
RESEARCH REPORTS AND NOTES

SIGNED WITH AN X:

Methodology and Data Sources for Analyzing the Evolution of Literacy in Latin America and the Caribbean, 1900–1950

Javier Núñez

Universidad de Chile

Received 2-8-2004; Revise and Resubmit 4-22-2004

Received Revised 9-10-2004; Final Acceptance 9-21-2004

Abstract: This article studies the evolution of literacy in Latin America and the Caribbean from 1900 to 1950. A methodology is developed to overcome the lack of census data for half of the countries in the region for 1900, as well as the lack of comparability of the existing census data. Combining census data and literacy data gathered from marriage registrations, military recruits, crime statistics, and urban censuses, adult literacy estimates for twenty-two countries of the region are provided for 1900, which offer a new and more complete portrait of human capital formation from 1900 to 1950. There are wide variations across the region in literacy rates in 1900, as well as in the increase of literacy from 1900 to 1950, the latter being associated with variations in the expansion of primary education enrollment in different Latin American countries. However, countries also differ in their success in transforming school enrollment into adult literacy, which is partly associated with the prevalence of Amerindian populations.

INTRODUCTION: THE PROBLEMS OF LIMITED DATA PRE-1950

One of the most significant changes that occurred in Latin America during the first half of the twentieth century is the state-led expansion of elementary education. This transformation—especially the rise and

Latin American Research Review, Vol. 40, No. 2, June 2005

© 2005 by the University of Texas Press, P.O. Box 7819, Austin, TX 78713-7819

expansion of the *Estado Docente* in Spanish America—has been studied by a number of authors.¹ Yet, there is limited and debatable knowledge about the real impact that the expansion of elementary education had on the formation of human capital across the region. Measuring and studying the formation of human capital is crucial to better understanding the process of economic, social, and political modernization in Latin America. This work attempts to contribute to the study of human capital formation in the region in the first half of the twentieth century by focusing on the evolution of literacy, which is regarded as a fundamental constituent of human capital.²

The rigorous study of Latin American literacy in past times in the region has been limited by the unavailability of reliable and comparable data. The year 1950 can be regarded as the earliest date for which comparable literacy figures for all countries of the region are available. Encouraged by the newly created United Nations, in 1950 most countries of the region undertook national censuses in coordination, following a fairly homogeneous methodology, while the remaining countries undertook national censuses within the next few years.³ However, prior to 1950, census literacy data are available only for some countries of the region, and yet they are not directly comparable across countries due to differences in the definitions of literacy,⁴ census methodologies, and age groups for which literacy rates are reported.⁵ Comparability is further undermined because most censuses were taken in different years. The studies that report literacy rates prior to 1950, such as Newland (1994) and the recently launched Oxford Latin American History Database,⁶ although very valuable, suffer from these limitations, as they only draw on the available census data. More data and better estimates of literacy are needed to understand the process of human capital formation in the region prior to 1950.

Employing a new data set, this research note provides a comparative study of the evolution and determinants of literacy in Latin America and the Caribbean during the first half of the twentieth century. A methodology is developed in order to produce new estimates of literacy for twenty-two countries of the region for 1900, including eleven countries

1. Newland (1991, 1994) and Mariscal and Sokoloff (2000) provide good accounts of the expansion of elementary education during this period.

2. On the various theoretical and empirical linkages between literacy, human capital, labor productivity, and economic and human development, see Barro (1991), Dasgupta (1993), Hicks and Streeten (1979), Ranis, Stewart, and Ramirez (2000), and Sen (1983).

3. See table 6 below.

4. The most common literacy definitions are “ability to read only” versus “ability to read and write.” As we shall see, in this paper we employ only those censuses that employ the “read and write” criterion, which are reported in table 4.

5. See UNESCO (1953).

6. Available at <http://oxlad.qeh.ox.ac.uk/references.php>.

that lack census data circa 1900 (see table 4). The key aspect of the methodology is that it combines the existing census-based literacy rates with new literacy data coming from four alternative literacy sources, military recruits, crime statistics, and urban censuses, all of which have been gathered from a variety of primary and secondary sources.⁷ Using multivariate regression analysis, we show that these alternative literacy sources are remarkably good predictors of national literacy rates. Accordingly, these alternative literacy sources are employed to estimate the national literacy rate circa 1900. These estimates are further tested for consistency with cross-sectional information on primary education enrollment circa 1900, as suggested by theory and common sense.

Using the new 1900 estimates, this article provides a new and more complete portrait of the formation of human capital in the region from 1900 to 1950 than has been reported previously in the literature. The new data show high dispersion in literacy rates at the turn of the twentieth century, as well as a high degree of variation in the expansion of adult literacy across countries between 1900 and 1950, ranging from as low as 10 to more than 50 percentage points, an increase that is unrelated to the countries' starting literacy rates in 1900. These data show that this phenomenon reflects the variations in the expansion of elementary education enrollment in the region. However, there is also a great deal of variation across countries in the degree to which Latin American countries transformed elementary education enrollment into adult literacy. This suggests that the countries' ethnic profiles, measured as the proportion of Amerindian versus non-Amerindian population, stand out as key factors that shaped the heterogeneous effectiveness of the increase of primary enrollment in expanding the region's literacy in this period.

The rest of the research note is organized as follows: The second section details the methodology adopted and the literacy sources employed. The third section provides the literacy estimates for 1900. The fourth section examines the evolution of literacy and its determinants in the region between 1900 and 1950. Finally, the fifth section offers our conclusions.

ESTIMATING COMPARABLE NATIONAL LITERACY RATES FOR 1900

Methodology

This section describes the methodology developed in this article to obtain comparable literacy estimates for 1900. As explained earlier, this

7. The literacy data employed in this article as well as the references of the corresponding primary and secondary sources consulted have been incorporated into a statistical appendix, which is available from the author's website at <http://www.facea.uchile.cl/FrameArea.asp?cod=78>.

methodology attempts to overcome two main problems, namely i) the lack of census data for some countries around 1900, and ii) the lack of direct comparability of the available census data across countries around 1900.

Suppose that reliable census literacy data existed for a number of countries at a given time. Suppose also that there were literacy data for a specific population segment of these countries for a given year, for example middle-class, urban males. It is clear that the latter data would certainly provide a biased estimate of the "true" national literacy rate, as literacy is known to vary between gender, between urban and rural areas, as well as across socioeconomic groups. In particular, the raw literacy rate of this specific population segment is likely to be significantly higher than the national literacy rate. Yet, it is plausible that these data, although certainly not representative of the whole population, could be successfully employed to *predict* national literacy by means of a suitable statistical procedure, namely regression analysis. This procedure would require *variations* in national literacy to be associated with *variations* in the literacy of a given population segment across countries.⁸ For example, gender-specific literacy rates, or urban literacy rates are likely to be higher in a literate country than in a largely illiterate one. Assuming that the literacy rate of some specific population segment can provide a good prediction of the national literacy rate, the former can be employed to estimate the national literacy rate of those countries that lack census literacy statistics around 1900.

This methodology can also help tackle the second problem presented above, namely the lack of comparability of the available census data.⁹ To illustrate, suppose that a census literacy rate grossly overestimates the country's "true" literacy rate, due for example, to a particularly lax definition of literacy, a data manipulation error, or a faulty census implementation, such as underreporting of poor or rural areas. In this case, using an alternative national literacy estimate obtained from the procedure outlined above can help detect "outlier" or distorted census literacy figures.

The methodology described above rests on the assumption that the literacy rate of a specific population segment can indeed provide a good prediction of the national literacy rate. This is ultimately an empirical matter that must certainly be established robustly before estimates can be computed. This matter is discussed below.

8. Which can be related, for example, to variations in the degree of development of elementary education across countries, among other factors.

9. This may happen due to differences in the definitions of literacy employed in the censuses, differences in census implementation, or even data manipulation errors.

Sources

Census Data As argued above, inspection of the methodologies and implementation procedures of the existing censuses circa 1900 suggests that they are too diverse to safely assume direct comparability of their literacy rates. In particular, the most significant obstacles to tackle are i) the use of different definitions of literacy, and ii) differences in age tabulation of the reported literacy data. To overcome the first problem, this research note employs only literacy data from censuses that define literacy as the percentage of a population "able to read and write."¹⁰ Regarding age tabulations, censuses prior to 1950 do not report literacy for a common age threshold. However, some censuses report data for various age thresholds simultaneously.¹¹ From these censuses we computed the average ratio between literacy rates for ages 15+, and literacy rates for ages 0+, 6+, and 10+. All census literacy rates reported for all ages 0+ and 6+ were multiplied by the resulting average ratios 1.24 and 1.07 respectively in order to obtain comparable adult literacy rates for ages 15+, which is the definition of adult literacy most widely employed nowadays.¹² Table 4, below, presents the census literacy figures derived from this procedure. (Note that these data differ somewhat from the raw census literacy data employed in the related literature.)

Literacy Data for Specific Population Segments In order to implement the methodology outlined above, literacy data for specific population segments were obtained for most of the countries in Latin America and the Caribbean. These alternative literacy sources are the following: i) marriage registries, ii) military recruiting data, iii) crime statistics, and iv) urban censuses. All of the literacy information gathered as part of this investigation, as well as their corresponding sources have been incorporated into a statistical appendix, which has been made publicly available.¹³ It must be noted that, according to the sources, the literacy criteria

10. This is the definition most widely used at that time and nowadays, and it is presumably the most credible, as well as the most comprehensive and relevant from a human capital perspective.

11. The most common thresholds used are 0, 6, 10 and 15 years of age. The censuses are: Caracas, 1891; Montevideo, 1900; Brazil, 1900, 1920, 1940, 1950; Chile 1952; Cuba 1899; Trinidad & Tobago, 1946; Uruguay, 1900; and Argentina, 1914. In what follows, the + sign will be employed to mean "above" a certain age.

12. The standard deviations for these average factors are very low; 0.0016 ($n=8$) and 0.0009 ($n=5$), respectively. The difference between literacy for ages 10+ and 15+ turned out to be negligible.

13. See footnote 7. The sources are located in various libraries, including the British Library in London, The Bodleian Library in Oxford, The University Library at Cambridge, UK, the Library at the London School of Economics in London, the Library of Congress in Washington, DC and the national libraries of Argentina, Chile, and Perú.

employed in each alternative measure of literacy were similar across countries. Illiteracy in marriage statistics was assessed as the number of brides and bridegrooms who signed the marriage register with an X. Military recruits' reading and writing skills were assessed in order to identify those eligible for extra instruction courses. Finally, literacy of criminal offenders was, with rare exceptions, defined explicitly as "able to read and write" in the sources employed. Moreover, each population segment shares similar features across countries. Brides and bridegrooms in the registries belong to a similar age group and are likely to overrepresent urban areas. Military recruits are mostly young men around eighteen to twenty years of age, and criminal offenders are mainly young males. In any case, if there were large differences in either the data collection methods or the characteristics of the population segments across countries, this should be reflected in a poor or nonexistent statistical association between national literacy rates and the literacy rates from these alternative sources. The evidence, however, does suggest that the literacy rates of different population segments are indeed highly correlated. Table 1 reports correlation coefficients among literacy rates of different sources, including national census data.

The data employed in table 1 correspond to pairs of literacy rates from different sources but from one country in a common (or neighboring) year within 1878–1960, as reported in Part I of the statistical appendix.¹⁴ All correlation coefficients are remarkably close to unity, and statistically significant at more than 99 percent confidence. This is clear evidence of a very high degree of association among all the alternative literacy sources, despite the fact that they are not a representative sample of the corresponding national adult population. It is also noteworthy that these alternative literacy measures are highly correlated with the census literacy data. These results suggest that each of these alternative literacy sources possesses a great deal of information that can be used to estimate national adult literacy rates. For this purpose, adult census literacy rates were regressed against all three literacy sources employing the same data used for the correlation analysis reported in table 1. The variable "year" was introduced to control for any possible effect related to the time in which each observation was measured, which turned out not to be statistically significant in all three models, as shown in table 2. The evidence from table 2 implies that the statistically significant coefficients of the regressions can be employed to obtain literacy estimates for 1900.

14. Although our goal is to obtain literacy estimates for 1900, we employ data from this wider period in order to employ more observations. However, as shown later in table 2, the statistical associations between the alternative literacy sources are stable throughout this period and therefore independent of time.

Table 1 Correlation Coefficients among Alternative Literacy Sources: Latin America, Selected Years 1890–1950

	Convicts and Criminal Offenders	Military Recruits	Brides and Bridegrooms	Brides Only	Bridegrooms Only
Census 15 +	0.97 (22)	0.98 (12)	0.94 (19)	0.94 (15)	0.92 (15)
Convicts and Criminal Offenders		0.93 (6)	0.95 (11)	0.93 (8)	0.96 (8)
Military Recruits			n. a.	n. a.	n. a.
Brides and Bridegrooms				0.99 (18)	0.99 (18)
Brides Only					0.99 (18)

Note: Number of observations in parenthesis.

All coefficients are significant at 99 percent confidence.

Table 2 Regressions with Robust Standard Errors, Dependent Variable: Census Adult Literacy Data Latin America, Selected Years 1878–1960

Variable	1	2	3	4	5	6
Convicts and Offenders	1.296* (21.81)	1.244* (20.71)	—	—	—	—
Marriage Registries	—	—	0.713* (10.52)	0.733* (11.20)	—	—
Military Recruits	—	—	—	—	0.993* (15.30)	1.007* (14.95)
Year	-0.093 (-2.02)	—	0.081 (1.24)	—	0.069 (1.14)	—
Intercept	158.2 (1.79)	-18.11* (-5.99)	-150.64 (-1.19)	3.042 (0.69)	-132.56 (-1.12)	-1.893 (-0.43)
Number of observations	22	22	19	19	12	12
R-squared ^a	0.94	0.93	0.90	0.89	0.96	0.96

Note: *t*-statistics are given in parenthesis.

^a The robust errors estimation method employed does not report adjusted *R*-squared.

* Significant at 99 percent confidence.

The coefficients of all three alternative literacy sources are positive and stable across different specifications. Moreover, these coefficients have remarkably high *t*-values systematically over 10, being significant at a more than 99 percent confidence level in all specifications. The remarkably high *R*² of all regressions (a measure of goodness-of-fit) indicates that each alternative literacy figure is individually able to “explain” about 90 percent or more of the variation in national adult literacy across countries. Alternative functional specifications were tested, which did not improve the performance of the linear specification reported in table 2.

The regressions in table 2 have robust standard errors, as do the remaining regressions reported in this work.¹⁵

Literacy Estimates from Urban Census Data for 1900 Several countries of the region undertook urban censuses around 1900, typically in their respective main capital or principal cities. These literacy data have often been neglected or underemployed in previous studies that assess the evolution of human capital in the region. All the urban census data gathered and their respective sources are included in Part II of the statistical appendix. Note from the appendix that, not surprisingly, in all cases urban literacy is higher than the corresponding national literacy. However, as with the previous three alternative literacy sources discussed earlier, urban literacy can also be employed to predict national literacy by means of regression analysis. In particular, urban literacy is expected to be associated with national literacy for two distinct reasons. First, the main capital city contains a significant share of a country's population. Second, urban literacy is likely to be higher in a literate country than in an illiterate one. Hence, using regression analysis, urban literacy can be used to predict national literacy in a manner similar to the previous section, despite the fact that urban literacy is certainly higher than national literacy. For this purpose, national adult literacy rates were regressed against urban (main capital city) census literacy rates using all cases where data for urban and national literacy were available for a country in a common year (see Part II of the statistical appendix). The regressions also included other variables that were expected to affect national literacy depending upon a given level of urban literacy, namely the percentage of Amerindian population, the share of the national population living in the main capital city, and a dummy variable intended to capture the geographical extent of the urban area for which the urban literacy estimate was provided.¹⁶ Table 3 presents the results of alternative regression specifications.

The urban literacy coefficient is positive, stable, and statistically significant in all specifications, suggesting a solid basis for employing urban literacy data to estimate national literacy rates. The coefficients of the main capital city's share of the national population are also systematically significant and stable, suggesting that, all other factors remaining constant, countries with a larger share percentage of urban populations have (on average) higher literacy rates. On the other hand, the dummy variable for greater urban area and the percentage of

15. Which corrects any potential heteroskedasticity in the data.

16. Given that literacy is expected to be lower at the outskirts of the cities than in the center, national literacy is expected to be lower if a given urban literacy figure is obtained from a low urban coverage census than a census with wider urban coverage.

Table 3 Regressions with Robust Standard Errors, Dependent Variable: Census Adult Literacy, Latin America, Selected Years 1876–1926

Variable	1	2	3	4	5
Urban Adult Literacy (Main City)	0.209* (3.10)	0.222* (3.50)	0.413* (3.42)	0.282* (3.21)	0.310* (3.36)
Dummy (1= Greater Urban Area)	6.856* (3.25)	7.367* (3.27)	-4.286 (-1.04)	—	—
Percent of Population in Main City	1.455* (10.43)	1.528* (11.80)	—	1.114* (6.68)	1.196* (7.55)
Percent of Amerindian Population	-0.087 (-1.51)	—	-0.314** (-2.59)	-0.136 (-1.60)	—
Intercept	0.911 (0.21)	-1.751 (-0.46)	11.768 (1.54)	4.739 (1.09)	0.959 (0.22)
Number of Observations	24	24	26	25	25
R-squared	0.90	0.90	0.52	0.84	0.82

Note: *t*-statistics are given in parenthesis

* Significant at 99 percent confidence

**Significant at 95 percent confidence

Amerindian populations are not statistically significant across all specifications. Regression 5 of table 3, which includes only the former two variables, has a high R^2 (goodness-of-fit) that suggests that 82 percent of the variation in national adult literacy across countries is “explained” by these variables. This shows that urban literacy can be employed to obtain a prediction of national literacy rates, even though urban population is not representative of national population.

ROBUST ESTIMATES OF NATIONAL ADULT LITERACY RATES, CIRCA 1900

The coefficients of regressions 2, 4, and 6 of table 2, and the coefficients of regression 5 of table 3 were employed to obtain literacy predictions for twenty-two Latin American and Caribbean nations circa 1900, using the data available from the statistical appendix.¹⁷ These regressions have a high goodness-of-fit (R^2), and they include only the variables that are systematically stable and statistically significant in all regressions. Columns 1 to 5 of table 4 show the estimates obtained from this procedure. Each literacy estimate in table 4 is accompanied by the year for which it is provided, which corresponds to the year of the alternative literacy rate employed for the estimate.

17. The countries excluded did not undertake any census circa 1900, and did not have sufficient reliable literacy from alternative sources.

Table 4 shows that census literacy rates are generally very similar to the estimates from the alternative sources, as one would expect from the high goodness-of-fit (R^2) of the regressions in tables 2 and 3. For those countries where the census and some of the alternative estimates are many years apart, such as Costa Rica, Colombia, Paraguay, Venezuela, and Peru, the existing differences in literacy rates may be explained partly by an upward trend in literacy, which would imply somewhat lower values for the earlier figures.¹⁸ On the other hand, the sign and magnitude of the discrepancies between census and literacy estimates obtained from the alternative sources can also result from some differences in the censuses' methodologies and procedures, as discussed earlier.

It will be observed that table 4 often provides more than one literacy estimate for a given country in addition to the census figure, generally for different years. In order to obtain a single literacy estimate for each country circa 1900, all the literacy estimates, including the census figures, and their corresponding years were averaged. This procedure is justified for two reasons. First, for many countries the census figures are too far from 1900 to claim that they are a reasonable estimate of literacy for this year, particularly for Trinidad and Tobago, Costa Rica, Colombia, Paraguay, Venezuela, and Honduras. In these cases, taking the closest available census figure to 1900 would not achieve comparable literacy rates across countries. Note, however, that for all these countries the procedure of taking the average of all literacy estimates (including census data) generates literacy estimates for each country that are much closer to 1900, which enhances comparability across countries. Secondly, as this research note argued earlier, some of the census data may be biased for various reasons. Taking an average of all literacy estimates may help reduce those possible biases in the census data. There are not big discrepancies between the census data circa 1900 and the estimates derived from the alternative literacy sources. This implies that following the alternative strategy of using only the existing census data around 1900 whenever available would yield a very similar portrait of the extent of literacy across the region. Column 6 of table 4 shows the average of the literacy estimates reported in columns 1 to 5 accompanied by the average of the years of those estimates, which for most countries ends up being only within a few years of 1900. The twenty-two nations in table 4 are ordered from highest to lowest literacy circa 1900 according to column 6. However, this ranking must be taken only as an approximation, as the literacy estimates do not belong exactly to a common year.

Table 4 shows that, on average, the prevalence of literacy in the region at the turn of the twentieth century was only about one-third of the adult

18. This seems particularly likely in the case of Costa Rica, Venezuela, Paraguay, and Peru, whose census figures are much earlier than many of the alternative literacy estimates.

Table 4 National Adult Literacy Estimates: Latin America, circa 1900

Country	National Census Adult Literacy		Estimates from Convicts and Criminal Offenders		Estimates from Marriages		Estimates from Military Recruits		Urban Literacy		Estimates from Averages (1)–(5)	
	(1)	(2)	(3)	(4)	(5)	(6)	%	Year	%	Year	%	Year
Uruguay	58.1	1900	59.4	1900	60.0	1900			56.9	1900	58.6	1900
Argentina	57.0	1895							52.8	1895	55.4	1898
									56.4	1904		
Trinidad/ Tobago	50.4	1911			52.9	1901					51.7	1906
Cuba	43.2	1899	50.6	1907	49.3	1899			44.2	1899	46.2	1901
			43.7	1899								
Jamaica	45.4	1901			41.3	1895					43.4	1898
Guyana*	39.2	1901			45.4	1901					42.3	1901
Costa Rica	38.9	1892	49.7	1907					32.8	1892	40.3	1899
									40.0	1904		
Colombia	42.4	1918	47.4	1917	38.8	1917	22.4	1878			37.8	1907
Panama	30.1	1911							41.6	1911	35.9	1911
Chile	36.0	1895	31.5	1907			28.3	1901	29.8	1895	31.5	1899
			31.7	1895								
Paraguay	18.7	1886	51.6	1915					22.9	1886	31.1	1896
Brazil	34.7	1900					28.3	1890	24.5	1906	29.2	1895
Venezuela	19.1	1891	23.4	1894	39.0	1909			25.0	1891	28.4	1899
			35.4	1909								
Nicaragua									27.9	1906	27.9	1906
Ecuador			20.5	1897	25.8	1897			30.0	1906	25.4	1900
Peru	16.7	1876							26.0	1905	24.8	1896
									31.5	1908		
Honduras	25.4	1888	27.3	1888	19.3	1888					23.3	1895
			20.7	1915								
Mexico	22.3	1900							19.3	1900	20.8	1900
El Salvador			19.6	1902	22.1	1902					20.8	1902
Puerto Rico	18.3	1899									18.3	1899
Bolivia	17.8	1900	14.5	1900					14.4	1900	15.6	1900
Guatemala	12.1	1893	10.4	1898					12.5	1893	14.1	1894
			9.51	1893			16.3	1893	22.3	1893		
Average	32.7	1897									32.8	1900

Note: * Averages between the 1891 and 1911 census figures.

population. However, there was a high degree of dispersion in literacy across the region. Uruguay and Argentina were well ahead in the region, where somewhat less than two-thirds of the adult population were literate. They are followed by Trinidad and Tobago, Jamaica, Guyana, and Cuba. In all these countries literacy rates ranging from between 42 and 52 percent of the adult population.¹⁹ Next come a group of independent Latin

19. The notable exception is Puerto Rico, a Spanish possession up to 1899 that ranked at the bottom of the table.

American republics that exhibited adult literacy around one-third of the adult population, all of which were characterized by an indigenous population that made up a relatively small proportion of the total population. Finally, the lagged last group is composed of independent Latin American republics and Puerto Rico, which generally have had a relatively large Amerindian population.²⁰ In this group of countries literacy ranged from only one-seventh of adult population in Guatemala, to about one-fourth in Peru.

Despite its diversity, the prevalence of literacy in all Latin American and Caribbean countries is substantially below contemporaneous literacy rates in most of Northwestern Europe and North America, where literacy ranged between 80 and 90 percent at the turn of the twentieth century.²¹ However, the Latin American countries in the upper part of table 4 exhibit literacy rates fairly similar to those of many Eastern and Southern European countries at the time. For example, adult literacy rates in the Russian Empire, Bulgaria, and Hungary in 1900 were 28, 30, and 60 percent, respectively.²² On the other hand, adult literacy in Italy and Spain were 53 and 44 percent, respectively, while in the less literate Portugal and Greece the respective figures were 27 and 39 percent.²³

However, Latin American literacy rates seem, on the whole, above the contemporaneous prevalence of literacy in other regions of the now-called "developing world," such as Africa and parts of Asia. For example, literacy in India and Egypt around 1900 was only about 5 to 7 percent (UNESCO 1953), which is even lower than the figures for the less literate countries in Latin America at the time, as shown in table 4.²⁴

The prevalence of literacy around 1900 in Latin America appears to be very low considering that during the nineteenth century most nations of the region formally enacted free and compulsory primary education laws (see n.7). However, as shown in table 5 literacy rates circa 1900 are positively associated with the prevalence of primary school enrollment, as expected.²⁵ Table 5 also shows a positive association of

20. With the exception of Puerto Rico and Honduras, which have a relatively small proportion of Amerindians.

21. For example, in 1871 Prussia already had 88 percent literacy, and in 1900 adult literacy in France and Belgium was 82.5 and 80 percent, respectively. In the same year adult literacy was 83 and 90 percent respectively in Canada and the United States (Cipolla 1969, UNESCO 1953). In the latter, however, the African American population's literacy rate was only 60 percent (Cipolla 1969, 14).

22. UNESCO (1953) and Cipolla (1969).

23. UNESCO (1953).

24. These are the few developing countries that have fairly comparable censuses around 1900 (UNESCO 1953). However, they are likely to provide some indication about the prevalence of literacy in other Asian and African countries at the time.

25. The enrollment data comes from Mitchell (1983), which was divided by the population in ages 6–14 years of age in each country obtained from the same source.

Table 5 Correlation Between Adult Literacy versus Amerindian and Urban Population, Latin America circa 1900

	Primary School Enrollment c. 1900	Amerindian Population	Urban Population*	Population in Capital City
Literacy 15+	0.63 (22)	- 0.58 (22)	0.65 (18)	0.86 (21)

Note: Number of observations in parenthesis.

*Population in cities and towns larger than 10,000 inhabitants, Sánchez-Albornoz (1973). All coefficients are significant at 99 percent confidence.

literacy and the percentage of urban population, and a negative association between literacy and the prevalence of Amerindian population. This latter finding may be related with the higher difficulty of teaching and learning reading and writing skills in a second language, an issue that we shall analyze in more depth below.

THE EVOLUTION OF LITERACY IN LATIN AMERICA AND THE CARIBBEAN, 1900–1950

Table 6 summarizes the comparative evolution of literacy and primary school enrollment in the region during the first half of the twentieth century. The first column of table 6 reports the adult literacy estimates circa 1900 as in column 6 of table 4. The measure of change in literacy in the period is expressed in column 5 of table 6 as the average change in literacy per decade, in order to have a common comparison across countries.

As a whole, between 1900 and 1950 Latin America and the Caribbean increased its literacy rate from one-third to nearly two-thirds of adult population, with a corresponding decennial literacy change of about 5.7 percentage points per decade. Moreover, all countries in the sample show improvements during the period, although with remarkable variation. Puerto Rico, Chile, Panama, Costa Rica, and Mexico stand out as the countries with the highest decennial change in literacy in the period. The exceptional case of Puerto Rico may be related to the substantial aid delivered by the United States following the Spanish-American War in 1898. A similar argument may explain the case of Panama (see Osuna 1923). Also, as appendix 1 (in original database, see n.7) shows, Chile is the very last country to establish free and compulsory education laws in the region, which had a large impact on schooling and on the quality of primary education after 1920. On the other hand, a group of Central American republics composed by Nicaragua, Honduras, Guatemala, and El Salvador, together with Bolivia and Brazil exhibit very poor literacy and primary education enrollment changes throughout the period. The poor performance of Brazil may be the result of a very specific feature—Brazil abolished compulsory primary education in 1891, and re-enacted it only in 1934, becoming with Colombia the only countries in the region to reverse compulsory

Table 6 Evolution of Primary Education Enrollment and Adult Literacy: Latin America, 1900–1950

	Literacy c. 1900 (1)	Year of Estimate (2)	Census Literacy c. 1950 (3)	Year of Census (4)	Decennial Literacy Change 1900–1950 (5)	Average* Primary Enrollment 1895–1900 (6)	Primary* Enrollment 1950 (7)	Total Enrollment Change 1900–50 (8)
Argentina	55.4	1898	86.4	1947	6.3	30.6	66.2	35.6
Bolivia	15.6	1900	32.1	1950	3.3	8.2	28.7	20.5
Brazil	29.2	1899	49.4	1950	3.9	10.6	32.7	22.1
Chile	31.5	1899	80.2	1952	9.1	20.7	58.2	37.5
Colombia	37.8	1907	62.3	1950	5.8	15.2	28.5	13.3
Costa Rica	40.3	1899	79.4	1950	7.6	28.8	47.3	18.5
Cuba	46.2	1901	77.9	1953	6.1	32.4	45.4	13
Ecuador	25.4	1900	55.7	1950	6.1	21.9	41.4	19.5
El Salvador	20.8	1902	39.4	1950	3.9	17.3	31.4	14.1
Guatemala	14.1	1894	29.4	1950	2.7	22.3	23.3	1.0
Guyana	42.3	1901	70.5	1946	6.3	41.6	78.7	37.1
Honduras	23.3	1895	36.3	1945	2.6	29.4	30.9	1.5
Jamaica	43.4	1898	72.4	1943	6.5	55.3	65.2	9.9
Mexico	20.8	1900	56.8	1950	7.2	20.7	38.6	17.9
Nicaragua	27.9	1906	38.4	1950	2.4	17.5	31.6	14.1
Panama	35.9	1911	69.9	1950	8.7	9.1	57.9	48.8
Paraguay	31.1	1896	65.8	1950	6.4	22.2	52.3	30.1
Peru	24.8	1896	47.0	1950	4.1	13.5	47.7	34.2
Puerto Rico	18.3	1899	74.4	1950	11.0	13.4	61.3	47.9
Trinidad/Tobago	51.7	1906	73.8	1946	5.5	44.9	83.2	38.3
Uruguay	58.6	1900	87.0	1950	5.7	32.1	64.8	32.7
Venezuela	28.4	1899	52.2	1950	4.7	15.9	39.8	23.9
Average	32.8	1900	60.8	1949	5.7	23.8	48.0	24.2

Note: * Enrollment rates are primary enrollment over population in ages 6–14, both taken from Mitchell (1983).

education legislation and maintain that situation for more than four decades (see appendix 1, n.7).

We now investigate some determinants of literacy change between 1900 and 1950. Table 7 reports the results of various correlations between measures of enrollment in primary education and adult literacy around 1900 and in 1950, and the change in both variables between 1900 and 1950. Table 7 shows that the increase in primary school enrollment during 1900–1950 is indeed significantly associated with the change in literacy across the region during this period. Moreover, in 1900 and 1950 the ranking of countries according to primary enrollment and literacy rates are positively associated, as well as the *changes* in the ranking of both variables during the interval 1900–1950.²⁶

Table 8 confirms these findings. Regressions 1–5 of table 8 show that the decennial change in literacy is associated with the increase in primary school enrollment, which explains about 42 percent of the

26. Rankings correlations are often employed as a robustness check to correlations between data that are suspected to be measured with errors or biases. See for example Dasgupta (1993).

Table 7 Correlation between the Spread of Primary Education Enrollment and Literacy: Latin America, 1900–1950

	Literacy Rank		Literacy Rank		Decennial Literacy Change	Literacy Rank Change
	c. 1900	c. 1900	c. 1950	c. 1950	1900–50	1900–50
Primary Enrollment, c. 1900	0.63*					
Enrollment Rank, c. 1900		0.60*				
Primary Enrollment in 1950			0.78*			
Enrollment Rank in 1950				0.79*		
Enrollment Difference, 1900–1950					0.59*	
Enrollment Rank Change, 1900–1950						0.39**

Notes: Number of observations: 22

* Significant at 99 percent confidence

** Significant at 95 percent confidence

variation in the increase in literacy across countries during this period. The coefficient of the change in primary school enrollment is stable in sign and magnitude and statistically significant at 99 percent confidence across all specifications. However, the evidence in table 8 also suggests that the efficacy of the expansion of school enrollment is associated with the countries' ethnic composition. In particular, the extent of a country's Amerindian population seems to have limited the impact of the enrollment change on adult literacy. As the data indicate, although the Amerindian variable coefficient is far from significant, the coefficient of the *interactive* variable constructed as the product of primary enrollment times the share of Amerindian population is negative and statistically significant when the Amerindian variable is dropped, as in models 4 and 5 of table 8. This result indicates that the *higher* the proportion of Amerindian population, the *lower* the impact of expansion in school enrollment on literacy. We provide the following interpretation for this finding. It seems likely that a significant share of the Amerindian population of the region during this period was either bilingual or spoke Spanish as a second language.²⁷ There is also ample evidence and literature on the greater difficulties and challenges

27. Note from the statistical appendix that the Amerindian population in the region is located mostly in Spanish-speaking countries.

Table 8 Regressions with Robust Standard Errors, Dependent Variable: Decennial Literacy Change 1900–1950

Variable	1	2	3	4	5
Enrollment Difference	0.106* (3.98)	0.111* (3.29)	0.107* (3.18)	0.118* (3.97)	0.112* (4.37)
Amerindian Population (%)		-0.021 (-0.83)	-0.011 (-0.50)		
Amerindian Population * Enrollment Difference Literacy c. 1900		-0.0011 (-1.32)	-0.0012 (-1.25)	-0.0018** (-2.46)	-0.0016** (-2.59)
Intercept	3.16* (4.51)	4.36* (2.87)	3.50* (3.65)	3.86* (3.74)	3.36* (4.87)
Observations	22	22	22	22	22
R-Squared	0.43	0.52	0.51	0.52	0.51

Note: *t*-statistics are given in parenthesis

* Significant at 99 percent confidence

** Significant at 95 percent confidence

associated with learning and teaching literacy skills in a second language, or to bilingual students.²⁸ As a consequence, the effectiveness of the expansion of primary education in transforming enrollment into literacy during this period would have been inversely associated with the prevalence of Amerindian population across countries.

The data in table 8 show that the countries' starting literacy rates in 1900 did not affect significantly the rate of increase of literacy in the following decades. This suggests that literacy rates across countries did not converge or diverge from each other during this period; having an initial advantage in adult literacy did not lead to a faster rate of increase in literacy in subsequent decades.

The regressions also indicate that, judging from the R^2 values, nearly half of the variation in literacy change is *not* explained by the change in schooling and the prevalence of Amerindian population alone. This suggests that there is ample room for other uncontrolled factors that may have had an important effect in expanding literacy across the region. These may be associated in particular with the remarkable variations in qualitative aspects and effectiveness of the increase of elementary education that seems to have existed across countries during the expansion of the Estado Docente.²⁹ The differences in literacy change during 1900–1950 across countries could have emerged from factors unrelated to the expansion of primary education, for example, the implementation of

28. See for example, Krashen (1981), Reynolds (1991), and Nunan (1999).

29. See, for example, Newland (1991, 1994) and Mariscal and Sokoloff (2000).

adult literacy campaigns, primary instruction received during military service, and instruction delivered by religious organizations.³⁰ Assessing the role and impact of these forms of instruction on adult literacy in the region remains as topics for future research.

Finally, it must be noted that the fact that literacy changes during 1900–1950 are associated with the increase in schooling and with the share of Amerindian population can be interpreted as further evidence of the robustness of the 1900 literacy estimates derived earlier.

CONCLUSIONS

Employing original data, this article examines the evolution and determinants of literacy in twenty-two countries of Latin America and the Caribbean from 1900 to 1950. The results show a wide dispersion of literacy rates among the countries of the region at the turn of the twentieth century, ranging from slightly above 10 percent to 60 percent of the adult population, the latter being fairly close to contemporaneous literacy rates in many European countries. This finding raises the question of how this significant dispersion in literacy rates emerged prior to 1900, which remains a subject for future research. The study also shows a significant disparity in literacy progress for the Latin American and Caribbean countries between 1900 and 1950. The evidence shows that this disparity is, as expected, largely associated with the rate of expansion of primary education policies and institutions in the period, as well as with country-specific features, most notably the proportion of the Amerindian population in each country. This feature may have conditioned the effectiveness of elementary education policies implemented in the era of the *Estado Docente* throughout the region.

The data generated by this research note can motivate future research in other related historical issues in Latin America. For example, employing literacy data as a proxy of human capital formation may help explain the variety of patterns of economic growth in the region during the twentieth century.³¹ In addition, given the established association between literacy and labor productivity, literacy data can shed light on the evolution and determinants of real wages in individual countries of the region during the period.

It is also hoped that the methodology offered here may stimulate future research on the study of literacy in other historical contexts where

30. In fact, Cipolla (1969) provides evidence that in nineteenth-century Europe, a significant fraction of literate adults learned to read and write outside schools, mainly in factory schools, during military service, and with the clergy. A similar situation may have happened in Latin American and the Caribbean during 1900–1950.

31. See for example Barro (1991) and Ranis et al. (2000).

census data may be scarce and/or questionable, for example prior to 1900 and *within* the period from 1900 to 1950. Finally, this methodology can be adapted to study the levels of literacy for males and females around 1900, as well as their evolution since then, although this would require gathering new data. Addressing these issues will contribute to a better assessment of the formation of human capital in the past, and to a better understanding of the process of economic, social, and political modernization in Latin America and the Caribbean.

REFERENCES

- Barro, R. J.
1991 "Economic Growth in a Cross-Section of Countries." *Quarterly Journal of Economics* 106 (20): 407–33
- Cipolla, C.
1969 *Literacy and Development in the West*. London: Penguin Books.
- Dasgupta, P.
1993 *An Inquiry into Well-being and Destitution*. New York: Oxford University Press.
- Gallardo, R.
1961 *Las Constituciones de El Salvador*. Madrid: Ediciones de Cultura Hispánica.
- Hicks, N., and P. Streeten
1979 "Indicators of Development: The Search for a Basic Needs Yardstick." *World Development* (7): 567–580.
- Krashen, S. D.
1981 *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.
- Lazcano y Mazón, A.
1952 *Las constituciones de Cuba*. Madrid: Ediciones Cultura Hispánica.
- Legarza, E.
1958 *Las constituciones de Nicaragua*. Madrid: Ediciones Cultura Hispánica.
- Mariñas Otero, L.
1962 *Las constituciones de Honduras*. Madrid: Ediciones Cultura Hispánica.
- Mariscal, E., and K. Sokoloff
2000 "Schooling, Suffrage and the Persistence of Inequality in the Americas, 1800–1945" in *Political Institutions and Economic Growth in Latin America*, edited by Stephen Haber, 159–218. Stanford: Hoover Institution Press.
- Mitchell, B. R.
1983 *International Historical Statistics*. London: The Macmillan Press.
- Newland, C.
1991 La educación elemental en Hispanoamérica: Desde la independencia hasta la centralización de los sistemas educativos nacionales, *Hispanic American Historical Review* 71 (2): 335–64.
1994 "The *Estado Docente* and its Expansion: Spanish American Elementary Education, 1900–1950." *Journal of Latin American Studies* 26 (2): 449–67.
- Nunan, D.
1999 *Second Language Teaching and Learning*. Boston: Heinle and Heinle.
- Osuna, J. J.
1923 *Education in Porto Rico*. New York: Teachers College, Columbia University.
- Peralta, H. G.
1962 *Las constituciones de Costa Rica*. Madrid: Instituto de Estudios Políticos.
- Ranis, G., F. Stewart, and A. Ramirez
2000 "Economic Growth and Human Development." *World Development* 28 (2): 197–219.
- Reynolds, A. G., ed.
1991 *Bilingualism, Multiculturalism, and Second Language Learning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Sánchez-Albornoz, N.
1973 *La población de América Latina desde los tiempos precolombinos al año 2000*. Madrid: Alianza Editorial.
- Sen, A.
1983 "Development: Which Way Now?" *The Economic Journal* 93: 745–62.
- Tenenbaum, B., ed.
1996 *Encyclopedia of Latin American History and Culture*. Vols. 1 and 2. London: Prentice Hall International.
- UNESCO
1953 "Progress of Literacy in Various Countries: Preliminary Statistical Study of Available Census Data since 1900." *Monographs on Fundamental Education*. Geneva: UNESCO.
1957 "World Illiteracy at Mid-Century." *Monographs on Fundamental Education*, Number 11. Geneva: UNESCO.
- Uzcategui, E.
1973 *Historia de la educación en Hispanoamérica*. Quito: Editorial Universitaria, Universidad Central del Ecuador.
- Valcarcel, C.D.
1975 *Breve historia de la educación peruana*. Lima: Editorial Educación.