used for the non-Native population to record the birthplace of one's father to indicate whether the American Indian was "full blood" or "mixed blood." The IPUMS complete count data set does not contain the information about blood quantum; we therefore rely on the IPUMS sample for data on 1930 blood quantum.

Table 11 documents a substantial relationship between blood quantum and literacy. In 1900, only 28 percent of American Indians with no White ancestry were literate as compared to fully 68 percent of those with 1-25 percent White blood. This roughly means that the difference in literacy between having 0 and 1 White grandparent is a substantial 40 percentage points. Smaller increments of literacy are added with the increase in White ancestry above the 25 percent mark, with the percentage literate reaching over 90 percent for those with 50 percent or more White blood. Large differentials also exist in 1910 and 1930.

Multivariate Logistic Regression Analyses

To obtain further insights into the levels and trends of literacy among American Indians and to provide summary indicators of the effects of the various predictors, we turn to multivariate logistic regression (Powers and Xie 2008; Long and Freese 2014; Muller and MacLehose 2014). In this

multivariate logit regression analysis, we investigate and summarize simultaneously the influence of birth cohort, gender, region, rural or urban residence, census year, integration into the Euro-American community, and White ancestry on literacy. Most importantly, the multivariate regression approach allows for the investigation of the effects of one variable in the analysis while controlling for the effects of the other variables.

This multivariate logistic regression approach requires that we make several transformations of the predictor variables used in the analysis. For gender, we code men with a value of zero and women with a value of one. For region, we create 8 dummy variables--one for each of the regions except the Mountain region; for each of these 8 regional dummy variables we code the people as one if they are in that region and zero otherwise. We treat the Mountain region as the reference region because it has a large population base. We coded residence as a dichotomy with urban coded as one and rural coded as zero. Integration into the Euro-American community was also coded as a dichotomy with the more integrated coded as one and less integrated coded as zero. For blood quantum we coded individuals into four dichotomous variables, as in Table 11, with more than 50% White ancestors as the omitted category.

In our analyses discussed earlier we considered the effects of three time variables on literacy: census year; birth cohort; and age. We would like to study simultaneously the effects of each of these variables on literacy while controlling the effects of the other variables. However, without making substantial assumptions about the effects of the variables, this is impossible because the combination of any two of these three variables produces the third one, thereby preventing estimation. Our initial approach to this problem is to examine only two of these variables and drop the third one, but more sophisticated approaches can be considered.

We have chosen to examine birth cohort and census year and not age because the analysis that we presented in Table 4 suggests that birth cohort is a very significant predictor of literacy. Our earlier analyses presented in Tables 5 and 6 also suggest that there is a definite effect of measurement year, probably due to differential coverage across census years. Our detailed analyses of Tables 5 and Table 6 also suggest that any effect of aging itself is probably very small, with the exception of aging from childhood into adulthood. This latter consideration led us not to include age as a variable directly in the analysis and to drop the age groups 10-14 and 15-19 from the population of analysis to control for pure aging effects. We recognize that to the extent that age does have independent effects, those effects will be represented in the birth cohort and/or census year variables. Nevertheless, our earlier analyses suggest that such distortions are likely to be small.

For census year we created three dummy variables, one each for 1900, 1920, and 1930. For each of these three dummy variables we coded people enumerated in that census year as one and all others as zero. There is no dummy variable for 1910 since 1910 is treated as the census year reference category. The decision of which year category to treat as the reference group is arbitrary; we wanted to use the earliest year, but did not want to use 1900 as the reference group because of the smaller population base produced by using the sample in that year.

As noted earlier in our discussion of Table 4, we constructed the cohort variable by subtracting age at the time of the census from the date of the census. We then divided the individual birth cohorts into five year birth-cohort categories ranging from the earliest—born in 1830 or earlier—to the latest—born between 1906 and 1910. This produced 17 five-year birth cohorts and we created 16 five-year birth cohort dummy variables to indicate all the 17 five-year cohorts except for the most recent cohort, those born between 1906 and 1910. The 1906-1910 birth cohort was treated as the reference cohort because it was the largest.

The logistic regression approach requires that we make a series of transformations on our dependent variable, literacy. We first transformed literacy into a 0-1 variable, with one specifying literate and zero specifying illiterate and subsequently transformed this dichotomous 0-1 variable into an odds ratio by dividing the percentage literate by the percentage illiterate. A third transformation was made by transforming the odds-ratio into a log-odds indicator through a logarithmic transformation. We then predicted the log odds simultaneously by the predictor variables, estimating effect parameters for the categories of the predictor variables while controlling for the other variables. We summarize the results of our logistic regression analysis by transforming the estimated effect parameters for the log odds effects on being literate into predicted probabilities of being literate for a person in each of our defined categories of birth cohort, census year, region, sex, rural-urban residence, integration into the Euro-American community, and ancestry. We did so by using the standardization method that calculates the predicted probabilities for the categories of one variable by holding each of the other variables at its observed value (Long and Freese 2014; Muller and MacLehose 2014). By multiplying, these predicted probabilities by 100, we have predicted percentages or adjusted percentages that can be compared directly to the unadjusted or raw percentages discussed earlier.

We present both the unadjusted and adjusted percentages literate in Table 12 for each of the categories of our seven predictor variables. Note that the unadjusted, or observed, percentage literate reported in Table 12 do not necessarily match those in the previous tables because Table 12 is limited to people ages 20 and older and the previous tables include those 10-19 as well as those 20 and older.

The presentation of results in Table 12 is complicated by the fact that we have to summarize three sets of analyses rather than one. This is necessitated by the fact that we can use the whole population of American Indians from 1900 through 1930 for only the variables of birth cohort, census year, region, rural-urban residence, and sex. The analysis for these variables across the entire period are

shown in the second and third columns of Table 12. The variable of Euro-American community integration was measured only in 1900 and 1910 and we have to rely on samples of the population for these analyses; the primary findings for this variable are displayed in Columns 4 and 5. The detailed variable for blood quantum was only measured in 1900 and 1910 and only for the group of American Indians that were determined by the Census Bureau to be resident in areas with high concentrations of Natives; the results for this variable are also based on a sample of the population and are shown in Columns 6 and 7 of Table 12. Note that we also estimate in Columns 4-7 effects for birth cohort, census year, region, rural-urban residence, and gender because we control the effects of these variables when examining the effects of Euro-American community integration and ancestry. However, we do not discuss these results as they are less generalizable than those in Columns 2 and 3, which will already have been discussed. Also note that some categories of birth cohort and census years are omitted in Columns 4-7 because these analyses are limited only to people participating in the 1900 and 1910 censuses.

We begin our summarization of the effects of our predictor variables estimated with multivariate logistic analysis with birth cohort. As we discussed in regard to Table 4, there are enormous changes in literacy from the earliest birth cohort to the most recent. Whereas only 10 percent of those born by 1830 or earlier were recorded as being literate in Table 12, this number had increased to 82 percent for those born between 1906 and 1910 (unadjusted column).

A comparison of the unadjusted and adjusted percentages literate for the various birth cohorts indicates that the controls for census year, region, rural-urban residence, and sex have only small effects on the results (Columns 2 and 3). Introducing these controls increases the adjusted percentages literate for the earliest cohorts (before 1871-1875) very slightly and decreases them very slightly for the later

cohorts (1871-1875 and after). This means that only a very small portion of the observed cohort trends from 1830 to 1910 was due to the correlation of birth cohort with the other variables.

We now shift to our second time variable, census year, with Table 12 showing the unadjusted and adjusted percentages literate for each of the census years (Columns 2 and 3). Note first that the unadjusted percentages literate in Table 12 are lower than the percentages reported as literate in Table 1. This is because Table 12 excludes the more literate young people—those ages 10-19--whereas Table 1 includes these young people. Nevertheless, the unadjusted column of Table 12 reconfirms the dramatic increases in literacy across the census years between 1900 and 1930.

What we learn new with the adjusted percentages in Table 12 is that the controls for birth cohort, region, rural-urban residence, and sex nearly eliminate the literacy differences across census years. Whereas the unadjusted difference of literacy between 1900 and 1930 was fully 35 percentage points, it was reduced to only 3 percentage points with the multivariate adjustments. This means that some combination of birth cohort, region, rural-urban residence, and sex explain the original census year effect and not something about census year itself. Further investigation revealed that this very large reduction in the census year effect with controls for other variables is almost entirely due to the fact that each successive census drew from increasingly later birth cohorts that have higher literacy. When this high correlation between census year and birth cohort was taken into account in our multivariate analysis, most of the census year effect disappeared. This means that the time trends we observe across census years are largely due to changes across successive birth cohorts rather than changes across census years.

The differences across census years that remain in the adjusted percentages in Table 12 are also consistent with our earlier discussion of the data in Tables 5 and 6 that there are differences across census years that are not due to birth cohort differences. Recall from that earlier discussion that

observed literacy was higher in 1910 and 1930 than in 1900 and 1920, even with birth cohort controlled. Although the adjusted differences across census years reported in Table 12 are quite small, they are also consistent with the earlier results in Table 5 that literacy was somewhat higher in 1910 and 1930 than in 1900 and 1920. In addition, Table 12 suggests that the 1920 adjusted literacy level was only a bit lower than the adjusted 1910 literacy level (51.2 versus 52.6 percent).

In our discussion of Tables 5 and 6, we interpreted the higher reported literacy rates for 1910 and 1930 relative to 1900 and 1920 as probably being due to the greater census coverage in 1910 and 1930 and the greater literacy we hypothesized to be associated with greater census coverage. The data in Table 12 allow us to extend this interpretation to suggest that census coverage and thus literacy levels were lowest in 1900, somewhat higher in 1920, next highest in 1930, and highest in 1910.

We now shift our attention from birth cohort and census year to geographic region. The unadjusted region percentages literate are shown in Column 2 of Table 12, while the adjusted percentages are shown in Column 3. Recall that the unadjusted percentages in Table 12 equal the estimates averaged across the four census years and for the population 20 years of age and older. The unadjusted differences in Table 12, as expected, line up closely with the percentages reported earlier in Table 8, but are not identical because Table 12 only includes people ages 20 and older while Table 8 includes people ages 10 and older.

There are some differences between the unadjusted and adjusted percentages for some regions in Table 2. The two biggest negative adjustments are for the South Atlantic region (from 55.5 to 49.7 percent) and for the West South Central region (from 72.3 to 68.2 percent). The largest positive adjustment is for the Pacific region (from 48.6 to 52.9 percent). These adjustments are primarily due to the population in the South Atlantic and West South Central regions coming from somewhat later birth cohorts, and therefore being more literate, and the Pacific region people coming from earlier birth

cohorts and therefore being less literate. These differences in the distribution of birth cohorts across regions are controlled for or adjusted in the multivariate analysis.

The adjusted results in Table 12 indicate that when birth cohort, census year, rural-urban residence, and sex are simultaneously taken into account, that average literacy levels from the highest to the lowest are: New England; Middle Atlantic; West South Central; East North Central; West North Central; Pacific; South Atlantic; East South Central; and Mountain. Of course, these estimates reflect overall literacy levels across the entire 1900-1930 censuses and do not indicate regional differences at any one time point.

We turn our attention from geographical region to rural-urban residence and see from Table 12 that the multivariate controls significantly reduce the substantial gap between rural and urban American Indians. Whereas, the observed difference in Column 2 is 37 percentage points, the adjusted difference in Column 3 is 20 percentage points, suggesting that a significant part of the rural-urban difference is due to other variables controlled in Table 12. Further examination demonstrated that urban living was highest in the New England and Middle Atlantic regions and lowest in the South Atlantic and Mountain Regions. These rural-urban differences across regions, along with the fact that the New England and Middle Atlantic Regions had particularly high literacy and the Mountain Region had particularly low literacy, helps to account for a substantial amount of the observed rural-urban difference in literacy.

In addition, urban residence is more common among more recent birth cohorts of American Indians. This fact, tied with the observation that more recent cohorts are more literate, helps to explain some of the difference in literacy of the urban compared to the rural.

Although the distribution of rural-urban residence and literacy across birth cohorts and regions can explain much of the observed difference in literacy across rural-urban residence, the adjusted difference after taking these factors into account remains at 20 percentage points. This observation

suggests that there are real distinctions in literacy among rural and urban residents. Either urban living increases literacy or the literate move to urban areas—or both.

We now shift our attention to differences across gender. The unadjusted column numbers in Table 12 indicate literacy levels for the combined 1900-1930 census years and for the population ages 20 and older—and so are somewhat different than the percentages reported in Table 7 that are for ages 10 and above. The unadjusted percentages in Table 12, however, are consistent with those in Table 7 in showing higher literacy for men than for women. However, Table 12 suggests that the difference between men and women is greater (being around 9 percent) than the differences suggested in Table 7 (between 3 and 8 percent). Further investigation indicates that this difference is largely due to the fact that the data in Table 12 are limited to those 20 and older while the data in Table 7 are for those 10 and older. More specifically, while the gender differential for those 20 and older in the 1900-1930 combined analysis is about 9 percent, the difference for those between 10 and 19 is virtually zero. That is, the reported percentage literate for people between the ages of 10 and 19 was 78.5 percent for both women and men. The smaller gender differential in literacy for the young people suggests that there was a decline in the gender differential across birth cohorts—a decline that we discussed earlier with Figure 4.

Table 12 also provides the additional insight that the introduction of multivariate controls does little to change the differences between men and women. More specifically, the difference shrinks only from the unadjusted difference of 9.2 percentage points to the adjusted difference of 8.7 percentage points. Further investigation revealed that women came from somewhat earlier birth cohorts than men, weighting their literacy percentages to be lower. When this birth cohort difference was controlled (or adjusted), the gap between women and men narrowed a bit.

We now turn to the multivariate analysis of Euro-American community integration, noting that here we shift populations from the entire 1900-1930 period to just 1900 and 1910 and use the sample data for both years. In Column 4 of Table 12, we see that the observed difference across community types was 25 percentage points, but that difference was reduced to 14 percentage points in Column 5 with the introduction of the multivariate controls, indicating that a substantial part of the integration difference was due to the other factors included in that equation. ,

More extensive investigation reveals that birth cohort, region, and rural-urban residence are the major factors helping to explain the higher literacy of those more integrated into the Euro-American community. The more integrated people are over-represented in the New England and Middle Atlantic regions where literacy is high and under-represented in the Mountain Region where literacy is low—a combination that helps to make the more integrated into Euro-American life more literate. Similarly, the more integrated are over-represented among the later birth cohorts and the urban residents, helping to account for their higher literacy levels.

Equally important, however, to the fact that birth cohort, region of residence, and rural-urban residence can explain a substantial part of the observed difference in literacy across community integration groups is the fact that a substantial difference in literacy existed across integration types even with these other factors taken into account. This indicates that residence in highly integrated communities leads to more literacy, that more literate folks migrate to such communities, or both.

Finally, we consider having white ancestry (blood). We earlier documented in Table 11 very large differences in literacy for those 10 and above across groups defined by degree of White blood. Table 12 shows that the differences across ancestry groups for those 20 and older are also very large, 65 percentage points from those with the least to the most white blood, but are reduced to 45 percentage points with the full controls (Columns 6 and 7). These findings suggest both that some of the very large

original differences across ancestry groups are due to other factors and that the differences remain very large even when those other factors are controlled. Clearly, having White ancestors is associated with high levels of literacy.

Further analysis demonstrates that White ancestry is highly related to birth cohort. In fact, White ancestry increases monotonically from a low of 7 percent for those born in 1830 or earlier to a high of 38 percent for the birth cohort of 1866-1890. Since literacy increases almost monotonically from 10 to 68 percent across these same cohorts, this confluence of factors helps explain the higher literacy of those with White ancestry. In addition, White ancestry is particularly high in the New England, Middle Atlantic, South Atlantic, East North Central, and West South Central Regions—all of which have generally high rates of literacy. And, White ancestry is particularly low in the East South Central, Mountain, and Pacific regions where literacy is low. This geographical distribution of White ancestry helps explain the difference in literacy across blood quantum groups. But, to reiterate a central conclusion of this part of the analysis: there is a full 45 percentage point gap between those with the least and most White ancestry with these other factors taken into account.

Conclusions and Discussion

We begin our conclusions by reminding readers of the important caveats mentioned earlier about the data. Although there are several data weaknesses that could affect results, we believe that the most important one probably concerns the coverage of American Indians in the decennial censuses. For the reasons we discussed earlier, we believe that the decennial censuses did not include and identify as American Indian all—perhaps even most—of the 1900-1940 descendants of those residing in what is now known as the United States during the year 1491. In addition, the actual coverage and identification of American Indians were likely less complete among those who were more integrated into Euro-American society. And, as we have established in this paper, the more integrated into Euro-American

society—having more urban residence, more living within high Euro-American concentration areas, and having at least one White ancestor are associated with higher levels of literacy. These considerations suggest that the levels of literacy we have documented from the 1900-1930 decennial censuses are probably lower than what would have been recorded if all descendants of the year 1491 residents had been enumerated and recorded as American Indian. Nevertheless, we believe that our estimates are roughly accurate for American Indians relatively closely connected to Native communities.

A second concern is the quality of the data collected for the people recorded in the censuses. As discussed above, there are several areas of concern, as there are in all data collections. Perhaps the largest concern of this type is the measurement of literacy in the censuses. We know that the literacy measure in the 1900-1930 censuses is very crude, with no clear definition of what it means to be able to read and write. In addition, we do not know how enumerators and people being enumerated would have responded to the census question about reading and writing. One partial reassurance in this regard is that there was relative consistency of the question wording across the 1900-1930 decennial censuses.

Our analyses of the linkage of literacy with schooling attainment, however, suggests that the literacy bar in the 1900-1930 censuses was quite low. In fact, we reported a nearly one-to-one association at the birth cohort level between literacy and having been to school for one or more years. The implication of this is that the threshold for reporting literacy in these censuses is only at a first grade level. As we noted earlier, this level of "literacy" was probably sufficiently low to preclude the reading and writing of basic documents for many. At the same time, there were undoubtedly many with higher levels of school attainment with functional literacy skills.

Another caveat concerns the interpretation of our multivariate analyses that allowed us to examine the influence of various factors on literacy while controlling other factors. While we believe that such multivariate analyses sharpen our understanding of the effects of various factors, we make no

claim that we controlled for all confounding influences. In fact, there are undoubtedly other factors that would be useful to control in testing the effects of the factors we investigated. Without those controls, our estimates provide only a beginning point in understanding the influence of various factors on literacy.

Despite these necessary caveats, we believe that we have roughly documented levels and trends of literacy—at least at a relatively rough low level--for the 1900-1930 period and for the birth cohorts from before 1830 to 1920 for a significant portion of the American Indian population. Our analyses have documented the rather substantial incorporation of American Indians into literacy across the years covered in this paper. We have also shown how these reported literacy increases among American Indians occurred primarily through the incorporation of new cohorts of people into literacy—with successive new cohorts becoming more literate. Literacy levels were very low for American Indians born during the first few decades of the 19th century—at around ten percent. Literacy increased during the first few decades of the 19th century—slowly at first but then rapidly for people born during the last part of the 19th century. Finally, by the beginning of the 20th century, literacy among American Indians became nearly universal—reaching 87 percent for those born early in that century. Our results suggest that these trends in literacy were closely related to increases in school levels during the periods observed.

At the same time that our analyses document very large across-cohort differences in American Indian literacy, they suggest very small, if any, increases within cohorts across the adult years. To be sure, our results suggest that literacy increased within birth cohorts after ages 10-14, but such increases are likely due to childhood school experiences after ages 10-14. When we consider changes over time within cohorts of American Indians after ages 20-24, there is very little evidence in these data of increasing literacy during adulthood. Thus, the story of increasing literacy, at least for the 1900-1930

period is one of between cohort changes that distinguish successive cohorts and not change within cohorts during adulthood. As we mentioned earlier, our data do not allow us to investigate whether the lack of within cohort literacy increases after 1930 also existed in earlier years—a topic we leave for future research.

Our analyses also document substantial differences in levels and trends in literacy across regions of the country. Several regions such as New England, the Middle Atlantic, and the West South Central began the early 19th century with higher American Indian literacy levels than existed elsewhere. Literacy in these regions increased rapidly across the subsequent years and reached levels of near-saturation by the late 19th or early 20th century. Although starting at lower levels, literacy in the East North Central, West North Central, and Pacific regions increased rapidly across the 19th century and approached saturation levels by the early 20th century. The Mountain region began the 19th century at low levels of literacy, maintained these low levels for several decades, and only entered a period of rapid increase in the middle of the 19th century. Of particular interest are the South Atlantic and East South Central regions that displayed relatively typical levels and trends until the Civil War birth cohorts, after which the upward literacy trajectory lost much or all of its momentum. This suggests that the Civil War may have had important disruptive effects on American Indian literacy in these regions—effects that continued for many decades.

We alluded earlier in this conclusion that high literacy levels are closely related to urban residence, living in areas of higher Euro-American concentration, and having White ancestors. While some of these effects are due to these attributes being correlated with other important factors related to high literacy, important literacy differences on these dimensions persist even with our full array of control variables. This suggests that being more highly integrated into Euro-American life may lead to

higher levels of literacy—although some of the influence could be in the opposite direction, with the more literate migrating to urban and high Euro-American concentration areas.

We also found lower overall literacy levels among women than among men, although that differential was not constant across the birth cohorts examined. Birth cohorts of women in the early 19th century had somewhat lower levels of literacy than men, and this gap widened during the middle of the 19th century—only to close again during the late 19th century. By the birth cohorts of the early 20th century, the gap had completely closed. We advocate further research to understand and explain these gender trends.

We come to our closing paragraphs with the acknowledgement that our research is a beginning rather than end point in the population-level analysis of levels and trends in American Indian literacy up to 1930. Our research has opened many questions and analyses for both clarifying our results and for new directions. Among the many new analyses that could be conducted include those addressing the previously mentioned issue of differential mortality by literacy levels—and how such differential mortality could affect our observed cohort trends, observed literacy trends within cohorts across the adult life course, and literacy differentials across such factors as gender and region. Such analyses could be facilitated by the newly-released IPUMS Multigenerational Longitudinal Panel data files that link together the lives of individuals across successive early 20th century censuses—allowing the examination of literacy changes within the lives of individuals rather than within birth cohorts (Helgertz et al. 2020a, 2020b; IPUMS USA No Date(d)).

We have documented important levels and trends in American Indian literacy and differentials across various social, economic, and demographic factors, but we have provided few insights into the factors explaining these levels, trends, and differentials in literacy. For example, we have documented differences in levels and trends across region, with particular emphasis on literacy levels in the

southeast after the Civil War. Understanding of these southeastern American Indian trends could possibly be increased by examining similar regional literacy levels and trends for African Americans to see if the trends are similar across groups. The various regional trends might also be illuminated by conducting analyses at the tribal level. Breaking the analyses down by tribe or by tribal group would also likely increase our understanding of other dimensions of American Indian literacy levels and trends.

Finally, we end with a comment concerning whether these trends in literacy are good or bad. The common wisdom in evaluations of literacy is that it is a good thing because it opens up numerous avenues and opportunities in life such as increased connectivity with the larger world and greater economic well-being. At the same time, as we mentioned earlier, we know that among the American Indian population school attendance and the accumulation of literacy was mixed in its motivation, processes, and outcomes. Many American Indians welcomed and sought after Euro-American schools, but in many settings schooling was coercive, harsh, and designed to eliminate Native languages, cultures, and lifestyles. Thus, we make no evaluation of whether the trends documented here were, on balance, good or bad, or worth the costs involved. In any event, we suggest that the documented changes in literacy were of very substantial import to American Indian communities, with very long-lasting ramifications.

References

Adams, David Wallace. 1995. *Education for Extinction: American Indians and the Boarding School Experience, 1875-1928*. Lawrence: University Press of Kansas.

American Anthropological Association. 1998. *AAA Statement on Race*. https://www.americananthro.org/ConnectWithAAA/Content.aspx?ItemNumber=2583. Copied on July 1, 2019.

Baird, Dennis, Diane Mallickan, and William R. Swagerty (2015). *Encounters with the People: Written and Oral Accounts of Nez Perce Life to 1858*. Pullman, Washington, Washington State University Press.

Bennett, Claudette. 2000. "Racial Categories Used in the Decennial Censuses, 1790 to the Present". *Government Information Quarterly*. Volume 17, Number 2, pages 161-180.

Bentley, Michael, Tracy Mattingly, Christine Hough, and Claudette Bennett. 2003. "Census Quality Survey to Evaluate Responses to the Census 2000 Question on Race: An Introduction to the Data". Unpublished paper, U. S. Census Bureau, Census 2000 Evaluation B.3, April 3, 2003

Berkhofer, Robert F. Jr. 1972. *Salvation and the Savage: An Analysis of Protestant Missions and American Indian Response, 1787-1862*. New York: Atheneum.

Blackbird, Andrew J. 1887. *History of the Ottawa and Chippewa Indians of Michigan; A Grammar of Their Language, and Personal and Family History of the Author*. Ypsilanti, Michigan: The Ypsilantian Job Printing House.

Census Office. 1900. Twelfth Census of the United States. Instructions to Enumerators. Washington: Government Printing Office.

Child, Brenda J. 1998. *Boarding School Seasons: American Indian Families 1900-1940*. Lincoln: University of Nebraska Press.

Churchill, Ward. 2004. *Kill the Indian, Save the Man: The Genocidal Impact of American Indian Residential Schools*. San Francisco: City Lights Books.

Cleland, Charles E. 1992. *Rites of Conquest: The History and Culture of Michigan's Native Americans*. Ann Arbor: The University of Michigan Press.

Coleman, Michael C. 1993. American Indian Children at School, 1850-1930. Jackson: University Press of Mississippi.

DeJong, David H. 2007. "'Unless They are Kept Alive: Federal Indian Schools and Student Health, 1978-1918", *The American Indian Quarterly*, Vol. 31, No. 2 (Spring 2007), pages 256-282.

DeWeaver, Norm. 2013. "Who Counts as Indian in the Census; The Multi-Racial Difference". Unpublished paper, April 24, 2013.

Ehle, John. 1988. Trail of Tears: The Rise and Fall of the Cherokee Nation. New York: Random House.

Eschbach, Karl. 1993. “Changing Identification among American Indians and Alaska Natives”, *Demography*, Vol. 30, No.4 (Nov., 1993), pages 635-652.

Eschbach, Karl, Khalil Supple, and C. Matthew Snipp. 1998 “Changes in Racial Identification and the Educational Attainment of American Indians, 1970-1990”, *Demography*, Vol. 35, No. 1 (Feb., 1998), pp. 35-43.

Fear-Segal, Jacqueline. 2007. *White Man’s Club: Schools, Race, and the Struggle of Indian Acculturation*. Lincoln: University of Nebraska Press.

Gram, John R. 2015. *Education at the Edge of Empire: Negotiating Pueblo Identity in New Mexico’s Indian Boarding Schools*. Seattle: University of Washington Press.

Hacker, J. David. 2013. “New Estimates of Census Coverage in the United States, 1850-1930”. Social Science History, 37(1), 71–101.

Hacker, J. David and Michael R. Haines. 2005 “American Indian Mortality in the Late Nineteenth Century: the Impact of Federal Assimilation Policies on a Vulnerable Population” *Annales de Démographie Historique*. 110 (2), pages 17-45.

Hacker, J. David and Michael R. Haines. 2011. “The Construction of Life Tables for the American Indian Population at the Turn of the Twentieth Century”. Pages 73-93 in Per Axelsson and Peter Sköld (editors), *Indigenous Peoples and Demography: The Complex Relation Between Identity and Statistics.* New York: Berghahn.

Helgertz, Jonas, Joseph R. Price, Jacob Wellington, Kelly Thompson, Steven Ruggles, and Catherine R. Fitch. 2020a. “A New Strategy for Linking Historical Censuses: A Case Study for the IPUMS Multigenerational Longitudinal Panel” IPUMS Working Paper No. 2020-03, September 2020. DOI:https://doi.org/10.18128/IPUMS2020-03

Helgertz, Jonas, Steven Ruggles, John Robert Warren, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Joseph P. Price, Evan Roberts, and Matthew Sobek. 2020b. IPUMS Multigenerational Longitudinal Panel: Version 1.0 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D016.V1.0.

Hirschman, Charles. 2004. “The Origins and Demise of the Concept of Race”, *Population and Development Review*. Vol. 30, No. 3. (Sep., 2004), pp. 385-415.

Hoxie, Frederick E. 2001. *Talking Back to Civilization: Indian Voices from the Progressive Era*. Boston: Bedford/St. Martin’s.

Hoy, Benjamin. 2015. “Uncertain Counts: The Struggle to Enumerate First Nations in Canada and the United States, 1870-1911” *Ethnohistory* 62: 4 (October 2015). Pages 729-749.

Huff, Delores J. 1997. *To Live Heroically: Institutional Racism and American Indian Education*. Albany: State University of New York Press.

IPUMS USA. No Date(a). “IPUMS Complete Count Data”.

IPUMS USA. No Date(b). “Descriptions of IPUMS Samples”. https://usa.ipums.org/usa/sampdesc.shtml.

IPUMS USA. No Date(c). https://usa.ipums.or/usa-action/variables/URBAN#description_section.

IPUMS USA. No Date(d). “MLP Linked Data Version 1.0”. https://usa.ipums.org/usa/mlp_data_description.shtml.

Jacobson, Matthew Frye. 2002. “History, Historicity, and the Census Count by Race”. Pages 259-262 in Joel Perlmann and Mary C. Waters (editors). *The New Race Question: How the Census Counts Multiracial Individuals*. New York: Russell Sage Foundation, 2002.

Jobe, Margaret M. 2004. “Native Americans and the U.S. Census: A Brief Historical Survey”. Journal of Government Information, 30(1), 66–80.

Josephy, Alvin M., Jr. 1965. *The Nez Perce Indians and the Opening of the Northwest*. Boston: Houghton Mifflin.

Karamanski, Theodore J. 2012. *Blackbird’s Song: Andrew J. Blackbird and the Odawa People*. East Lansing: Michigan State University Press.

Kesslen, B. 2019. “Native Americans, the census' most undercounted racial group, fight for an accurate 2020 tally”. NBC News, Dec. 29, 2019.

Liebler, Carolyn A. 2018. “Counting America’s First Peoples”. The ANNALS of the American Academy of Political and Social Science, 677(1), 180–190.

Liebler, Carolyn A. and Timothy Ortyl. 2014. “More than One Million New American Indians in 2000: Who are They?” *Demography*. 51: 1101-1130.

Littlefield, Alice. 1989. “The B.I.A Boarding School: Theories of Resistance and Social Reproduction”. *Humanity & Society*, Volume 13, Number 4, November 1989, pp. 428-441.

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

Lyons, Scott Richard. 2010. X-Marks: Native Signatures of Assent. Minneapolis: University of Minnesota Press.

Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. Third Edition, College Station, TX: Stata Press.

Lujan, Carol Chiago. 2014. “American Indians and Alaska Natives Count: The US Census Bureau’s Efforts to Enumerate the Native Population”. *The American Indian Quarterly*, Volume 38, Number 3, Summer 2014, pp. 319-341.

Meister, Cary. W. 1980. “Methods for Evaluating the Accuracy of Ethnohistorical Demographic Data on North American Indians: A Brief Assessment”. Ethnohistory, 27(2), 153–168.Meister. 1980.

Mihesuah, Devon A. 1993. *Cultivating the Rosebuds: The Education of Women at the Cherokee Female Seminary, 1851-1909*. Champaign: University of Illinois Press.

Minnesota Population Center. 2005. "Public Use Microdata Sample of the 1900 U.S. Census of American Indians: User's Guide" Unpublished paper, University of Minnesota, May 2005.

Muller, Clemma J. and Richard F. MacLehose. 2014. "Estimating predicted probabilities from logistic regression: different methods correspond to different target populations". *International Journal of Epidemiology*. Vol. 43, No.3. pp. 962-970.

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

Office of Management and Budget. 1997. "Revisions to the Standards for Classification of Federal Data on Race and Ethnicity". *Federal Register* 62(210), 58782-58790.

Parker, Arthur C. 1915. "The Status and Progress of Indians as Shown by the Thirteenth Census", *Quarterly Journal of the Society of American Indians,* Vol 3, No. 3, July-September, 1915. Pages 185-208.

Parkhurst, Melissa D. 2014. *To Win the Indian Heart: Music at Chemawa Indian School*. Corvallis: Oregon State University Press.

Passel, Jeffrey S. 1976. "Provisional Evaluation of the 1970 Census Count of American Indians", *Demography*, Vol. 13, No. 3 (Aug., 1976), pp. 397-409.

Passel, Jeffrey S. 1996. "The Growing American Indian Population, 1960-1990: Beyond Demography", Pages 79-102 in Gary D. Sandefur, Ronald R. Rindfuss, and Barney Cohen, Editors. *Changing Numbers, Changing Needs: American Indian Demography and Public Health*. Washington, D.C.: National Academy Press.

Powers, Daniel A. and Yu Xie. 2008. *Statistical Methods for Categorical DataAnalysis: 2nd E*dition. United Kingdom: Emerald Group Publishing.

Reyhner, Jon and Jeanne Eder. 2017. *American Indian Education: A History* (Second Edition). Norman: University of Oklahoma Press.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. IPUMS USA: Version 9.0 [dataset]. Minneapolis, MN. IPUMS, 2019. *https://doi.org/10.18128/D010.V9.0*.

Snipp, C. Matthew. 1989. *American Indians: The First of This Land.* New York: Russell Sage Foundation.

Snipp, C. Matthew. 1997. "Some Observations about Racial Boundaries and the Experiences of American Indians". *Ethnic and Racial Studies*, Volume 20, Number 4, October 1997.

Snipp, C. Matthew. 2003. "Racial Measurement in the American Census: Past Practices and Implications for the Future". *Annual Review of Sociology*. 2003, 29: pages 563-588.

Snipp, C. Matthew. 2002. “American Indians: Clues to the Future of Other Racial Groups”. In Joel Perlmann & Mary C. Waters (Eds.), The New Race Question: How the Census Counts Multiracial Individuals (pp. 189–214). New York: Russell Sage Foundation.

Sollors, Werner. 2002. “What Race Are You?”. Pages 263-268 in Joel Perlmann and Mary C. Waters (editors). *The New Race Question: How the Census Counts Multiracial Individuals*. New York: Russell Sage Foundation.

Superintendent of Indian Affairs. 1830. *Report from the Superintendent of Indian Affairs.*

Thornton, Arland, and Linda Young-DeMarco. 2021. “Federal Decennial Census Data for Studying American Indians”, Unpublished paper, University of Michigan.

Thornton, Arland, Linda Young-DeMarco, and Lindsey Willow Smith. 2021. “The Identification of American Indians in the Federal Decennial Censuses”, Unpublished paper, University of Michigan.

Thornton, Russell. 1987. American Indian Holocaust and Survivial: A Population History Since 1492. Norman: University of Oklahoma Press.

Trafzer, Clifford E., Jean A. Keller, and Lorene Sisquoc. 2006. *Boarding School Blues: Revisiting American Indian Educational Experiences*. Lincoln: University of Nebraska Press.

U. S. Bureau of the Census. 1915. Indian Population in the United States and Alaska, 1910. Washington: Government Printing Office.

U.S. Bureau of the Census. 1937. Fifteenth Census of the United States: 1930. The Indian Population of the United States and Alaska. Washington: United States Government Printing Office.

U.S. Bureau of the Census. 1953. *United States Census of Population: 1950, Special Reports, Nonwhite Population by Race,* Washington, D.C. Government Printing Office.

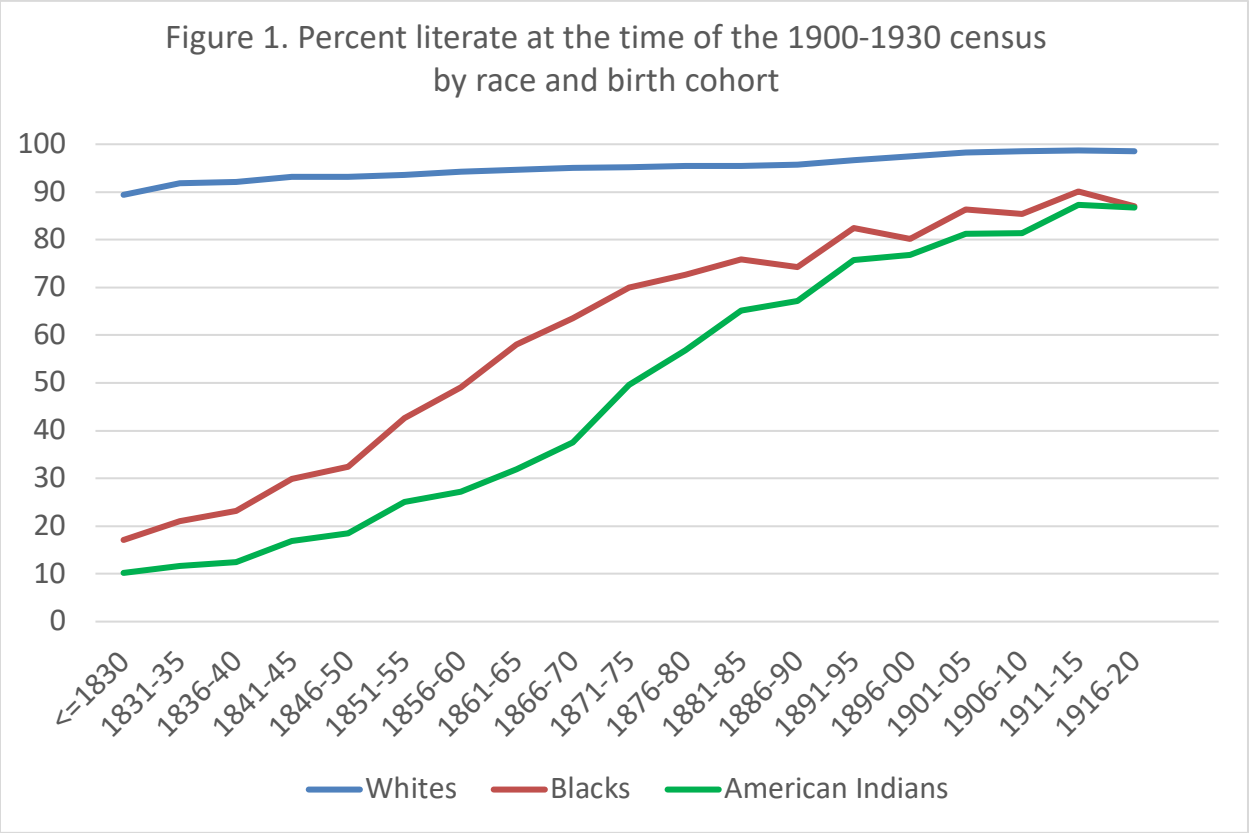
U.S. Bureau of the Census. No Date. *Measuring Race and Ethnicity Across the Decades: 1790–2010 Mapped to 1997 U.S. Office of Management and Budget Classification Standards.* https://www.census.gov/data-tools/demo/race/MREAD_1790_2010.html.

U.S. Bureau of the Census. 2007. “First, Second, and Total Responses to the Ancestry Question by Detailed Ancestry Code: 2000”. Report Number PHC-T-43, Census 2000, special tabulation. Retrieved from https://www.census.gov/data/tables/2000/dec/phc-t-43.html.U.S. Bureau of the Census. 2007.

Vučković, Myriam. 2008. *Voices from Haskell: Indian Students Between Two Worlds, 1884-1928*. Lawrence: University Press of Kansas.

Widder, Keith R. 1999. *Battle for the Soul: Mètis Children Encounter Evangelical Protestants at Mackinaw Mission, 1823-1837*. East Lansing: Michigan State University Press.

Figures and Tables



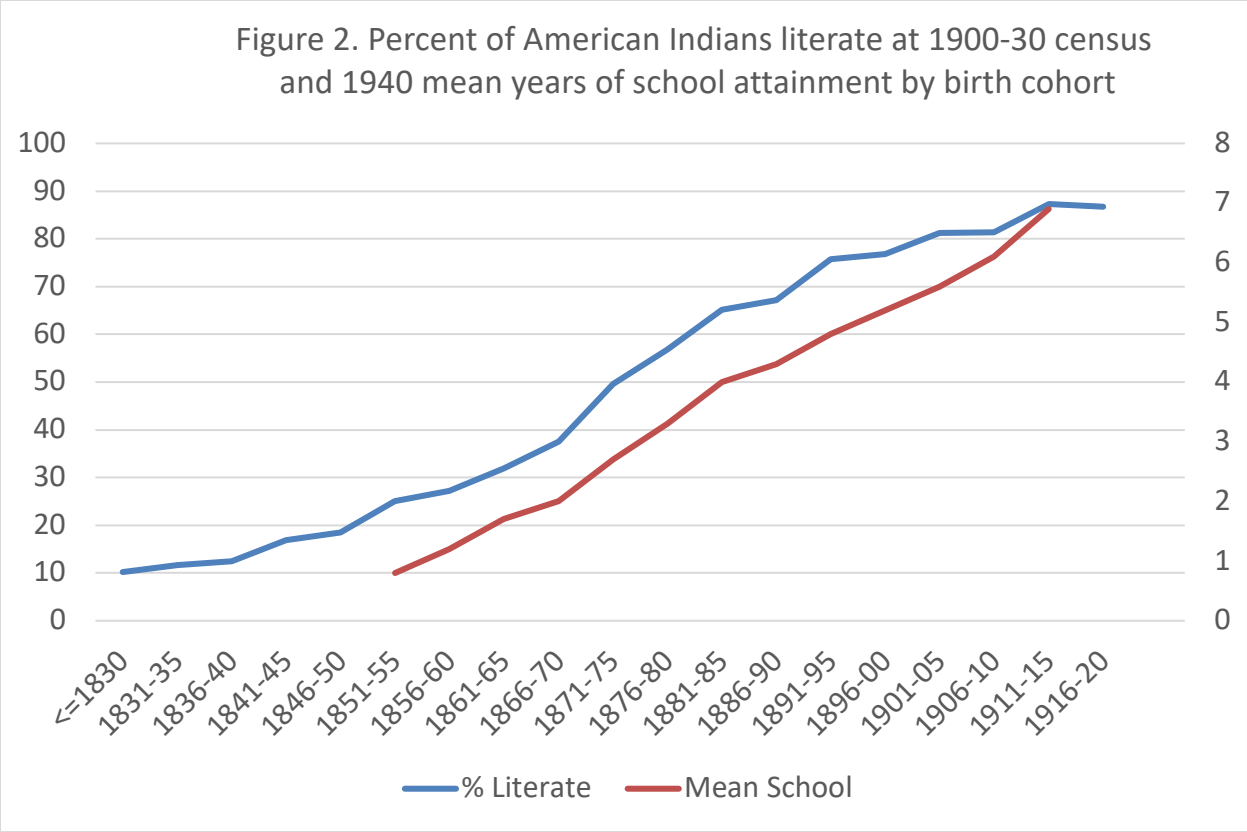


Figure 2. Percent of American Indians literate at 1900-30 census and 1940 mean years of school attainment by birth cohort
100
90
80
70
60
50
40
30
20
10
0
8
7
6
5
4
3
2
1
0
<=1830
1831-35
1836-40
1841-45
1846-50
1851-55
1856-60
1861-65
1866-70
1871-75
1876-80
1881-85
1886-90
1891-95
1896-00
1901-05
1906-10
1911-15
1916-20
% Literate
Mean School

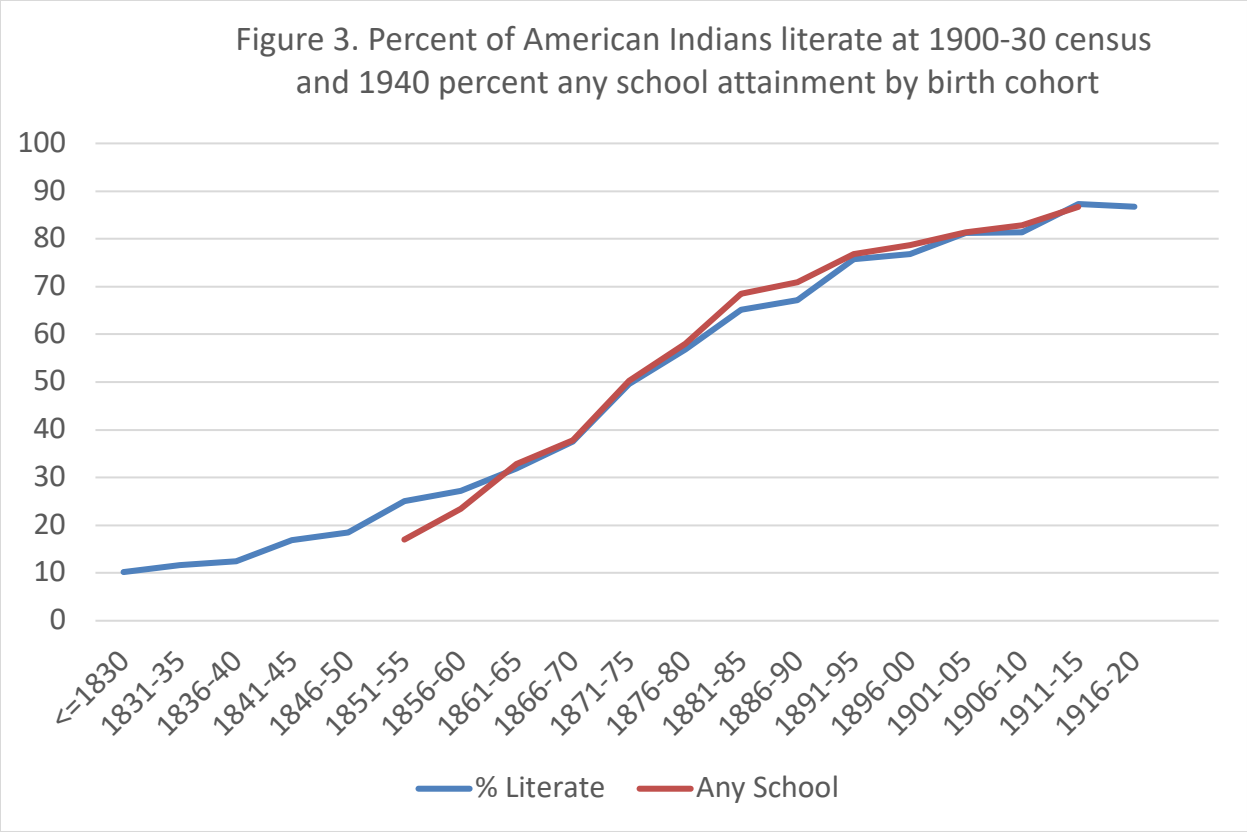


Figure 3. Percent of American Indians literate at 1900-30 census and 1940 percent any school attainment by birth cohort
100
90
80
70
60
50
40
30
20
10
0
<=1830
1831-35
1836-40
1841-45
1846-50
1851-55
1856-60
1861-65
1866-70
1871-75
1876-80
1881-85
1886-90
1891-95
1896-00
1901-05
1906-10
1911-15
1916-20
% Literate
Any School

Figure 4. Percent of American Indians ages 10 and older who are literate at the time of the census, by birth cohort and sex

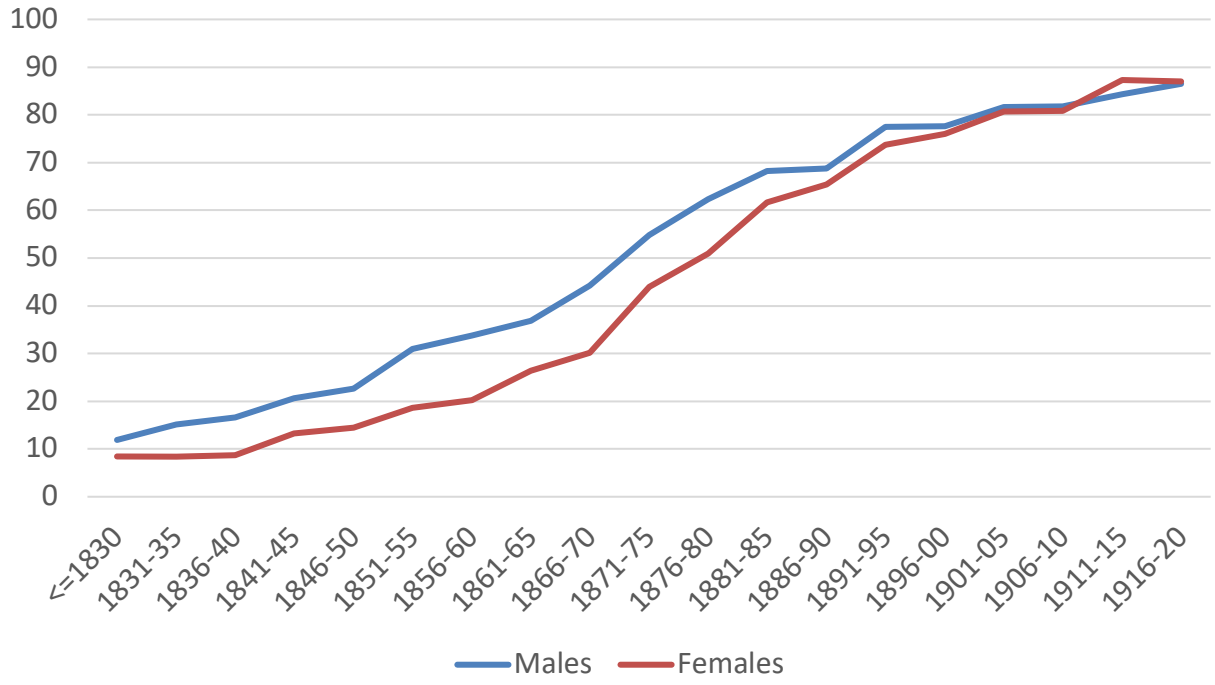


Figure 5. Percent of American Indians ages 10 and older who are literate at the time of the census, by birth cohort and region

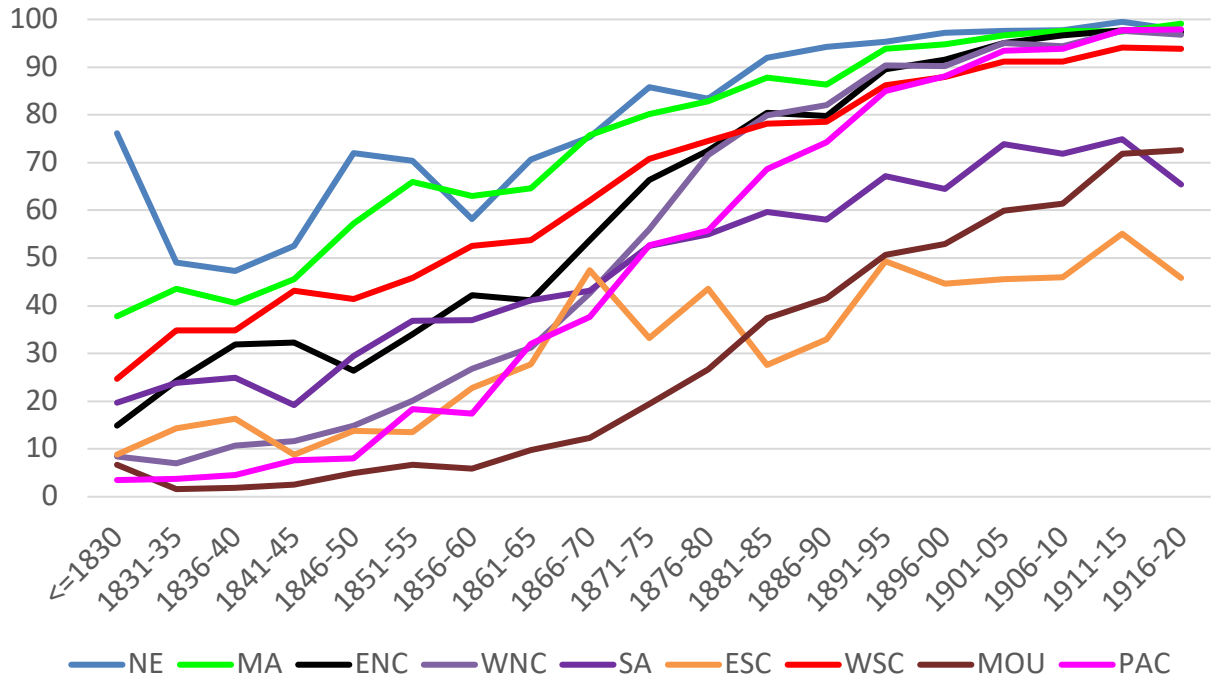


Table 1. Percent of American Indians ages 10 and older who are literate by Census Year and Data Source Type[a]

	1900			1910			1920			1930		
Literacy	Gov. Pub[a]	IPUMS Full	IPUMS Samp[b]	Gov. Pub[a]	IPUMS Full	IPUMS Samp[b]	Gov. Pub[a]	IPUMS Full	IPUMS Samp[b]	Gov. Pub[a]	IPUMS Full	IPUMS Samp[b]
Illiterate	56.2	62.4	58.1	45.3	43.7	42.7	34.9	37.2	36.4	25.7	26.4	25.1
Literate	43.8	37.6	41.9	54.7	56.3	57.3	65.1	62.8	63.6	74.3	73.6	74.9
Total	100	100	100	100	100	100	100	100	100	100	100	100

[a] Government publication percentages for 1900-1930 illiteracy are taken from 1930 Census of Population and Housing, Published Tables, Population, Vol. II, Section 16, Table 4, page 1223. Percent literate is derived by subtracting the percent illiterate from 100.

[b] Percentages shown for IPUMS 1900-1930 American Indian samples are weighted to adjust for differential sampling fractions.

Table 2. Percent of people **ages 10 and older** who are literate according to government published reports by census year and race[a]

		Census Year			
		1900	**1910**	**1920**	**1930**
American Indian	**% Literate**	43.8	54.7	65.1	74.3
	N	171,552	188,758	176,925	238,981
White	**% Literate**	93.8	95.0	96.0	97.3
	N	51,250,918	63,933,870	74,359,749	87,980,667
Black	**% Literate**	55.5	69.6	77.1	83.7
	N	6,415,581	7,317,922	8,053,225	9,292,556

[a] Government publication percentages for 1900-1930 American Indian literacy are derived from 1930 Census of Population and Housing, Published Tables, Population, Vol. II, Section 16, Table 4, page 1223. Percent literate is calculated by subtracting the percent illiterate from 100.

Table 3. Percent of American Indians ages 10 and older who are literate by Census Year and 5-year Age Groups[a]

		Census Year			
		1900	**1910**	**1920**	**1930**
Age Group					
10-14	**% Literate**	63.1	75.4	80.6	86.7
	N	5184	30708	29501	39457
15-19	**% Literate**	67.3	77.8	82.2	87.3
	N	4462	27801	24752	36439
20-24	**% Literate**	57.8	71.3	78.0	82.2
	N	3607	21811	19320	28644
25-29	**% Literate**	48.3	64.6	72.2	80.2
	N	3127	18146	16223	23256
30-34	**% Literate**	34.8	58.1	66.2	78.1
	N	2527	15360	13162	19153
35-39	**% Literate**	26.7	50.5	61.9	75.6
	N	2308	14755	13306	17896
40-44	**% Literate**	25.0	39.1	52.8	69.4
	N	2278	11962	11279	15012
45-49	**% Literate**	24.1	35.8	48.2	65.0
	N	1855	9887	10456	13332
50-54	**% Literate**	16.7	30.9	36.8	57.7
	N	1637	9185	8618	11503
55-59	**% Literate**	15.8	28.8	32.8	52.3
	N	1158	7023	6393	9178
60-64	**% Literate**	11.2	20.4	25.4	40.8
	N	1067	6445	5807	7483
65-69	**% Literate**	11.0	18.5	22.2	35.0
	N	684	4488	4565	5755
70-74	**% Literate**	8.2	14.6	18.9	27.7
	N	577	3343	3359	4173
75-79	**% Literate**	16.9	12.2	16.0	23.8

	N	466	2072	2135	2790
80-84	**% Literate**	11.1	8.1	13.2	19.3
	N	306	1530	1407	1784
85-89	**% Literate**	2.8	8.7	11.6	18.2
	N	160	647	679	914
90-94	**% Literate**	3.8	6.6	7.1	10.4
	N	82	440	393	454
95-99	**% Literate**	2.9	3.7	6.1	15.5
	N	34	164	180	207
100+	**% Literate**	6.2	6.6	26.9	10.3
	N	16	151	376	117

[a] Data for 1900 are from the American Indian sample, with literacy percentages weighted to adjust for differential sampling fractions. The sample Ns are unweighted. Data for 1910-1930 are from the complete-count data files.

Table 4. Percent of American Indians ages 10 and older who are literate at the time of the census, by birth cohort

	1900-1930[a]
Birth Cohort: 1830 or earlier	10.1
N	5639
1831-1835	11.6
N	3642
1836-1840	12.4
N	6271
1841-1845	16.9
N	8695
1846-1850	18.5
N	13225
1851-1855	25.0
N	16233
1856-1860	27.2
N	21443
1861-1865	31.9
N	24343
1866-1870	37.5
N	30590
1871-1875	49.6
N	37516
1876-1880	56.9
N	41749
1881-1885	65.1
N	49246
1886-1890	67.2
N	55169
1891-1895	75.7
N	61920
1896-1900	76.8
N	69181
1901-1905	81.2
N	48008
1906-1910	81.4
N	57749
1911-1915	87.3
N	36439
1916-1920	86.7
N	39457

[a] Data for 1900 are from the American Indian sample, with literacy percentages weighted to adjust for differential sampling fractions. The sample Ns are unweighted. Data for 1910-1930 are from the complete-count data files.

Table 5. Interdecadel Literacy Transition Ratios 1900-1930

	Census Years		
Age Groups	**1900-1910**	**1910-1920**	**1920-1930**
10-14 to 20-24	1.12	1.03	1.02
15-19 to 25-29	0.96	0.93	0.98
20-24 to 30-34	1.00	0.93	1.00
25-29 to 35-39	1.05	0.96	1.05
30-34 to 40-44	1.12	0.91	1.05
35-39 to 45-49	1.34	0.95	1.05
40-44 to 50-54	1.24	0.94	1.09
45-49 to 55-59	1.20	0.92	1.08
50-54 to 60-64	1.22	0.82	1.11
55-59 to 65-69	1.17	0.77	1.07
60-64 to 70-74	1.30	0.93	1.09
65-69 to 75-79	1.12	0.86	1.07
70-74 to 80-84	0.99	0.90	1.02
75-79 to 85-89	0.51	0.95	1.14
80-84 to 90-94	0.59	0.88	0.79
85-89 to 95-99	1.32	0.70	1.34
20-49 to 30-59	1.10	0.93	1.05

Table 6. 20 Year Interdecadel Literacy Transition Ratios, 1900-1930

	Census Years	
Age Groups	**1900-1920**	**1910-1930**
10-14 to 30-34	1.05	1.03
15-19 to 35-39	0.92	0.97
20-24 to 40-44	0.91	0.97
25-29 to 45-49	0.99	1.01
30-34 to 50-54	1.06	0.99
35-39 to 55-59	1.23	1.04
40-44 to 60-64	1.02	1.04
45-49 to 65-69	0.92	0.98
50-54 to 70-74	1.13	0.90
55-59 to 75-79	1.01	0.83
60-64 to 80-84	1.18	0.95
65-69 to 85-89	1.05	0.98
70-74 to 90-94	0.87	0.71
75-79 to 95-99	0.36	1.27
20-49 to 40-69	1.02	1.01

Table 7. Percent of American Indians ages 10 and older who are literate by Census Year and Sex[a]

		Census Year			
Sex		**1900**	**1910**	**1920**	**1930**
Males	**% Literate**	46.0	59.8	65.3	75.2
	N	18637	95415	88562	122842
Females	**% Literate**	37.8	52.6	60.1	71.8
	N	12898	90503	82953	114705

[a] Data for 1900 are from the American Indian sample, with literacy percentages weighted to adjust for differential sampling fractions. The sample Ns are unweighted. Data for 1910-1930 are from the complete-count data files.

Table 8. Percent of American Indians ages 10 and older who are literate by Census Year and Region[a]

		Census Year			
		1900	**1910**	**1920**	**1930**
Region					
New England	**% Literate**	72.8	82.6	87.7	93.6
	N	148	1462	1291	1867
Middle Atlantic	**% Literate**	74.2	86.8	84.2	88.5
	N	983	5859	4192	5639
East North Central	**% Literate**	48.0	67.5	78.6	86.7
	N	1819	13089	10804	14226
West North Central	**% Literate**	48.4	59.6	72.8	83.4
	N	5987	30217	26693	34493
South Atlantic	**% Literate**	47.2	53.8	63.2	64.4
	N	755	6813	9019	12905
East South Central	**% Literate**	26.4	39.9	40.4	45.3
	N	303	1630	1245	1438
West South Central	**% Literate**	62.9	75.7	79.3	87.4
	N	8121	49798	39063	66812
Mountain	**% Literate**	19.0	30.6	40.2	51.1
	N	9367	52930	55885	72605
Pacific	**% Literate**	30.7	54.9	66.0	81.6
	N	4052	24120	23323	27562

[a] Data for 1900 are from the American Indian sample, with literacy percentages weighted to adjust for differential sampling fractions. The sample Ns are unweighted. Data for 1910-1930 are from the complete-count data files.

Table 9. Percent of Americans Indians **ages 10 and older** who are literate by Census Year[a] and Rural/Urban status

		Census Year			
		1900	**1910**	**1920**	**1930**
Urban	**% Literate**	77.9	84.9	88.4	95.2
	N	669	8217	11291	25159
Rural	**% Literate**	40.8	55.0	61.0	71.0
	N	30866	177701	160224	212388

[a] Data for 1910-1930 are from IPUMS complete count census file years. 1900 data are from IPUMS 1% sample with oversamples. Percentages for that year are weighted and Ns are unweighted.

Table 10. Percent of American Indians **ages 10 and older** who are literate by Census Year[a] and Integration into Euro-American Society

		Census Year	
		1900	**1910**
More Integrated Euro-American Society	**% Literate**	57.6	76.0
	N	111	147
Less Integrated Euro-American Society	**% Literate**	40.8	55.0
	N	31424	33562

[a] Percents for 1900-1910 IPUMS American Indian oversample data are calculated from two of the four sub-samples included within that data set. The sub-samples included were the 1% National sample (all records within this sub-sample were administered the general population schedule only), and the American Indian oversample (all records in this sub-sample were administered both the general and the American Indian schedules). It is unknown whether one or both schedules were completed for individuals included in the remaining two sub-samples (Preston and Hispanic); these two-subsamples are excluded from this analysis.

Table 11. Percent of American Indians **ages 10 and older** who are literate by Census Year and Percentage of White Blood

		Census Year			
% White Blood		**1900**	**1910**		**1930**
None	**% Literate**	27.8	40.4	**None**	60.5
	N	21968	22133	**N**	5568
>0 and <=25%	**% Literate**	68.1	72.7	**Some**	91.0
	N	2126	2035	**N**	4542
>25 and <=50%	**% Literate**	73.3	76.5		
	N	3490	3693		
>50%	**% Literate**	91.4	93.4		
	N	1802	4995		

[b] Percents for 1900 and 1910 are calculated from the 1900-1910 IPUMS American Indian oversample data (all records in this sub-sample were administered both the general and the American Indian schedules). Percents for 1930 are calculated from the IPUMS 1930 5% sample data. Percentages are weighted and Ns are unweighted.

Table 12. Percentage Literate for American Indians **ages 20 and older** at the time of the Census, Unadjusted and adjusted through logistic regression

	Complete Count Data[a]		American Indian Oversample[b]		American Indian Oversample[c]	
	Unadjusted	Adjusted[d]	Unadjusted	Adjusted[d]	Unadjusted	Adjusted[d]
Birth Cohort: 1830 or earlier	10.2	13.3	10.1	12.9	9.3	17.1
1831-1835	11.6	14.2	12.7	15.1	11.8	16.4
1836-1840	12.4	16.0	12.4	15.7	11.4	16.3
1841-1845	16.9	18.9	16.8	18.4	16.7	18.9
1846-1850	18.5	20.7	17.6	19.4	17.0	19.3
1851-1855	25.0	25.8	25.6	25.4	22.8	24.4
1856-1860	27.2	28.4	27.0	27.2	25.5	26.4
1861-1865	31.9	33.2	29.8	30.6	29.4	30.5
1866-1870	37.5	40.0	37.1	39.5	36.4	36.9
1871-1875	49.6	49.8	49.3	49.2	47.4	45.2
1876-1880	56.9	56.4	57.9	56.0	55.9	52.3
1881-1885	63.9	62.2	65.0	61.0	63.9	57.0
1886-1890	69.4	67.5	71.0	65.6	70.4	63.0
1891-1895	74.0	73.2	--	--	--	--
1896-1900	78.0	77.6	--	--	--	--
1901-1905	80.2	79.0	--	--	--	--
1906-1910	82.2	81.2	--	--	--	--
Census Year: 1900	31.9	49.4	31.9	37.6	30.4	36.4
1910	47.0	52.6	46.6	41.2	45.1	38.3
1920	54.3	51.2	--	--	--	--
1930	67.2	52.3	--	--	--	--
Region: New England	81.4	78.8	66.7	66.7	59.0	51.8
Mid Atlantic	78.5	75.7	72.9	66.6	66.9	62.0
East North Central	64.1	65.2	49.7	50.2	48.6	45.2
West North Central	56.1	59.3	41.8	44.3	40.8	42.3
South Atlantic	55.5	49.7	47.4	42.3	45.9	34.9
East South Central	36.3	37.0	26.9	22.7	20.0	23.1
West South Central	72.3	68.2	63.6	59.4	63.1	50.4
Mountain	27.8	28.8	15.3	16.8	14.7	21.9
Pacific	48.6	52.9	31.2	33.6	29.0	35.1
Residential Location: Rural	49.2	50.6	38.3	39.1	37.3	37.3
Urban	86.6	70.5	76.3	53.8	74.0	47.2
Sex: Female	46.8	47.0	33.7	34.2	32.6	32.8
Male	56.0	55.7	45.1	44.6	43.4	42.0
Euro-American Integration: More	--	--	62.9	52.8	--	--
Less	--	--	38.0	38.7	--	--
% White Blood: None					25.1	29.6
>0 and <=25%					63.6	49.7
>25 and <=50%					68.0	57.2
>50%					90.6	74.4

[a] 1900 IPUMS weighted American Indian oversample data and 1910-1930 IPUMS 100% Data.

[b] Percents for 1900-1910 IPUMS American Indian oversample data are calculated from two of the four sub-samples included within that data set. The sub-samples included were the 1% National sample (all records within this sub-sample were administered the general population schedule only), and the American Indian oversample (all records in this sub-sample were administered both the general and the American Indian schedules). It is unknown whether one or both schedules were completed for individuals included in the remaining two sub-samples (Preston and Hispanic); these two-subsamples are excluded from this analysis.

[c] Percents for 1900-1910 IPUMS American Indian oversample data are calculated from subsample of all records for individuals that were administered both the general and the American Indian schedules.

[d] Adjusted percentages are calculated as predicted probabilities (times 100) in a logistic regression equation with all the predictor variables listed in the equation.