

REGULAR ARTICLE

# Gender gap and spatial disparities in the evolution of literacy in Spain, 1860-1910

José Manuel Gutiérrez

Universidad de Salamanca, Spain, [jmgut@usal.es](mailto:jmgut@usal.es), ORCID iD: 0000-0002-2576-378X

Gloria Quiroga

Universidad Complutense de Madrid, Spain, [mariagloria.quiroga@pdi.ucm.es](mailto:mariagloria.quiroga@pdi.ucm.es), ORCID iD: 0000-0002-9825-0416

*Received: January 12, 2025. Returned: June, 30, 2025 Revised: July, 21, 2025. Accepted: October 3, 2025.*

---

**Abstract:** This article considers the dynamics of Spanish literacy in the period 1860-1910, characterized by local councils' responsibility of public elementary education. To this end, it is built a harmonized series of the literacy of the population aged ten or over, disaggregated by sex and province. Marked spatial differences and a very large gender gap can be observed. Five clusters are determined according to the male literacy rates of the provinces in 1860; these clusters prove to have explanatory power all along the period and for both sexes. A parsimonious statistical model of the evolution of male literacy during the period, introducing linguistic variables, shows a considerable temporal stability of the spatial distribution of male literacy. The model of the evolution of female literacy presents similarities with that of male literacy, although now the initial state (in 1860) is not described by female literacy, but yet by male literacy. All in all, the evolution of literacy in Spain between 1860 and 1910 did not follow the spatial pattern of the economic modernization process. Besides, there was no correlation between birth rates and literacy rates of children, for both sexes, and the same can be said of the correlation between urbanization and literacy. Considering the West European context, the Spanish literacy process during the period 1860-1910 was a failure, except for the geographical area of the top cluster.

**Keywords:** Historical censuses, literacy, nineteenth century, Spain, Official Statistics

**MSC:** 91F10, 62P25

---

## 1 Introduction

Human capital is a fundamental determinant of long-term economic and human development (Hanushek and Woessmann, 2010). In this sense, female illiteracy is a particular obstacle to economic

progress<sup>1</sup>. Sandberg (1982) argued that literacy rates, as a variable capturing the stock of human capital of the population, anticipate future increases in per-capita income. In general, all Western countries, except for Britain<sup>2</sup>, greatly improved their literacy rates during the early stages of modern economic growth. In principle, industrialization requires previous physical and, above all, human capital (as Galor's unified growth theory points out<sup>3</sup>), and a fast accumulation of them. Literacy allows the accumulation of information.

We consider herein the dynamics of literacy in Spain during the period between mid-nineteenth century and the First World War. From 1860 the Spanish censuses report reliable literacy data. The first aim of the article is to build a harmonized series of the literacy of the population aged ten or over, disaggregated by sex and province (NUTS-3 level)<sup>4</sup>, for the period 1860-1910. Vilanova Ribas and Moreno Julià (1992) compiled a harmonized series from 1887, but before 1887 the literacy data provided by the censuses are not broken down by age. In fact, for the period 1887-1910, we make use also of the censal literacy data not only disaggregated by sex and province, but also by age intervals.

This paper deals with the *evolution* of literacy between 1860 and 1910. Apart from the understanding that it provides *per se*, disaggregation by sex turns out to be a useful tool to study the dynamics of the literacy process during the period.

Not all historical data are of equal quality. Census data are usually the best when available and we have favoured them. For example, census data on the sectoral distribution of the labour force are the best available indicators of economic modernisation for the period under consideration. Exploratory data analysis allows for the consideration of variables that present problems of collinearity with other variables or whose available data are of inferior quality.

The period 1860-1910 is characterized by institutional factors that impinge on the funding of public elementary education. The comprehensive Public Instruction Law (1857) (known as Moyano Law after the incumbent minister), valid throughout this span of time, declared primary education compulsory (at least in theory) for all children between the ages of six and nine, and also free in public schools for the certified poor. The financing of public elementary education was left to municipalities. On the other hand, the *desamortización* of 1855 had confiscated the assets of (generally Catholic) educational foundations of all kinds<sup>5</sup> and most of the land belonging to the municipalities (until then a major source of the income of local councils). Near the end of the period, in 1901, the (central) state assumed the direct payment of teachers' salaries (although taking a percentage of municipal taxes in exchange), and in 1909 compulsory education was extended to the ages between six and twelve. Until 1910, the funding of public elementary education fell fully on the shoulders of municipalities<sup>6</sup>.

<sup>1</sup>See Bowman and Anderson (1963), Núñez (1992, 2003a) and Sarasúa (2002, 2019).

<sup>2</sup>A paradox appears in the English case. Although there is no complete consensus on the level of English literacy rates before and during the early industrial revolution, it seems that they improved substantially between 1642 and 1750, and then stagnated until 1815-1830, especially in industrial centres. This evolution led many researchers to assume the hypothesis that education was not an essential requirement for modernization. All the same, that stagnation was a consequence of the industrialization process itself, which initially needed unskilled labour (also regrettably including children, making higher the opportunity cost of attending school), and of the intense urbanization process, which was not accompanied by a similar increase in the educational supply. Once this first phase of industrialization had been overcome, the need for a more qualified workforce and the extension of the franchise meant that from 1840 the literacy process became widespread in England (see West (1978); Schofield (1973); Mitch (1993, 2013); Pleijt et al. (2020)).

<sup>3</sup>Galor (2011)

<sup>4</sup>The 1887 census is the only one providing literacy data by age at a higher level of disaggregation, that of judicial districts (*partidos judiciales*).

<sup>5</sup>These were assets that had survived the *desamortizaciones* of 1836-1837 and 1841, which affected the properties of the Catholic Church.

<sup>6</sup>See Terrón Bañuelos (1997).

Only from that year public primary education was made free for all, and the state began to contribute to the financing of primary education<sup>7</sup>.

The Moyano Law must be seen in the context of the laws which organised the alphabetization process in most Western European countries: the state establishes a national system of universal elementary education, imposing on local authorities the obligation to create and maintain schools. Financing is essentially dependent on local authorities and on the fees paid by parents<sup>8</sup>, although it is free for the poor. The differences between countries lie in the time of commencement of the process and the extent to which the legal provisions were actually implemented.

C.E. Núñez<sup>9</sup> has carried out relevant studies on literacy and education in Spain, showing that, although Spanish literacy rates performed poorly at the national level in the period 1860-1910, large regional differences can be observed. According to Núñez, the northern half of the country, except for Galicia, was more literate than the south-eastern Mediterranean coast, and some areas had in 1860 male literacy rates similar to those of the most advanced countries of Europe, whereas others displayed levels among the lowest in Western Europe. Fifty years later, even when male literacy rates had improved, the differences between provinces persisted. For their part, female literacy rates were low and much more homogeneous in 1860, which entailed a high gender differential overall, much higher in the more literate areas.

The spatial distribution of Spanish literacy reflected by the censuses, in which the highest levels of literacy correspond to rural areas of the northern plateau, has led to attempts at explanation. Beltrán Tapia and Martínez-Galarraga (2018) builds a formal model in which the ratio of day labourers to the total agricultural population turns out to be significant to explain literacy. On the other hand, Reher (2023) stresses the existence of important regional differences in the perceived value of literacy and education and argues that these cleavages go beyond the importance of economic structures and have deep historical roots.

Beltrán Tapia et al. (2021) studies the spatial convergence of literacy in Spain with data disaggregated by municipality, showing a positive correlation between the (global) literacy rates in 1860 and the change in the literacy rates between 1860 and 1900, and a negative correlation between the literacy rates in 1900 and the change in the literacy rates between 1900 and 1930.

We focus here on the evolution of literacy in the period 1860-1910. As for any analysis of the dynamics of a system, an essential point is how much the initial state explains the final state. Five clusters are determined according to the similarity measure naturally provided by the male literacy rates of the provinces in 1860 (see Section 4). The resulting clusters turn out to be spatially contiguous to a high degree.

The study of the evolution of literacy is facilitated by considering its age structure (available after 1887), as low literacy rates in 1860 may obscure the literacy effort if only “biologically linked” literacy rates are used in later years. Two literacy rates may have a *biological link*, in so far as they share part of their underlying populations. For example, the male rate (for men 10 years old and over) in 1900 has a biological link with the same rate in 1910, because the men 10 years old and over in 1900 still surviving in 1910 are part of the men 10 years old and over in 1910. In contrast, the male rate in 1900 has no biological link with the male rate for boys (between 11 and 15 years old) in 1910. Assuming that the acquisition of literacy for those older than a certain age becomes increasingly unlikely, a very

<sup>7</sup>Centralization in primary school systems fostered eventually literacy in Latin European countries like Portugal (see Reis (1993); Nunes (2003); Gomes and Machado (2020)), Italy (see Zamagni (1993) and Cappelli and Vasta (2020)) and Spain (see Núñez (1992)).

<sup>8</sup>Fees were eliminated later: in 1881 in France, in 1889 in Prussia and from 1910 in Spain. The Casati Law of Italy (1859) established full gratuity (for the first two years of education) upon its enactment.

<sup>9</sup>Núñez (1992, 1993a, 1997, 2003a,b, 2010).

low literacy rate in some year holds back the biologically linked literacy rates in later years (this is particularly the case with female rates).

The clusters considered above prove to have explanatory power all along the period and for both sexes. The 6 provinces of the first cluster, the “Castilian core”, had surpassed the threshold of 75% male literacy in the 1870s, and still headed the list in 1910. All the 12 provinces of the fifth cluster, “South and East”, were among the bottom 14 provinces by male literacy in 1910, with values between 30% and 45%. The statistical results show a considerable temporal stability of the spatial distribution of male literacy. We obtain a rather parsimonious model of the evolution of male literacy during the period, with coefficient of determination round 90% and only three regressors: the initial state (in 1860) and two linguistic variables. There is no influence of economic modernization: the proportion of the male active population working in agriculture, or the same proportion of those working in industry, turn out to be non-significant.

The harmonized literacy series for the period 1860-1910 facilitates especially the analysis of the female literacy development. The overall female rate in 1860 was low enough, 11.2%, in contrast with the certainly lacklustre male rate, 38.9%. Unlike male literacy, there is no clear spatial pattern of female literacy in 1860. Generally speaking, the main feature of the dynamics of female literacy in the period is that the greatest increase in female literacy did not occur in the most economically developed or urbanized areas, but rather in those provinces that in 1860 had the highest levels of male literacy<sup>10</sup>. In fact, the female literacy of 1910 is predicted quite well by the male literacy of fifty years before. The resulting model of the evolution of female literacy during the period presents similarities with that of male literacy, although now the initial state (in 1860) is not described by female literacy, but yet by male literacy. Indeed, female literacy in 1860 turns out to be non-significant for the evolution of female literacy in the period 1860-1910.

The evolution of literacy in Spain between 1860 and 1910 did not follow the spatial pattern of the economic modernization process, as it is usually the case<sup>11</sup>. On another note, birth rates and urbanization are two issues impinging on the financial constraints of parents and local councils, the essential funders of elementary education during the period. In this sense, there was no correlation between birth rates and literacy rates of children, for both sexes, and the same can be said of the correlation between urbanization and literacy (data of 1910). Certainly, the Moyano Law allowed small villages to maintain mixed-sex schools and opt for less paid (and qualified) teachers. But, ultimately, the list of per capita investors in public primary education was headed by rural provinces with high literacy (data of 1908).

All in all, the Spanish literacy process during the period 1860-1910 was a failure. If we consider eight West European countries able and willing to provide reliable censal literacy data during the 19th century, Spain had the fifth literacy level and the widest gender gap at the beginning of the period, and the eighth literacy level and still the widest gender gap at the end. Certainly, there were drastic regional differences, and, by 1910, literacy was almost universal among girls in the Castilian core, while only around one quarter of girls were literate in the South and East cluster.

The article is organized as follows. Section 2 examines the level of Spanish literacy in the period 1860-1910 within the European context. Section 3 deals with the methodology used to build a harmonized series of the literacy of the population aged ten or over; the series is provided in Appendix 1. Section 4 and Section 5 present the evolution of male and female literacy, respectively, and discuss

<sup>10</sup>See Núñez (1992) p.122: “the progress [of female literacy] was more intense where male literacy was more widespread”. Cf. also Beltran Tapia et al. (2021) (footnote 43): “female literacy grew rapidly in municipalities where male literacy was already high”.

<sup>11</sup>See Smith (1976. Book I-Chapter 10), Hanushek and Woessmann (2010) and Sandberg (1982). On the other hand, see Reher (1997).

the link between female literacy at the end of the period and male literacy 50 years earlier. Section 6 concludes.

## 2 The European context

Literacy concerns primarily communication. An individual who can communicate with another by means of written language is regarded as a “literate” person, and one who does not possess this ability is considered an “illiterate” (see UNESCO (1957)). Thus, literacy comprises both reading and writing<sup>12</sup>.

In order to estimate the literacy level of a population, we face considerable problems, conceptual and practical, especially when we go back in time<sup>13</sup>. The advent of modern censuses in the mid-19th century opened new possibilities for the measurement of literacy. In modern censuses data were obtained on all individuals present in the household on the specified census day. Information was self-reported by the household heads through household forms (later individual forms). A field force of professional enumerators was employed to assist in the process from house to house (especially if there was no one in the house who could write) and collect the forms<sup>14</sup>.

All modern censuses had a similar basic methodology, and thus comparisons between countries are made easier<sup>15</sup>. It must be considered when the data refer to literacy (ability to read and write) or semi-literacy (ability to read). We shall refer here to literacy data in Western Europe. Sometimes the first censuses gave literacy data without distinguishing ages, setting or not setting a minimum age to obtain the data (4, 5 or 6 years); in these cases, obtaining a true literacy rate (from 10, 11, 12 or 15 years old) requires estimation work (which can be quite precise if the necessary auxiliary data are available, as it is usually the case).

The first (modern) census in Western Europe with literacy data is that of Ireland in 1841. Then we have the censuses of Spain (1860), Italy (1861), Belgium (1866) and France (1866). More censuses are added later, carried out in tune with the political borders of their time. Some European countries have never included questions on literacy in their censuses (e.g., the United Kingdom (except Ireland) and Denmark), or have done so very late (e.g., Sweden, in 1930).

Taking into account the political entities in Western Europe where censal literacy data are available already in the nineteenth century (Ireland, Spain, Italy, Belgium, France, Prussia, Cisleithania<sup>16</sup> and Finland), there are three countries where the literacy rate had not reached the 75% level before

<sup>12</sup>The modern UNESCO proposed definition reads: “A person is considered *literate*, who can both read with understanding and write a short simple statement on his everyday life” (see UNESCO (1957); the proposal was made by a committee in 1951).

<sup>13</sup>As for the pre-statistical age, the main tool of analysis is considering who could sign and who could not sign in documents (such as marriage certificates, deeds, wills, etc.), and even the quality of the signatures. Apart from the issue of how representative of the population is the sample in each case, the ability of an individual to write his/her name does not entail, in principle, a general ability to read or write, although there can be statistical correlations (see Furet and Sachs (1974)).

<sup>14</sup>See Baffour et al. (2013) about modern censuses and their evolution.

<sup>15</sup>See UNESCO (1953) about problems arising in censal literacy data. Besides, when literacy is self-reported there are attendant issues of possible upward bias. A test was implemented in 1864 to check the accuracy of the literacy self-report of the conscripts in France, with the result that their statements were highly reliable (see Furet and Ozouf (1977)).

<sup>16</sup>The term “Cisleithania” denotes the northern and western part of Austria-Hungary, containing Austria proper, present-day Czechia and other *crown lands* (“Kronländer”). After the *Compromise* (“Ausgleich”) of 1867, the Austrian Empire was transformed into the dual monarchy of Austria-Hungary, constituted by two parts, with their respective parliaments and governments: *Cisleithania* (the Austrian part) and *Transleithania* (lands of the “Archiregnum Hungaricum”).

the First World War: Spain, Italy<sup>17</sup> and Finland<sup>18</sup>. In fact, the 75% threshold was reached in the three of them by the 1930-1940 interval, beyond the time span considered in this paper.

A comparative picture of the literacy processes in Western Europe on the eve of and during the Second Industrial Revolution is provided in Gutiérrez (2024), taking censal literacy rates as a yardstick to measure and compare literacy in different countries<sup>19</sup>. If only partial or insufficient censal data are available, literacy is assessed as if given by full censal data. A set of comparable (as far as possible) literacy data is built. The non-Spanish data in this section are in that paper or directly taken from the censuses referenced there.

It is worth highlighting the large gender gap in Spanish literacy rates. At the beginning of the period, in 1860, it is 27.7 points in Spain, while it is 15 points in Ireland (1861), 16.4 in Italy (1861), 7.5 in Belgium (1866) and 11.4 in France (1866). At the end of the period, in 1910, the gender gap is still 19 points in Spain, to be compared with 1.0 in Ireland (1911) or 3.4 in Belgium, but 12.4 in Italy (estimation, see Table 1 below).

On the other hand, Spain lost ground to Italy and Finland during the period 1860-1910. Table 1 focuses on the literacy rates of Spain, Italy and Finland. For each country and census, three percentages of literacy are stated: for men and women, separated by a hyphen, and the overall percentage (marked bold) in the bottom row. The minimum age limits are indicated in brackets under the name of every country (the figures from the 1861 census in Italy are for the population aged 12 or over, and those from the 1881, 1901 and 1911 censuses for the population aged 10 or over; data from the 1880 census in Finland are for the population aged 10 or over, and those from the 1900 and 1910 censuses for the population aged 15 or over). The dates of the censuses have been made homogeneous by linear interpolation. The problem arises that in Italy there are no data on literacy after 1881, but only on semi-literacy. In the table, a crude estimate (written in italics) of the Italian literacy rates in 1900 and 1910 has been given, subtracting from the (interpolated) Italian census percentages of semi-literates (out of the population aged 10 or more) the estimates of the percentages of those individuals able only to read (but not to write), assuming that the latter percentages are equal to those of Spain in the same year. These estimates are accurate for the purpose of the comparison between Spain and Italy (using them is equivalent to considering semi-literacy rates in both countries), but they are not so accurate beyond this comparison<sup>20</sup>.

At the beginning of the period<sup>21</sup>, Spain had the highest literacy level of the three countries. At the end of the period, it had the lowest. The comparison with Italy is particularly relevant (in the

<sup>17</sup>Spain and Italy were in this period predominantly agricultural economies and followed a *Latin pattern of modernization* (Tortella (1994), p. 5): relative backwardness in the nineteenth century and recovery in the twentieth century.

<sup>18</sup>Note that Finland followed the Swedish model (based on home instruction of the ability to read known texts: a set of selected religious texts, emphasizing submission to authority), with high *restricted semi-literacy* and low literacy until modern school systems were introduced. See Johansson (1977) and Tveit (1991).

<sup>19</sup>There is a typo in Table 1 (p. 43), in the last row of the "Slovenia" column: it says 88.5 and should read 81.5.

<sup>20</sup>The percentage of semi-illiterates (people who can read but not write) was small in Italy (in 1861 it was 3.9% for men, 5.5% for women and 4.7% overall, and in 1881 it was 1.2% for men, 3.4% for women and 2.3% overall) and in Spain (in 1887 it was 2.2% for men, 4.5% for women and 3.4% overall, and in 1910 it was 1.0% for men, 2.3% for women and 1.7% overall (Vilanova Ribas and Moreno Julià, 1992).

<sup>21</sup>Prior to 1860, there are no safe data on literacy in Spain, but research based on the counting of signatures in various sources and regions (see Bennassar (1985), Rodríguez and Bennassar (1978) and Larquí (1981)) seems to show that the literacy level in Spain was similar to that in France during much of the Old Regime. At any rate, after the Spanish War of Independence against Napoleon (1808-1814), arguably the bloodiest event in Spain's modern history, it is clear that Spanish literacy entered a period of relative decline.

case of Finland it must be considered that in 1880 almost the whole Finnish population had at least restricted semi-literacy<sup>22</sup>).

	Spain ( $\geq 10$ )	Italy ( $\geq 12, \geq 10$ )	Finland ( $\geq 10, \geq 15$ )
1861	39.2-11.6 <b>25.1</b>	30.4-14.0 <b>22.17</b>	
1880	44.9-19.4 <b>31.7</b>	44.8-26.8 <b>35.8</b>	16.2-10.2 <b>13.1</b>
1900	52.7-30.5 <b>41.2</b>	56.1-41.1 <b>48.4</b>	41.1-36.5 <b>38.8</b>
1910	57.6-38.6 <b>47.7</b>	66.4-54.0 <b>60.0</b>	57.4-53.3 <b>55.3</b>

Table 1: Literacy rates in Spain, Italy and Finland.

However, as we shall see, in a predominantly rural area corresponding approximately to the original Castile, the threshold of 75% in the male literacy rate had been reached already in the interval 1871-1880 (in the 1877 census), whereas in a fifth of Spanish provinces it was less than 30%. At any rate, low female literacy was a burden spread throughout the country, to a greater or lesser degree, which meant that the literacy gender gap was very large (particularly in the most literate provinces), even when compared to Italy.

It is worth considering the internal spatial differences in literacy levels of other countries around 1870.

The most literate areas of France in 1866 were, apart from Paris and its surroundings, the north-eastern part of the country: Alsace, Lorraine, Franche-Comté and Champagne, with literacy levels above 75%. There was a gulf with the least literate areas, located in the eastern Pyrenees, much of Brittany and a strip in central France covering Perigord, Berry, Bourbonnais and the east of Limousin; in all of them, the percentage of literates did not reach a third of the population in 1866.

In Italy (1871 census, semi-literacy rates for the population aged 6 or over) literacy decreased from north to south, with the highest values in Piedmont (66.3% for males and 49.2% for females) and Lombardy (59.3% for males and 50.3% for females), and very low values in the south, with the smallest values in Basilicata (19.1% for males and 5.3% for females) and Calabria (20.9% for males and 5.3% for females)<sup>23</sup>.

Despite the high literacy of Prussia already in 1871, a swathe of land along the far east of the country (Prussia proper, Posen and Upper Silesia) had literacy rates below 75%, with a minimum of 57.1% in Bromberg. This area corresponded to the districts with a sizeable Polish-speaking minority (Prussia proper) or a Polish-speaking majority (Posen and Upper Silesia). The rest of the country had literacy rates above approximately 90%, except for part of Pomerania (83.3% in Köslin and 84.1% in Stralsund). The highest values were in Berlin (97.4%) and the rural district of Sigmaringen (97.2%).

Unlike France and Prussia, in Italy there were still intense spatial differences at the end of the period (1911 census, semi-literacy rates for the population aged 6 or over): in Piedmont the literacy rates were 90.9% for males and 87.2% for females, whereas in Calabria they were 40.5% for males and 21.9% for females.

<sup>22</sup>Semi-literacy was high in Finland: in 1880 it was 81.0% for males, 87.6% for females and 84.4% overall, and in 1910 it was 41.3% for males, 45.7% for females and 43.6% overall (Myllyntaus, 1990). These Finnish figures should be understood in the sense of restricted semi-literacy (see the footnote above), as Finland followed the Swedish model.

<sup>23</sup>See Noble (1965) (p. 300).

### 3 Censuses and literacy rates

The first aim of this paper is to provide a census-based time series from 1860 to 1910 of the Spanish literacy rates, disaggregated by sex and province, for the population 10 years old and over (see Appendix 1)<sup>24</sup>.

There are five “complete censuses” with literacy data in the period: 1860, 1877, 1887, 1900 and 1910 (the “incomplete censuses” of 1857 and 1897 are not relevant here)<sup>25</sup>. From 1887 onwards combined data of literacy and age appear in the censuses; in 1860 and 1877 these two kinds of information are given separately. There are 49 provinces in Spain throughout the period<sup>26</sup>. The city of Ceuta is always included in the province of Cádiz<sup>27</sup>, and the city of Melilla is grouped with other small “plazas de soberanía” and is treated for census purposes as one more province, with which 50 divisions appear in the censuses. Colonial data are not considered in this paper.

Both the *de facto population* and the *de jure population* (usually resident population) are provided in all censuses (except in 1860, where only the *de facto* population is given). Literacy figures are taken from the *de facto* population. The number of individuals unspecified for literacy is also provided in all censuses (except in 1860), and it is always low (0.04% over the population aged 10 or over in 1877, 0.1% in 1887, 0.1% in 1900, 0.3% in 1910).

We have not excluded from the population those individuals unspecified for literacy when calculating the literacy rates (which is equivalent to considering them illiterate). Obviously, these literacy rates are lower than when individuals unspecified for literacy are excluded from the reference population (as it is done most frequently).

Now we set up some notation. Let us consider a certain group of people (Spain, a province, the women of that province, etc.), which is clear from the context.  $P_{10}$  is the population aged 10 or over in this group, and  $A_{10}$  is the literate population aged 10 or over. In general,  $P_k$  is the population aged  $k$  or over and  $A_k$  is the literate population aged  $k$  or over. We now define  $T_k = \frac{A_k}{P_k}$ , the literacy rate of individuals aged  $k$  or over. Similarly,  $P_{11-15}$  is the population aged 11-15 (inclusive) and  $A_{11-15}$  is the literate population aged 11-15. In general,  $P_{m-n}$  is the population between  $m$  and  $n$  years (inclusive) and  $A_{m-n}$  has the obvious meaning. The literacy rate for individuals aged between  $m$  and  $n$  (inclusive) is  $T_{m-n} = \frac{A_{m-n}}{P_{m-n}}$ .

Our purpose is to obtain  $T_{10}$  for each province, for men and women, in the 1860, 1877, 1887, 1900 and 1910 censuses. Also, the child literacy rates  $T_{11-15}$  in the 1887, 1900 and 1910 censuses are to be calculated.

The values of  $P_{10}$ ,  $P_{11-15}$ ,  $A_{10}$  and  $A_{11-15}$  can be found immediately, for each province, for both men and women, in the 1887, 1900 and 1910 censuses, and thus  $T_{10}$  and  $T_{11-15}$  can be calculated for these censuses. The rest of this section provides an exposition of how we estimate the value of  $T_{10}$  (through the values of  $P_{10}$  and  $A_{10}$ ) in the 1860 and 1877 censuses, in each province, for men and for women.

In the 1877 census we can calculate  $P_{10}$  directly. On the other hand, in the 1860 census the only relevant values given are  $P_0$ ,  $P_{0-5}$  and  $P_{6-10}$ . We estimate  $P_{10-10}$  (and, from there,  $P_{10}$ ) by calculating

<sup>24</sup>Individuals who do not state their level of literacy are considered illiterate in this paper.

<sup>25</sup>See Melón (1951), Cusidó i Vallverdú and Gil-Alonso (2012), for comparative analyses of the Spanish censuses of the period.

<sup>26</sup>This geographical division has remained stable since its creation in 1833 until 1927, when the Canary Islands split into two provinces, *Las Palmas* and *Santa Cruz de Tenerife*, and has remained so until today. We shall use the official names of the provinces (as appearing in the censuses of the period) throughout.

<sup>27</sup>In the 1887 census the literacy data of Ceuta are not even provided separately and are subsumed into those of the judicial district (*partido judicial*) of Algeciras.

the coefficient  $\frac{P_{10-10}}{P_{6-10}}$  from the 1877 census and using it as an estimate of the same coefficient in the 1860 census.

What is left is to estimate  $A_{10}$  in 1860 and 1877, in each province, for men and for women (as for literacy data, only  $A_0$  is provided in these censuses). In order to do this, if we write  $T'_{0-9} = \frac{A_{0-9}}{P_{6-9}}$ , it suffices to estimate  $T'_{0-9}$ . Setting  $T'_6 = \frac{A_0}{P_6}$  (a value known in 1860 and 1877), we now propose the simple regression model (SRM), separately for men and women:

$$T'_{0-9} = \beta_0 + \beta_1 T'_6 + \epsilon$$

where the variables run through the values of the provinces. The first census in which the values of  $T'_{0-9}$  are known is that of 1887. We now estimate the values of the parameters  $\beta_0$  and  $\beta_1$  with the data from 1887, intending to use them thereafter to “predict backwards” the values of  $T'_{0-9}$  in 1860 and 1877<sup>28</sup>. The results are:

- Men:  $\beta_0 = -0.1039$ ,  $\beta_1 = 0.6848$ , with  $R^2 = 0.8312$
- Women:  $\beta_0 = -0.0364$ ,  $\beta_1 = 0.7796$ , with  $R^2 = 0.8487$

We have now a strictly increasing functional relation. In order to assess the “backwards predictive” capacity of the model with these parameters, we apply the equation with the estimated parameters to predict the values of  $T'_{0-9}$  in 1900 and 1910, the two censuses after 1887. In these censuses we know the true values of  $T'_{0-9}$ , and we can see to what extent the prediction deviates from the true value. For 1900 the resulting coefficients of determination are as follows:

$$\begin{aligned} \text{Men : } R^2 &= 0.8432 \\ \text{Women : } R^2 &= 0.8248 \end{aligned} \tag{1}$$

For 1910 the  $R^2$  values are 0.7476 for men and 0.7840 for women. As these coefficient of determination results are rather good at predicting the  $T'_{0-9}$  values in the two subsequent censuses (1900 and 1910), with the parameters calculated using the 1887 data, it can be plausibly concluded that the “backward prediction” in the two antecedent censuses (1877 and 1860) will also be good. It must be taken into account that, as it is discussed below, literacy changed less in the period 1860-1887 than in the period 1887-1910<sup>29</sup>, and thus the backward predictions for 1860 and 1877 are not likely to be worse than the predictions for 1900 and 1910.

We have tried other methods to estimate the  $T'_{0-9}$  values in 1877, using two regressors, apart from the independent term. Specifically, we have considered the multiple regression model (MRM):

$${}^n T'_{0-9} = \beta_0 + \beta_1 {}^n T'_6 + \beta_2 {}^{n+1} T'_{0-9} + \epsilon$$

where  ${}^n T'_{0-9}$  and  ${}^n T'_6$  indicate values in the  $n$ -th census and  ${}^{n+1} T'_{0-9}$  in the subsequent  $(n + 1)$ -th census.

<sup>28</sup>Melilla presents atypical characteristics and is not included in the estimation of the model parameters, here and in subsequent regression models (and correlation coefficients) appearing in the paper. Accordingly, the  $T'_{0-9}$  values of Melilla in 1860 and 1877 are estimated, for men and women, by multiplying the  $\frac{T'_{0-9}}{T'_6}$  ratio of Melilla in 1887 by the  $T'_6$  value of Melilla in 1860 and 1877, respectively. On the other hand, the values of  $T'_6$  for women in 1860 of five provinces are very low and far out of the range of the values of 1887, and this results in slightly negative estimations of  $T'_{0-9}$  which have been taken zero instead; at any rate, the affected figures are negligible.

<sup>29</sup>Although the  $T_{10}$  rates will be used later for this discussion (which for 1860 and 1877 are obtained using the parameters considered now), there is no circularity in the reasoning, since the result is the same using the cruder rates  $T'_6$ , which can be obtained directly from the census data.

We now estimate the values of the parameters  $\beta_0$ ,  $\beta_1$  y  $\beta_2$  with the data from 1887 ( $n$ -th census) and 1900 ( $(n + 1)$ -th census). The results are:

- Men:  $\beta_0 = -0.0538$ ,  $\beta_1 = 0.2056$ ,  $\beta_2 = 0.7585$ , with  $R^2 = 0.9246$  and  $\bar{R}^2 = 0.9214$
- Women:  $\beta_0 = -0.0251$ ,  $\beta_1 = 0.2937$ ,  $\beta_2 = 0.5746$ , with  $R^2 = 0.9254$  and  $\bar{R}^2 = 0.9221$ .

The multiple regression model (MRM) certainly provides a better fit to the data (from 1887 and 1900) than the simple regression model (data from 1887), for both men and women. Another question is its ability to “predict backwards” the values of  $T'_{0-9}$  in 1877 (the values of the regressor  ${}^{n+1}T'_{0-9}$  are known, as they correspond to the 1887 census). In order to compare this ability in the two models, we apply the multiple regression equation with the parameters now estimated to predict the values of  $T'_{0-9}$  in 1900, the post-1887 census. In 1900 we know the true values of  ${}^nT'_{0-9}$  (as we know those of the regressor  ${}^{n+1}T'_{0-9}$ , now corresponding to 1910) and we can see how far the prediction deviates from the true value. The coefficients of determination are as follows:

$$\begin{aligned} \text{Men: } R^2 &= 0.6806 \quad \text{and} \quad \bar{R}^2 = 0.6667, \\ \text{Women: } R^2 &= 0.7704 \quad \text{and} \quad \bar{R}^2 = 0.7605. \end{aligned} \tag{2}$$

Comparing (1) and (2), the results of the multiple regression model (MRM) are worse than those of the simple regression model (SRM) at predicting the values of  $T'_{0-9}$  in 1900, for both men and women. Therefore, we choose the simple regression model to “predict backwards” the values of  $T'_{0-9}$  in 1877<sup>30</sup>. Even more so, we discard a similar multiple regression model (in which a regressor of the type  ${}^{n+2}T'_{0-9}$  would have to appear) to estimate the values of  $T'_{0-9}$  in 1860, and we also maintain in this case the simple regression model.

## 4 The evolution of male literacy

### 4.1 The starting point

Spain in 1860 is a predominantly agrarian country, with little modern industry, except in the province of Barcelona. Politics is unstable, marked by *pronunciamentos* and uprisings, and it will continue to be so until 1876, with the end of the third Carlist (civil) War, at the beginning of the Bourbon Restoration. Spanish is the predominant language, although Basque, Catalan-Valencian and Galician are also spoken (see below).

Based on census data, we can estimate that the Spanish literacy rate for males (aged 10 and over) was 38.9% in 1860. This global level of literacy says nothing about an important part of reality: a heterogeneous and unusual spatial distribution of male literacy.

We introduce a clustering of the 49 Spanish provinces according to their male literacy rates in 1860 (i.e., the similarity measure is the modulus of the difference between the literacy rates). In fact, we proceed by ranking the provinces in descending order of their male literacy rates, assigning proportional indices to the male literacy rates of the provinces (with 100 corresponding to the mean

<sup>30</sup>Two alternative multiple regression models have also been considered, in which the second regressor is  ${}^{n+1}T_{10-10}$  or  ${}^{n+1}T_{16-20}$ , instead of  ${}^{n+1}T'_{0-9}$ , but in them the coefficients of determination and the corrected coefficients of determination are worse than those of the multiple regression model (MRM), for both men and women, and thus they have been ruled out.

Spanish male literacy rate), and then taking the indices 160, 130, 100 and 70 as dividers between clusters<sup>31</sup>. The following five clusters are obtained (see Figure 1<sup>32</sup>):

- *Castilian core*, with male literacy rates above 65%. It is made up of 6 provinces, geographically contiguous, roughly corresponding to the County of Castile becoming a (more or less) independent entity in the mid-tenth century. By 1877, five of these provinces (Álava, Burgos, Palencia, Santander and Soria) had surpassed the threshold of 75% male literacy (the sixth province, Segovia, reached 72.1% in that census). It is a predominantly rural area, with a prevalence of small villages and no town with more than 20,000 inhabitants, except for Burgos with 25,000 and Santander with 30,000.
- *Northern Plateau*, with male literacy rates between 50% and 62%. It is made up of 8 provinces located in the Northern Plateau (*Meseta Norte*) or on its edge, around the Castilian core. Here is situated Madrid, the capital and largest city of Spain, with 300,000 inhabitants, but also the very rural region of León (provinces of León, Zamora and Salamanca; the university city of Salamanca has only 16,000 inhabitants).
- *Sundry North*, with male literacy rates between 40% and 49%. It is the only cluster without geographical unity, made up of 6 provinces located at different points in the northern half of Spain. Here is situated Barcelona, the second city in Spain, with almost 200,000 inhabitants. It is the most linguistically diverse cluster, including areas with a predominance of the Spanish, Catalan, Basque or Galician languages.
- *Transition*, with male literacy rates between 28% and 39%. It is made up of 17 provinces, which (except for three of them) constitute a continuous swathe of land from the eastern Pyrenees to the southwest coast in the Atlantic, separating the first three clusters from the fifth. Here is situated Seville, the third Spanish city, with almost 120,000 inhabitants, and three other cities with more than 50,000 inhabitants.
- *South and East*, with male literacy rates below 27%. It is made up of the Balearic and Canary Islands and 10 provinces that constitute a continuous strip along the Mediterranean coast, except for the northern part of it. Here you can find some of the richest agricultural areas and six of the twelve Spanish cities with more than 50,000 inhabitants, including Valencia, the fourth Spanish city, with almost 110,000 inhabitants.

Our clustering procedure is straightforward, dealing with attribute similarity, where the attribute is one-dimensional and naturally ordered. The resulting clusters turn out to be spatially contiguous to a high degree. Moreover, the spatial distribution of literacy shows a mostly concentric pattern with the province of Burgos at the centre. The whole first cluster is included in the first ring formed by Burgos and the bordering provinces. If we consider the 19 provinces of the second ring (the first ring and its bordering provinces), it turns out that it coincides with the 19 provinces with the highest male literacy rate (with only two exceptions).

The spatial pattern of the male literacy rate in 1860 is remarkable. There is no positive correlation between male literacy and the level of urbanization (measured by the percentage of the population living in the provincial capital or in towns with more than 30,000 inhabitants): the correlation coefficient is  $\rho = -0.0096$  (we always consider disaggregation by province). There is also no appreciable correlation between male literacy and the level of industrialization, whether the latter is measured

<sup>31</sup>In this case, our procedure is equivalent to clustering by shading (see Johnson and Wichern 1998). The number of clusters have been determined by inspection of the distribution of the indices.

<sup>32</sup>The present official names of the provinces are used in the maps: "Oviedo" is now "Asturias", "Santander" is "Cantabria", "Logroño" is "La Rioja", "Guipúzcoa" is "Gipuzkoa", "Gerona" is "Girona", "Lérida" is "Lleida", "Coruña" is "A Coruña" and "Orense" is "Ourense".

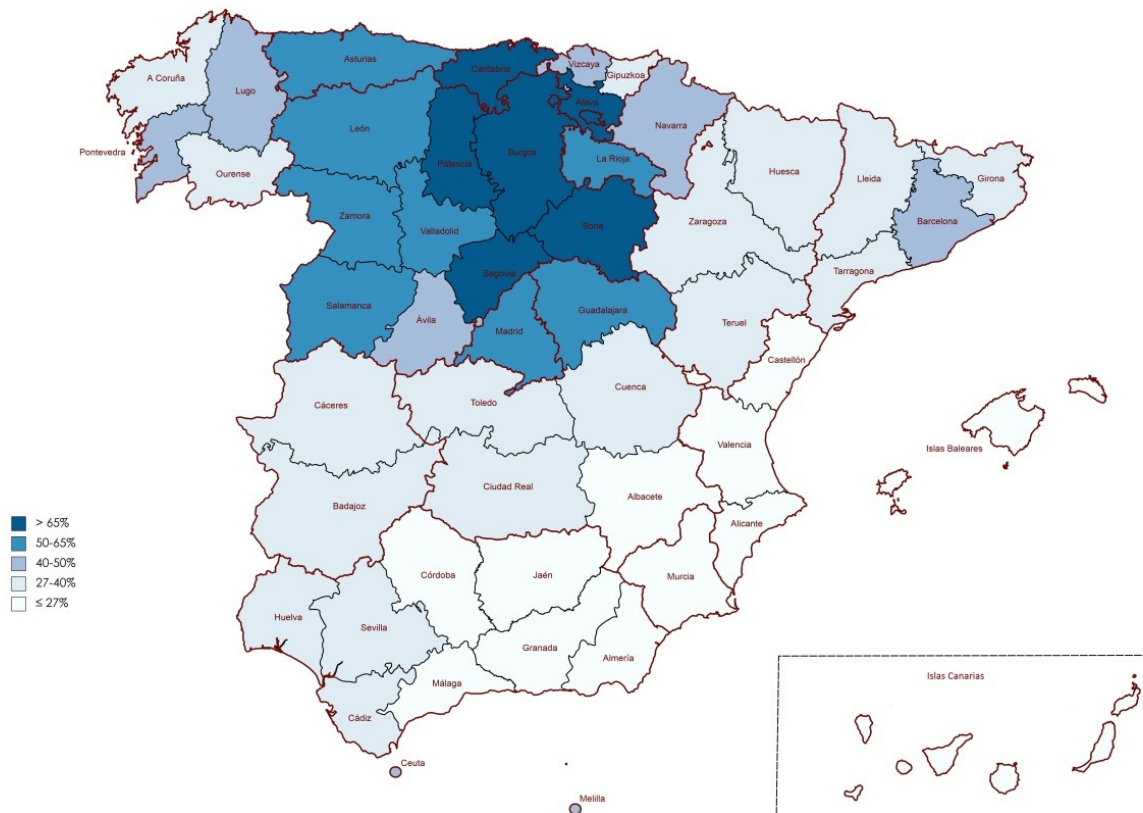


Figure 1: Literacy rates (males aged 10 and over) in 1860.

by the ratio of “workers in factories” (“jornaleros en las fábricas”) to the total population, or if both miners and workers in factories are in the numerator: the correlation coefficients are  $\rho = 0.0017$  and  $\rho = -0.0784$ , respectively.

Building explanatory models of literacy in 1860 would require the analysis of data much earlier than 1860. The purpose of this paper is rather to explain the *evolution* of literacy between 1860 and 1910, taking literacy rates in 1860 as initial conditions.

## 4.2 Evolution in the period 1860-1910

Literacy levels are the result of decisions made, on the one hand, by local authorities in the municipalities, where the provision of schools and teachers is established, and, on the other hand, by individuals, who decide whether they send their children to school, and for how long, or whether they try to become literate as adults. As in any decision problem, both types of decision makers, local authorities and individuals, have preferences and constraints.

The preferences of decision makers are part of their *mentality* and are marked above all by the value they give to education (see e.g. Reher (2023)).

The most important constraints are the economic ones, and particularly the financial resources available to the local councils, which are responsible for public primary education during this period. According to the *Moyano Law* (1857), the funding of primary education fell to the municipalities, which meant in practice a highly decentralized system. Irrespective of different historical contexts and levels of advancement in the literacy process, the Moyano Law shared the decentralized approach to the funding of public elementary education with the relevant laws in other Western European countries: the *Allgemeines Landrecht* (1794) of Prussia<sup>33</sup>, the Guizot Law (1833) and the *Falloux Law* (1850) of France<sup>34</sup>, the *Casati Law* of Italy (1859)<sup>35</sup>.

In general, the resources available for primary education in Spain were conditioned throughout this period by the *desamortizaciones* of 1836-1837, 1841 and 1855 (see Bennassar (1985)), which confiscated the properties of the Catholic Church, the assets of educational foundations of all kinds, and most of the land belonging to the municipalities<sup>36</sup>. The impoverished Spanish local councils were saddled from 1857 with the obligation to support public primary education, in a context where educational charities were deprived of all their means.

Two literacy rates may have a *biological link*, in so far as they share part of their underlying populations<sup>37</sup>. Assuming that the acquisition of literacy for those older than a certain age becomes increasingly unlikely, a very low literacy rate in some year holds back the biologically linked literacy

<sup>33</sup>The previous Prussian *Generallandschulreglements* for Protestant schools (1763) and for Catholic schools (1765) were advanced for their time, but poorly enforced. In 1819/20 the (central) state participation in elementary education spending was 6.2 percent; it was 4.5 percent in 1861, when universal alphabetization of children had been reached already (see Zilch (2014)). Then the state participation began to grow (28.8% in 1911), especially when elementary education fees were eliminated in 1889.

<sup>34</sup>The Guizot Law required municipalities to establish elementary schools for boys, and the Falloux Law extended this obligation to schools for girls. See Diebolt et al. (2005) on the rise of mass schooling in France and its funding.

<sup>35</sup>See Cappelli and Quiroga (2020, 2021) and Bray (1991).

<sup>36</sup>The special public debt securities issued by the state as a partial compensation for the municipalities were nowhere near enough and became eventually worthless (the low yields were reduced even more and often not paid). See Moral Ruiz (1984) (p. 30-31, 106-107) and Comín (1996).

<sup>37</sup>For example, rate  $T_{10}$  in 1900 has a biological link with rate  $T_{10}$  in 1910, because the people 10 years old and over in 1900 still surviving in 1910 are part of the people 10 years and over in 1910. In contrast,  $T_{10}$  in 1900 has no biological link with  $T_{11-15}$  in 1910.

rates in later years<sup>38</sup>. In this connection, age-specific rates (whenever available) such as  $T_{11-15}$  may provide explanatory power.

There are three *res ipsa loquitur* features of Spanish male literacy in the period 1860-1910:

1. The global failure of the literacy process. The male literacy rate grew by only 18.7 percentage points over a 50-year period, from a comparatively modest 38.9% in 1860 to a comparatively low 57.6% in 1910. This slow growth occurred not only in the phase of political instability up to 1876 (the male literacy rate is still 43.5% in 1877), but also in the relatively stable span of the Bourbon Restoration (growth of 14.1 points in the 33 years between 1877 and 1910).
2. The large spatial differences. As we shall see later in more detail, the spatial structure of 1860 is maintained. The six provinces of the Castilian core are still the top six provinces by male literacy rate in 1910. The 19 provinces of the second ring around Burgos still coincide in 1910 with the 19 provinces with the highest male literacy rate (now with only one exception). All the 12 provinces of the fifth cluster are among the bottom 14 provinces by male literacy rate in 1910. The following table shows the evolution of male literacy in the five clusters<sup>39</sup> (for the last three censuses  $T_{10}$  is given in the top row and  $T_{11-15}$  in the bottom row; the rates higher than 75% are marked bold):

	1860	1877	1887	1900	1910
	Men	Men	Men	Men	Men
	$T_{10}$	$T_{10}$	$T_{10}, T_{11-15}$	$T_{10}, T_{11-15}$	$T_{10}, T_{11-15}$
Castilian Core	69.01	<b>76.63</b>	<b>80.84</b> <b>82.93</b>	<b>83.04</b> <b>82.43</b>	<b>87.93</b> <b>88.28</b>
Northern Plateau	56.57	64.54	69.39 66.94	74.58 71.17	<b>78.95</b> <b>76.41</b>
Sundry North	45.31	51.28	57.80 55.58	62.03 59.94	70.62 68.48
Transition	33.94	38.54	43.00 40.15	47.88 44.75	52.32 49.02
South and Est	23.68	25.67	30.18 24.22	34.38 28.89	38.91 31.56
SPAIN	38.90	43.51	48.18 44.33	52.69 49.11	57.55 53.43

Table 2: Spanish male literacy rates by cluster.

3. The appreciable percentage of men becoming literate after school age. Beginning in 1887, censuses provide data of literacy by age<sup>40</sup>. Despite the underlying trend of growing child literacy, in all censuses the maximum male literacy corresponds to the group of adults between 31 and 35 years old, with literacy rates approximately 10 points higher than those of boys between 11 and 15 years old<sup>41</sup>:

<sup>38</sup>For example, a very low  $T_{10}$  in 1900 makes impossible for  $T_{10}$  in 1910 to be very high.

<sup>39</sup>We can order the 1910 male literacy rates and construct 5 levels parallel to the 1860 clusters (i.e. the first 6 provinces would form level 1, as the first cluster has 6 elements, the next 8 would form level 2, as the second cluster has 8 elements, etc.). The 1910 level 1 provinces would be exactly those of the first 1860 cluster, and the changes between the other 1910 levels and their 1860 counterpart clusters would be minimal: 5 provinces would move up one level (all of them where Basque or Catalan-Valencian was spoken), and displace 5 others that would move down one level.

<sup>40</sup>The age intervals considered by the three censuses do not coincide, but a homogeneous series can be built with 5-year intervals for the ages between 11 and 50 years old. Gabriel (1997) conducts a similar study to ours for the ages between 6 and 30 years old, and in Gabriel (1998) 10-year intervals are considered for all ages from 11 years onwards. The conclusions are essentially in line with our own, both for males and females (see below).

<sup>41</sup>Cohort analysis filters out the trend of growing child literacy, even when it is still susceptible to other distortions: the cohorts decrease through mortality and change through migration, and both phenomena may affect the literate and the illiterate to different degrees (see Cipolla (1969)). At any rate, cohort analysis seems to confirm our conclusion. The cohort

Census	Men $T_{11-15}$	Men $T_{16-20}$	Men $T_{21-25}$	Men $T_{26-30}$	Men $T_{31-35}$	Men $T_{36-40}$	Men $T_{41-45}$	Men $T_{46-50}$	Men $T_{51-60}$	Men $T_{61-70}$
1887	44.33	49.99	52.06	51.40	54.31	49.82	51.27	47.16	45.65	42.33
1900	49.11	54.75	57.12	55.71	58.11	54.93	57.67	52.63	50.15	45.42
1910	53.43	59.89	62.40	61.07	63.28	59.64	61.70	57.16	56.23	50.01

Table 3: Spanish male literacy rates by age.<sup>42</sup>

As age increases, the significance of parents in the literacy process gives way to that of the concerned individual. The means to implement late literacy were varied. Village schools allowed the not-so-young to attend. Adult schools were segregated by sex<sup>43</sup> and in 1900 legislation was passed organizing night classes for workers, as an instrument to achieve “a solid knowledge leading to capable and intelligent workers and teachers, who contribute to the development and progress of the arts and industries of the country”<sup>44</sup>. It was also non-negligible the literacy work that the army carried out on recruits during their military service<sup>45</sup>.

Figure 2 considers the male child literacy  $T_{11-15}$  at the end of the period (data in Appendix 2). This map shows how varying were the deeds of local authorities and parents in different areas of the country. It is a snapshot of the advancement of the modernization process in 1910, as far as male literacy is concerned. Besides, the correlation coefficient between  $T_{10}$  and  $T_{11-15}$  in 1910 is  $\rho = 0.9782$ .

Now we are to establish a model of the dynamics of male literacy in the period 1860-1910. Firstly, we must remark the high correlation between male literacy  $T_6$  in 1860 and male child literacy  $T_{11-15}$  in 1910 (even though there is no biological link between the two rates): the correlation coefficient is  $\rho = 0.8824$ . The values of 1860 represent well the inherited historical substratum, particularly with regard to the mentality of the decision makers in each area of Spain. Secondly, during the period 1860-1910 Spain experienced social and economic changes, whose influence on the literacy process must be considered.

The linguistic factor may also be relevant in the literacy process. In several areas of Spain, the usual language of the majority of the population was not Spanish in 1860. Basque was spoken in three provinces (Vizcaya, Guipúzcoa and the north of the province of Navarra; by the mid-nineteenth century Basque was hardly used in Álava). Catalan-Valencian was spoken in 8 provinces (Barcelona, Gerona, Lérida, Tarragona, the Balearic Islands, Castellón and a large part of the provinces of Valencia and Alicante). Galician was spoken in four provinces (La Coruña, Pontevedra, Lugo and Orense). These languages have different degrees of closeness to Spanish. Basque is not even an Indo-European language. Catalan-Valencian is a Romance language and has a significant mutual intelligibility with Spanish in written form and has partial or low intelligibility in spoken form (the latter varies greatly according to dialect)<sup>46</sup>. Galician is very close to Spanish and both languages are mutually intelligible<sup>47</sup>. The Moyano Law of 1857, in force throughout the period, established the obligatory teaching

---

of those aged 11-15 in 1900 had a literacy rate, in that year, of 44.11, and in 1910 the members of that cohort recorded in the census (now aged 21-25) had a literacy rate of 62.40. This represents a variation of +18.29. An analogous calculation for the cohort aged 16-20 in 1900 gives a variation from 1900 to 1910 of +6.32. In turn, the variations from 1900 to 1910 for the 21-25, 26-30, 31-35 and 36-40 age cohorts in 1900 are, respectively: 6.16, 3.93, 3.59 and 2.23.

<sup>42</sup>The value of  $T_{15-15}$  is only available in 1887, and it is 46.96.

<sup>43</sup>Data of 1880 indicate that 91% of adult schools were for males and 94% of the students were male altogether (see Dirección General de Instrucción Pública (1883)).

<sup>44</sup>Real Decreto of 25 May 1900.

<sup>45</sup>See Quiroga (1999) and Dirección General del Instituto Geográfico y Estadístico (1914) (p. 366).

<sup>46</sup>See Juge (2007) for a general study of the distance of Catalan-Valencian to the other Romance languages.

<sup>47</sup>See Ramallo (2007) (p. 28); also Regueira (1999) about the closeness to Spanish of the different varieties of Galician.

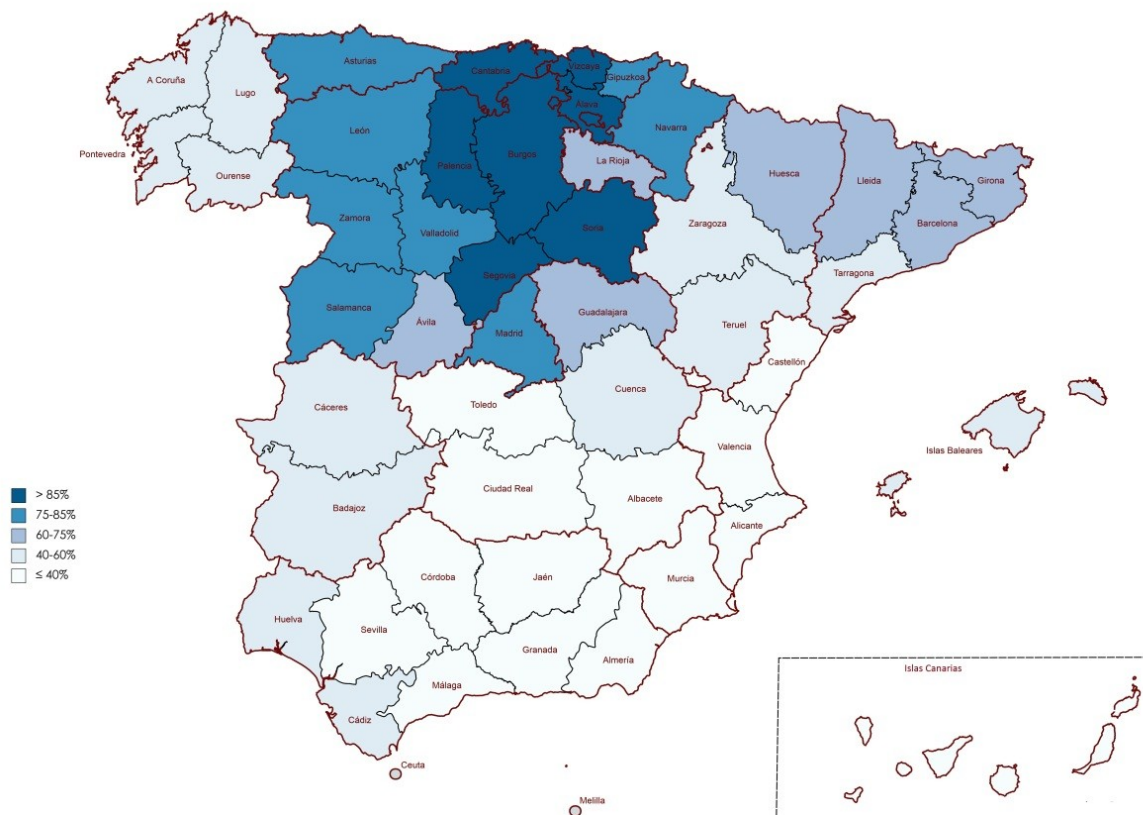


Figure 2: Literacy rates (males aged 11-15) in 1910.

of Spanish, but did not require it to be the compulsory language of instruction. In fact, it was not, in whole or in part, in quite a few municipalities of Spain, particularly in Catalonia<sup>48</sup>. In 1902, Spanish was imposed as the sole language of instruction by a decree<sup>49</sup>, which was to be made ineffective in practice by a ministerial decree one month later<sup>50</sup>.

The economic transformations, especially the industrialisation process, may affect literacy. Globally, the proportion of the agricultural active population to the total active population stayed at around 72% during the period (among men; it is difficult to estimate the composition of the female active population in predominantly agrarian economies)<sup>51</sup>. However, the process of establishing a modern industry advanced, with a marked tendency towards spatial concentration. Barcelona and Vizcaya became important industrial hubs<sup>52</sup>.

We propose the following regression model, with the variables running through the values of the provinces:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon$$

where  $Y$  is the male child literacy rate  $T_{11-15}$  in 1910 and  $\epsilon$  is the error term. The regressors are:

- $X_1$ : male literacy rate  $T_{10}$  in 1860.
- $X_2$ ,  $X_3$ , and  $X_4$ : dummy dichotomic variables, taking the value 1 if Basque (for  $X_2$ ), Catalan-Valencian (for  $X_3$ ), or Galician (for  $X_4$ ) is spoken in the province, and the value 0 otherwise.
- $X_5$ : proportion of the male active population working in agriculture (including forestry and fishing) in 1900<sup>53</sup>.

Table 4 presents the estimates of four models (the corresponding p-values are shown in parentheses). The first model is the initial one, with 5 regressors. The values of the coefficient of determination  $R^2$  and the adjusted coefficient of determination  $\bar{R}^2$  are high. The regressors  $X_4$  and  $X_5$  may be successively removed as not at all significant. In the resulting second model, the coefficient of determination is hardly altered in relation to the first model, and all the regressors are significant.

The estimated model we arrive at is

$$Y = 0.0034 + 1.2929X_1 + 0.2412X_2 + 0.1436X_3 + \epsilon$$

whose coefficient of determination is round 90%. On the other hand, the simple regression model

$$Y = 0.0811 + 1.1959X_1 + \epsilon$$

has coefficient of determination  $R^2 = 0.7786$ . Consequently, male child literacy in 1910 is to a large extent explained by male literacy 50 years earlier. Basque and Catalan-Valencian being spoken are also two significant variables, where the former is more influential.

<sup>48</sup>See Diario de las sesiones de Cortes (1896, 1902); González Ollé (1985); Gabriel (2019).

<sup>49</sup>Decree of 21/11/1902. The incumbent minister was the Count of Romanones. Shortly after the publication of the decree, Sagasta's Liberal government fell and was replaced by Silvela's Conservative government.

<sup>50</sup>Ministerial decree of 19/12/1902, signed by the new Minister of Public Instruction in Silvela's government, Manuel Allendesalazar. Part of Romanones's decree was declared void as being *contra legem*, and the rest was reinterpreted so as to make it unenforceable in the relevant cases.

<sup>51</sup>The available rates are 72.14% for 1877, 72.26% for 1887, 72.20% for 1900 and 71.64% for 1910 (see the corresponding censuses and Nicolau (2005)).

<sup>52</sup>The proportion of the male active population working in industry in 1900 was 31.86% in Barcelona and 37.06% in Vizcaya. On the other hand, this rate was only 6.74% in Burgos, the province with the highest male literacy in Spain.

<sup>53</sup>An alternative model has also been considered, with  $X_5$  standing for the proportion of the male active population working in industry (including mining, energy and construction) in 1900. There is no appreciable alteration in the results.

Regressors	$\beta_0$ intercept	$\beta_1$ M. 1860	$\beta_2$ Basque	$\beta_3$ Cat-V.	$\beta_4$ Galician	$\beta_5$ agric.	$R^2$	$\bar{R}^2$
$X_1, X_2, X_3, X_4, X_5$	0.0138 (0.883)	1.2915 (6E-22)	0.2387 (9E-6)	0.1423 (4E-5)	-0.0013 (0.974)	-0.0125 (0.909)	0.8979	0.8860
$X_1, X_2, X_3$	0.0034 (0.914)	1.2929 (4E-23)	0.2412 (7E-7)	0.1436 (9E-6)			0.8978	0.8910
$X_1, X_2$	0.0715 (0.046)	1.1867 (3E-19)	0.2181 (1E-4)				0.8407	0.8338
$X_1$	0.0811 (0.051)	1.1959 (5E-17)					0.7786	

Table 4: Regression models for male literacy.

Some explanations of the significance of these linguistic variables can be hypothesized. On the one hand, learning Spanish was perceived as increasingly important by local authorities and parents with the process of economic modernization, especially in the case of Basque; this learning went hand in hand with the acquisition of literacy. On the other hand, higher growth of literacy during the period for those not having Spanish as mother tongue might also be attributable to catching up from a relatively low level of literacy before 1860, *caeteris paribus*, even if the mother tongue was partially or totally the language of instruction. At any rate, further study of the influence of linguistic variables suggests itself, considering local data (e.g. on the language of instruction) and events such as the Carlist wars and the ensuing settlements. In contrast to political entities like Cisleithania (see Urbanitsch (2021)), France (see Furet and Ozouf (1977)) or Prussia (see Belzyt (1998)), not much attention has been paid to the influence of linguistic factors on the Spanish literacy process.

## 5 The evolution of female literacy

### 5.1 The starting point

Based on census data, we can estimate that the Spanish literacy rate for females (aged 10 and over) was 11.2% in 1860. Figure 3 displays the spatial distribution of female literacy in that year.

There is not as clear a spatial pattern for female literacy in 1860 as for male literacy. There are only 4 provinces above 20%. Half of the provinces have literacy rates that are concentrated between 7.1% and 11.1%. Nine provinces are below 7%, including the four provinces of Galicia.

Emigration to America is arguably the main reason for the low sex ratios (ratio of male to female for the population between 16 and 50 years old) in the Canary Islands and the provinces of the northwestern coast of Spain<sup>54</sup>. Here Galicia is included, but also Oviedo, Santander and Vizcaya<sup>55</sup>; the latter two provinces have relatively high female literacy rates (considering the dismal

<sup>54</sup>See Gozávez Pérez and Martín-Serrano Rodríguez (2016). Internal migrations were mostly within the bounds of each province during the period, with exception of the immigration to Madrid and Barcelona (see Nicolau (2005) and Juif and Quiroga (2019)).

<sup>55</sup>The lowest sex ratios in 1860 were the following: Pontevedra 65.76%, Oviedo 71.29%, Coruña 71.53%, Canarias 71.83%, Lugo 78.43%, Santander 79.36%, Orense 82.37%, Vizcaya 85.87%; the national average was 95.69%. In the 1910 census the six lowest ratios were as follows: Pontevedra 59.95%, Coruña 64.65%, Canarias 72.03%, Oviedo 76.45%, Orense 76.66%, Lugo 76.70%; the ratio of Santander was 83.36% and that of Vizcaya 92.44%; the national average was 91.69%.



background)<sup>56</sup>. At any rate, low sex ratios tend to impair the position of women, although other factors may countervail the effect<sup>57</sup>.

Contrary to what might be expected, there is not a high correlation between the male and female literacy rates in 1860, with  $\rho = 0.5377$ ,  $\rho^2 = 0.2891$  (we always consider disaggregation by province), a figure similar to that of the correlation between the female literacy rate and the urbanization rate ( $\rho = 0.5444$ ,  $\rho^2 = 0.2963$ ). The (squared) multiple correlation coefficient for the female literacy rate with respect to the male literacy rate and the urbanization rate is 0.5912.

There is hardly any correlation between female literacy and the level of industrialization, whether the latter is measured by the ratio of industrial workers to the total population ( $\rho = 0.1970$ ), or if both miners and industrial workers are in the numerator ( $\rho = 0.1490$ ).

## 5.2 Evolution in the period 1860-1910

In parallel to subsection 4.2, we consider these three features of the Spanish female literacy in the period 1860-1910:

1. As in the male case, the global failure of the literacy process. The female literacy rate grew by only 27.4 percentage points over a 50-year period, from a comparatively low 11.2% in 1860 to a comparatively very low 38.6% in 1910. This slow growth occurred not only in the phase of political instability up to 1876 (the female literacy rate is still 17.9% in 1877), but also in the relatively stable span of the Bourbon Restoration (growth of 20.7 points in the 33 years between 1877 and 1910).
2. The spatial distribution of female literacy gradually approaches along the period the pattern of male literacy in 1860. On the one hand, the correlation coefficients of female literacy (in 1877, 1887, 1900 and 1910) with female literacy in 1860 decrease continuously along the period (in 1877 is  $\rho = 0.9634$ , in 1910 is  $\rho = 0.7369$ ). This is not surprising. In contrast, the correlation coefficients of female literacy (in 1860, 1877, 1887, 1900 and 1910) with male literacy in 1860 increase steadily along the period (in 1860 is  $\rho = 0.5377$ , in 1877 is  $\rho = 0.6543$ , in 1910 is  $\rho = 0.8134$ ). All in all, the spatial distributions of female and male literacy become closer over time ( $\rho = 0.5377$  in 1860,  $\rho = 0.8815$  in 1910).

The clusters introduced in Section 4 continue to help us now. The following table, parallel to Table 2, shows the evolution of female literacy in the five clusters:

<sup>56</sup>Apart from the emigration to America, substantial in the aforementioned provinces, there was an important emigration from Almería to Argelia (the sex ratio of Almería was 87.38% in 1860 and 79.41% in 1910). It was predominantly a temporary migration, with most emigrants returning eventually (see Nicolau (2005)).

<sup>57</sup>Out-of-wedlock birth rates are to be interpreted very cautiously, and considering the society of 1860, but some figures are worth noticing (data of the *Anuario Estadístico de España 1860-1861*). Canarias (20.4%) and Lugo (18.4%) had the highest rates in the country, whereas the rates of Santander (4.0%), Vizcaya (2.4%) and Almería (3.4%) were below the national average (5.6%).

	1860	1877	1887	1900	1910
	Women $T_{10}$	Women $T_{10}$	Women $T_{10}, T_{11-15}$	Women $T_{10}, T_{11-15}$	Women $T_{10}, T_{11-15}$
Castilian Core	15.92	27.81	37.47 57.91	51.42 67.42	66.80 <b>82.06</b>
Northern Plateau	16.03	25.55	32.18 42.52	43.32 52.66	54.05 65.69
Sundry North	10.86	19.54	26.24 36.75	35.14 46.13	45.63 58.12
Transition	10.01	15.85	20.14 26.32	26.47 34.02	33.64 41.59
South and Est	8.55	12.37	15.30 17.18	20.36 23.17	25.20 26.10
SPAIN	11.14	17.86	22.84 29.58	30.54 38.07	38.55 45.90

Table 5: Spanish female literacy rates by cluster.

As an exception to the dismal global evolution during the period, the female literacy process was successful in the Castilian core. There was also some closing of the gap with male literacy in the North of Spain, especially in the Northern Plateau. Certainly, the low initial female literacy  $T_{10}$  rates “burden” the later  $T_{10}$  rates through biological link, but not the later  $T_{11-15}$  rates; it is thus disappointing that  $T_{11-15}$  is almost as bad as  $T_{10}$  in the South and East cluster in 1910.

A measure of the level of improvement of female literacy during the period is given by the difference between  $T_{11-15}$  in 1910 and  $T_{10}$  in 1860: 66.14 percentage points for the Castilian core, 49.67 for the Northern Plateau, 47.26 for Sundry North, 31.57 for Transition and 17.56 for South and East, with 34.75 for Spain overall.

3. Less women than men became literate after (extended) school age. In all censuses the maximum female literacy corresponds to the interval of those between 16 and 20 years old; in this interval literacy rates are little different from the rates of girls between 11 and 15 years old<sup>58</sup> (these results are to be interpreted considering the trend of growing child literacy)<sup>59</sup>:

Census	W. $T_{11-15}$	W. $T_{16-20}$	W. $T_{21-25}$	W. $T_{26-30}$	W. $T_{31-35}$	W. $T_{36-40}$	W. $T_{41-45}$	W. $T_{46-50}$	W. $T_{51-60}$	W. $T_{61-70}$
1887	29.58	30.73	29.10	25.91	25.50	21.01	19.56	15.96	13.82	11.84
1900	38.07	39.77	38.17	33.88	33.42	28.80	28.71	24.13	19.98	16.10
1910	45.90	48.40	47.31	42.78	42.17	37.11	37.10	31.65	28.36	21.86

Table 6: Spanish female literacy rates by age.

Figure 4 displays the female child literacy  $T_{11-15}$  in 1910 (data in Appendix 2)<sup>60</sup>. This map shows the spatial distribution of a most relevant indicator of cultural modernization. Out of

<sup>58</sup>In 1887, disaggregated data for the 16-20 interval are available, and thus  $T_{16-16} = 30.78$ ,  $T_{17-17} = 32.13$ ,  $T_{18-18} = 30.52$ ,  $T_{19-19} = 31.96$  and  $T_{20-20} = 28.79$ .

<sup>59</sup>Cohort analysis seems to confirm our conclusion, comparing with the corresponding data for men (see above). The cohort of those aged 11 – 15 in 1900 had a literacy rate, in that year, of 38.07, and in 1910 the members of that cohort recorded in the census (now aged 21 – 25) had a literacy rate of 47.31. This represents a variation of +9.24. An analogous calculation for the cohort aged 16-20 in 1900 gives a variation from 1900 to 1910 of +3.01. In turn, the variations from 1900 to 1910 for the 21 – 25, 26 – 30, 31 – 35 and 36 – 40 age cohorts in 1900 are, respectively: 4.00, 3.23, 3.68 and 2.85.

<sup>60</sup>The correlation coefficient between  $T_{10}$  and  $T_{11-15}$  in 1910 is  $\rho = 0.9543$ .

49 provinces, there are 9 provinces above the threshold of 75%, 15 provinces between 50% and 70%, and 25 provinces under 44%.

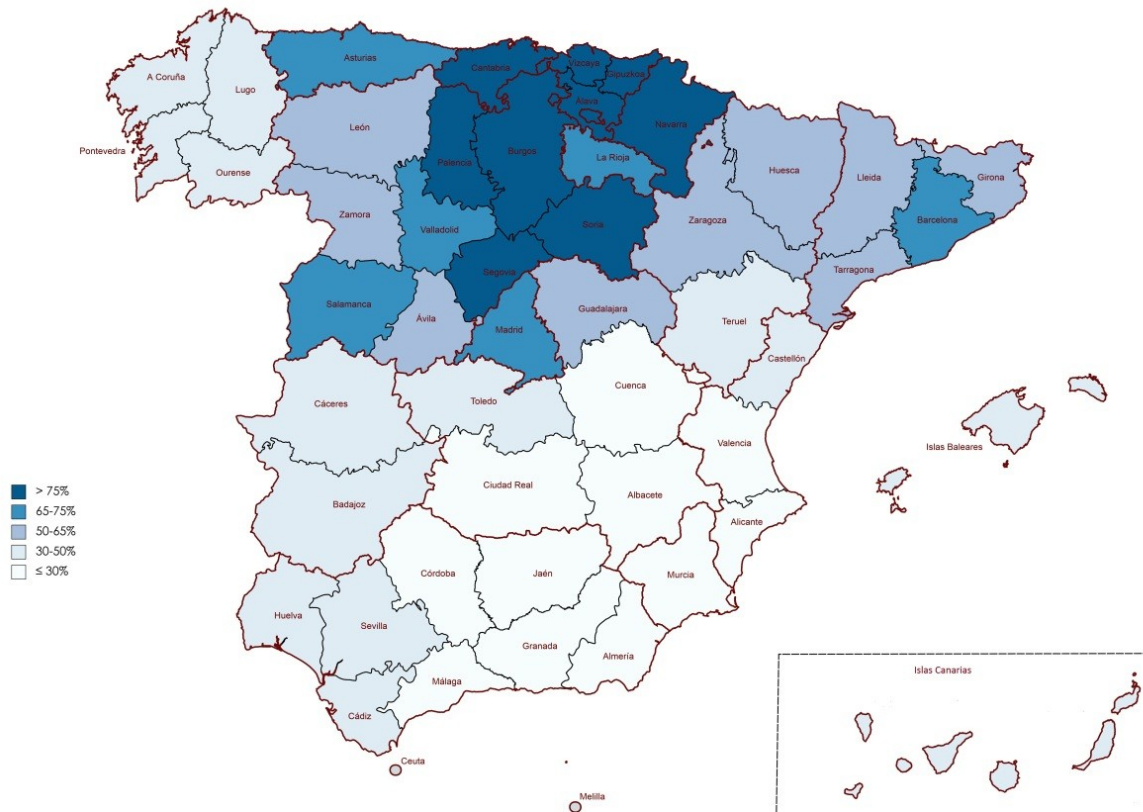


Figure 4: Literacy rates (females aged 11-15) in 1910.

Next, we establish a model of the dynamics of female literacy in the period 1860-1910. The same regressors as for male literacy (in Section 4) are considered, with the addition of a further regressor representing female literacy in 1860.

We propose the following regression model, with the variables running through the values of the provinces:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon$$

where  $Y$  is now the female child literacy rate  $T_{11-15}$  in 1910,  $\epsilon$  is the error term and  $X_1, X_2, X_3, X_4, X_5$  are as in Section 4. We incorporate the regressor:

$X_6$  : female literacy rate  $T_{10}$  in 1860.

Table 7 presents the estimates of six models (the corresponding  $p$ -values are shown in parentheses). The first model is the initial one, with 6 regressors<sup>61</sup>. The values of the coefficient of

<sup>61</sup>An alternative initial model has also been considered, with  $X_5$  standing for the proportion of the male active population working in industry (including mining, energy and construction) in 1900. There is no appreciable alteration in the results.

determination  $R^2$  and the adjusted coefficient of determination  $\bar{R}^2$  are high. The regressors  $X_6$  and  $X_5$  may be successively removed as non-significant. It is remarkable the non-significance of  $X_6$ , in view of the very strong significance of  $X_1$ . In the resulting model with 4 regressors, the coefficient of determination is little altered in relation to the first model. All the regressors are strongly significant, except for  $X_4$  (Galician), whose  $p$ -value is between 0.01 and 0.05; on the other hand, the estimated coefficient of  $X_4$  is negative, in contrast to the other linguistic regressors.

Regressors	$\beta_0$ intercept	$\beta_1$ M. 1860	$\beta_2$ Basque	$\beta_3$ Cat-V.	$\beta_4$ Galician	$\beta_5$ agric.	$\beta_6$ F. 1860	$R^2$	$\bar{R}^2$
$X_1, X_2, X_3, X_4, X_5, X_6$	0.2620 (0.13)	1.2651 (7E-16)	0.2602 (9E-6)	0.0982 (0.014)	-0.1063 (0.026)	-0.3361 (0.086)	-0.4891 (0.241)	0.8815	0.8646
$X_1, X_2, X_3, X_4, X_5$	0.0963 (0.347)	1.1872 (3E-19)	0.2647 (6E-6)	0.1207 (8E-4)	-0.0831 (0.054)	-0.1562 (0.190)		0.8775	0.8633
$X_1, X_2, X_3, X_4$	-0.0301 (0.400)	1.2029 (9E-20)	0.2936 (2E-7)	0.1334 (2E-4)	-0.0954 (0.026)			0.8725	0.8609
$X_1, X_2, X_3$	-0.0434 (0.241)	1.2105 (2E-19)	0.3037 (2E-7)	0.1444 (9E-5)				0.8570	0.8475
$X_1, X_2$	0.0251 (0.521)	1.1037 (3E-16)	0.2805 (1E-5)					0.7980	0.7892
$X_1$	0.0374 (0.432)	1.1155 (2E-13)						0.6929	

Table 7: Regression models for female literacy.

The estimated model we arrive at is

$$Y = -0.0301 + 1.2029X_1 + 0.2936X_2 + 0.1334X_3 - 0.0954X_4 + \epsilon$$

whose coefficient of determination is round 87%. On the other hand, the simple regression model

$$Y = 0.0374 + 1.1155X_1 + \epsilon$$

has coefficient of determination round 70%. Consequently, female child literacy in 1910 is to a considerable extent explained by male literacy 50 years earlier. The values of male literacy of 1860 represent well the inherited historical substratum. Catalan-Valencian being spoken is also a relevant variable, and Basque being spoken is very relevant.

## 6 Conclusions: towards an explanatory model of the evolution of Spanish literacy in the period 1860-1910.

During the second half of the 19th century, Spain was characterized by slow progress in literacy rates, with much worse results than other nearby European countries such as Italy, which had overcome Spain in this regard by 1880. A partial explanation for this poor performance is that the (central) state confiscated most of the land belonging to the municipalities by the *desamortización* of 1855. Yet the funding of public primary education was assigned to these municipalities by the Moyano Law of 1857. The lack of resources of Spanish local councils meant that aggregate spending on primary

education remained stagnant, while the school population grew. In contrast, Italy, which also had a decentralized system, increased the spending significantly<sup>62</sup>.

As for male literacy, despite the slow progress in the country as a whole, there were great spatial disparities that remained essentially stable throughout the period, only with exceptions related to minority languages. In general terms, during these fifty years, provinces increased their rates by around 20 percentage points and thus those starting at the first positions reached almost complete universal male literacy by 1910, whereas some backward provinces ended up below 40% male literacy, the minimum threshold for sustained economic development to begin. The 6 top provinces in 1860 (still heading the list in 1910) coincided with the “Castilian core” of the country. Some explanations of their high male literacy values can be traced back to historical processes of medieval origin related to the *Reconquista*, including the model of land distribution and mental processes of imitation of the behaviour of the nobility and the clergy as means of social advancement, involving a high social valuation of education (see Pérez Moreda (1997)); at any rate these provinces tended to be those devoting more public resources to primary education.

Figure 5 shows the per capita public expenditure on primary education at the end of the period (1908). Certainly, the commitment of local authorities to popular education remained uneven across the country<sup>63</sup>. Rich Madrid and Barcelona were not at the top, but rural high literacy provinces.

In addition to this geographic disparity, Spain had large differences between male and female literacy rates. In 1860, just over 10% of women over the age of 10 could read and write compared to almost 40% of men. A partial explanation of this difference might be economic. The return on investment in education depended on its degree of use: if parents anticipated a different participation in the labour market according to sex, they would allocate family resources (including time) in a biased way; therefore, the family unit would be maximizing the return on its investment by having boys more educated than girls. All the same, the literacy gender gap was very large in comparative West European terms (see Section 2), especially in the more literate provinces, where it could exceed 50 percentage points (obviously, it was smaller in the little literate provinces). The abysmal literacy gender gap in the most literate Spain at the beginning of the period remains a puzzle, only to be addressed by analysing data prior to 1860.

Although the improvement of female literacy was globally less than mediocre during the period, female literacy somewhat reduced distances with male literacy, certainly in an uneven way. The literacy gender gap remained stable in the low literacy provinces. In contrast, female literacy grew steeply in the high literacy provinces, and so in 1910 girls were almost universally literate in the Castilian core. Statistical data show a lack of correlation between female literacy rates in 1910 and 1860 but a very high correlation between female literacy rates in 1910 and male literacy rates in 1860. At last literate fathers were willing to assume the economic cost of providing literacy to their daughters.

Certainly, the institutional framework eased the efforts of the overburdened small villages of rural Spain to fund primary education. The Moyano Law allowed the existence of mixed-sex “incomplete schools” and “seasonal schools” in villages with less than 500 inhabitants, and of incomplete schools for girls in villages with less than 2000 inhabitants. These schools had fewer subjects and less paid (and qualified) teachers, thus lowering the costs in education and making cheaper the extension of literacy to girls in small municipalities<sup>65</sup>. On the other hand, in small villages the school routine

<sup>62</sup>Cappelli and Quiroga (2020), Figure 3.

<sup>63</sup>Franchise in local elections was mostly selective (property based) until 1890 and universal (for males) afterwards. Suffrage was already universal after 1868 in villages under 100 inhabitants.

<sup>64</sup>Source: Dirección General del Instituto Geográfico y Estadístico (1913).

<sup>65</sup>Prima facie at the expense of quality, but the positive action of the bandwagon effect cannot be overlooked.

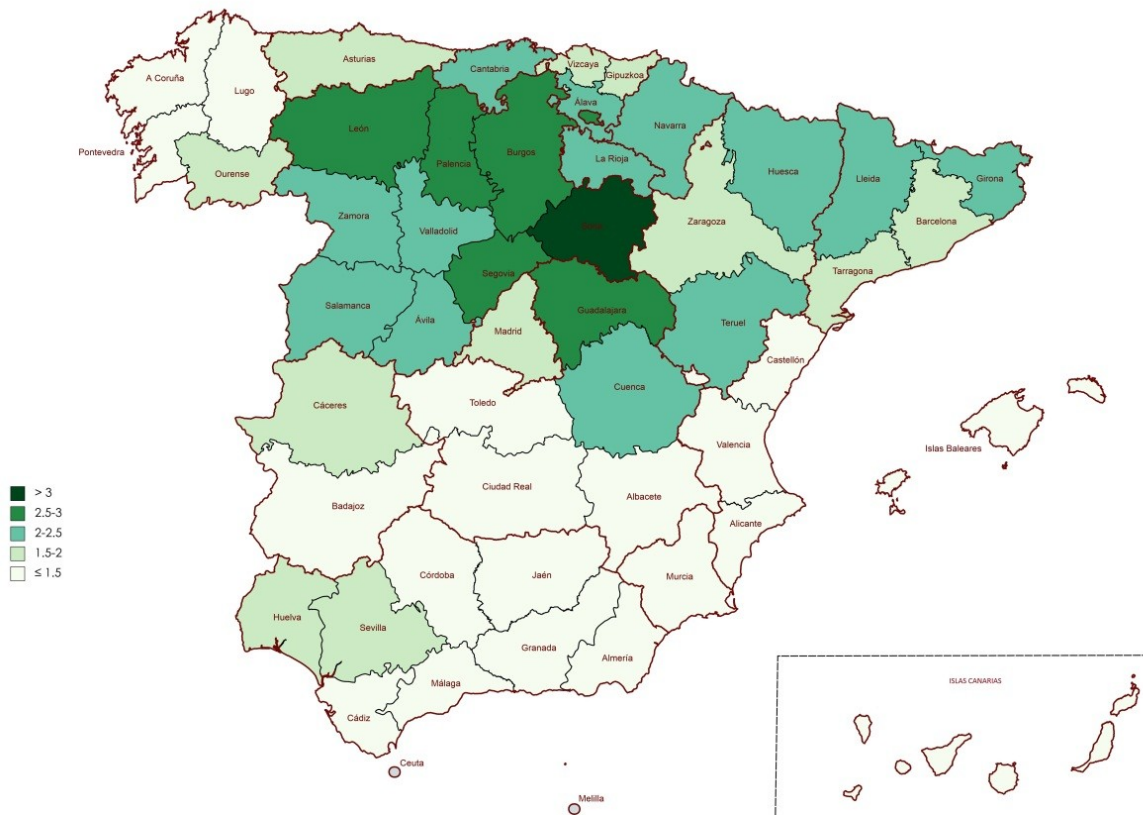


Figure 5: Per capita public expenditure in elementary education in 1908 (current pesetas).<sup>64</sup>

could be adapted to the seasonal work of children in agriculture, which was seen as an advantage by families; this was even more the case with seasonal schools<sup>66</sup>. As an example, in 1860, more than half of the population of Soria, one of the provinces with the highest literacy both for men in 1860 and for girls in 1910, lived in municipalities with less than 500 inhabitants, and over 70% of the schools were mixed-sex all along the period 1860-1910. At any rate, the local councils of very rural Soria were the top investors in public primary education in the entire country (data of 1908)<sup>67</sup>.

The per capita public expenditure in elementary education in 1908 and the female child literacy  $T_{11-15}$  in 1910 are highly correlated: the correlation coefficient is  $\rho = 0.7515$ <sup>68</sup>. Figure 6 displays the corresponding scatter plot and regression line.

According to the legislation in force during the period, public primary education was free only for the certified poor. At any rate, the literacy gender gap was not only attributable to parents and local councils. As shown in Table 3 and Table 6, the maximum literacy of women was reached in the interval between 16 and 20 years old, whereas the peak in male literacy was in the interval between 31 and 35 years old. The requirements of the labour market provided inducements for men (more than for women) to acquire literacy (e.g., through adult schools, schools for workers or while doing military service). Available technology, economic backwardness and blatant prejudice limited the quantity and quality of jobs accessible to women.

It is to be pointed out that birth rates and literacy rates were uncorrelated in Spain, both in 1860 and 1910. Moreover, there was no correlation between birth rates and literacy rates for children (data of 1910)<sup>69</sup>. There was a positive correlation between literacy rates and the intensity of (Catholic) confessional allegiance (see Gutiérrez (2025)).

The evolution of literacy in Spain between 1860 and 1910 did not follow the spatial pattern of the economic modernization process. The high literacy rates of the Castilian core did not correspond to the relative income levels of the area at that time<sup>70</sup>. The most literate provinces were neither the most industrialized, nor the most urbanized<sup>71</sup>. The majority of them were agrarian provinces, with a predominance of small and medium property, and with the population living in small villages.

All in all, the dynamics of literacy during the period are predominantly explained by the initial *male* literacy (the initial female literacy turns out to be non-significant). To a lesser extent, Basque and Catalan-Valencian being spoken are also two significant variables, where the former is more influential. The corresponding parsimonious statistical models (only with three regressors) have

<sup>66</sup>Palencia, a province with high literacy, had the highest concentration of seasonal schools in Spain (see Núñez (1992) (p. 269)).

<sup>67</sup>Soria had an urbanization rate (percentage of the population living in the provincial capital or in towns with more than 30,000 inhabitants) of 4.75%, compared to 69.65% of Madrid or 50.54% of Barcelona (data of 1900). The financial effort of the municipalities of Soria has to be appreciated gauging the extent of the confiscation of municipal land by the *desamortización* of 1855 (see Marín Gutiérrez (2015)).

<sup>68</sup>The correlation for boys is a little higher than that for girls: the correlation coefficient between per capita public expenditure in elementary education in 1908 and the male child literacy  $T_{11-15}$  in 1910 is  $\rho = 0.7839$ .

<sup>69</sup>In 1860 the correlation coefficients were  $\rho = -0.1321$ , between birth rates and male literacy rates, and  $\rho = 0.0915$ , between birth rates and female literacy rates. In 1910 the correlation coefficients were  $\rho = 0.0519$  and  $\rho = 0.0027$ , respectively; as for children, the correlation coefficients between birth rates and  $T_{11-15}$  were  $\rho = -0.0544$  for boys and  $\rho = -0.0659$  for girls. The sources for the birth dates are Junta General de Estadística (1861-1862) for 1860 (baptisms are taken instead of births), and Dirección General del Instituto Geográfico y Estadístico (1916) for 1910.

<sup>70</sup>There are no estimations of the GDP of this period disaggregated by province, but only by (present) region, and the latter estimations can be considered tentative. The majority of the provinces of the Castilian core and half of the provinces of the second cluster (Northern Plateau) are in the present region of Castilla y León. In 1900, the GDP per capita of Castilla y León was 91.2% of the Spanish average (see Álvarez Llano (1986) and Carreras et al. (2005)).

<sup>71</sup>In 1910 there was essentially no correlation between urbanization and literacy: the correlation coefficients were  $\rho = -0.0030$ , between urbanization rates and male literacy rates, and  $\rho = 0.2065$ , between urbanization rates and female literacy rates.

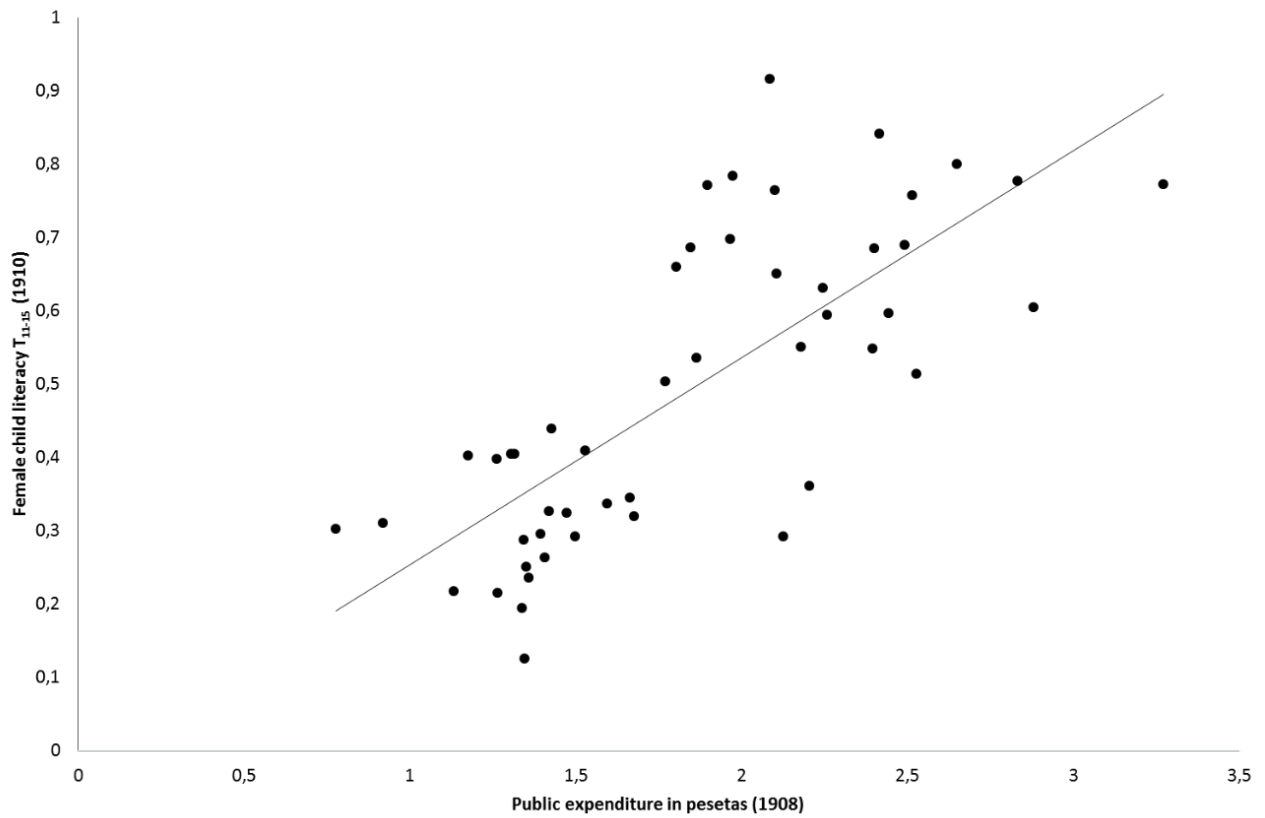


Figure 6: Public expenditure in elementary education in 1908 and female child literacy  $T_{11-15}$  in 1910.

coefficients of determination of round 90% for men and 86% for women. At any rate, further study, considering local data, of the influence of linguistic variables suggests itself.

## References

### Censuses

Junta General de Estadística: *Censo de la población de España, según el recuento verificado en 25 de diciembre de 1860 por la Junta General de Estadística*. Imprenta Nacional. Madrid (1863).

Dirección General del Instituto Geográfico y Estadístico: *Censo de la población de España, según el empadronamiento hecho en 31 de diciembre de 1877 por la Dirección General del Instituto Geográfico y Estadístico*. Imprenta de la Dirección General del Instituto Geográfico y Estadístico. Madrid (Tomo I, 1883; Tomo II, 1884).

Dirección General del Instituto Geográfico y Estadístico: *Censo de la población de España según el empadronamiento hecho en 31 de diciembre de 1887 por la Dirección General del Instituto Geográfico y Estadístico*. Imprenta de la Dirección General del Instituto Geográfico y Estadístico. Madrid (Tomo I, 1891; Tomo II, 1892).

Dirección General del Instituto Geográfico y Estadístico: *Censo de la población de España según el empadronamiento hecho en la península e islas adyacentes el 31 de diciembre de 1900*. Imprenta de la Dirección General del Instituto Geográfico y Estadístico. Madrid (Tomo I, 1902; Tomo II, 1903; Tomo III, 1907; Tomo IV, 1907).

Dirección General del Instituto Geográfico y Estadístico: *Censo de la población de España según el empadronamiento hecho en la península e islas adyacentes el 31 de diciembre de 1910*. Imprenta de la Dirección General del Instituto Geográfico y Estadístico. Madrid (Tomo I, 1913; Tomo II, 1916; Tomo III, 1917; Tomo IV, 1919).

### Other statistical sources

Dirección General de Instrucción Pública: *Estadística General de Primera Enseñanza correspondiente al decenio que terminó el 31 de diciembre de 1880*. Imprenta y Fundición de M. Tello. Madrid (1883).

Dirección General del Instituto Geográfico y Estadístico: *Anuario Estadístico de España. Año 1912*. Imprenta de la Dirección General del Instituto Geográfico y Estadístico. Madrid (1913).

Dirección General del Instituto Geográfico y Estadístico: *Reseña Geográfica y Estadística de España. Tomo III*. Talleres del Instituto Geográfico y Estadístico. Madrid (1914).

Dirección General del Instituto Geográfico y Estadístico: *Movimiento Natural de la Población de España. Año 1910*. Talleres del Instituto Geográfico y Estadístico. Madrid (1916).

Junta General de Estadística: *Anuario Estadístico de España 1860–1861*. Imprenta Nacional. Madrid (1861-1862).

## Further References

- Álvarez-Llano, R. (1986): “Evolución de la estructura económica regional de España en la historia: una aproximación”. *Situación*, 1, p. 5–61.
- Baffour, B.; King, T.; Valente, P. (2013): “The Modern Census: Evolution, Examples and Evaluation”. *International Statistical Review*, 81-3, p. 407–425.
- Beltrán Tapia, F. J., Díez-Minguela, A., Martínez-Galarraga, J. & Tirado-Fabregat, D. A. (2021): “The uneven transition towards universal literacy in Spain, 1860–1930”. *History of Education*, 50(5), p. 605–627.
- Beltrán Tapia, F.J.; Martínez-Galarraga, J. (2018): “Inequality and education in pre-industrial economies: Evidence from Spain”. *Explorations in Economic History*, 69, p. 81–101.
- Bennassar, B. (1985): “Las resistencias mentales”. In: B. Bennassar et al., *Orígenes del atraso económico español*, p. 147–163. Ariel. Barcelona.
- Belzyt, L. (1998): *Sprachliche Minderheiten im preußischen Staat 1815–1914. Die preußische Sprachenstatistik in Bearbeitung und Kommentar*. Verlag Herder-Institut. Marburg.
- Bowman, M.J.; Anderson, C.A. (1963): “Concerning the role of education in development”. In: Geertz, C. (ed.), *Old Societies and New States*, p. 247–263. The Free Press. Glencoe, Illinois.
- Bray, M. (1991): “Centralization versus decentralization in educational administration: regional issues”. *Educational Policy*, 5, p. 371–385.
- Cappelli, G.; Quiroga, G. (2020): “Literacy and schooling in Italy and Spain (1860–1921): a comparative analysis”. *Rivista di storia economica*, XXXVI-1, p. 87–123.
- Cappelli, G.; Quiroga, G. (2021): “Female teachers and the rise of primary education in Italy and Spain, 1861–1921: evidence from a new dataset”. *Economic History Review*, 74-3, p. 754–783.
- Cappelli, G.; Vasta, M. (2020): “Can school centralization foster human capital accumulation? A quasi-experiment from early twentieth-century Italy”. *Economic History Review*, 73, p. 159–184.
- Carreras, A.; Prados de la Escosura, L.; Rosés, J.R. (2005): “Renta y Riqueza”. In: Carreras, A. and Tafunell, X. (eds.), *Estadísticas Históricas de España: Siglos XIX–XX*, 2nd edition, p. 1297–1376. Fundación BBVA. Bilbao.
- Cipolla, C.M. (1969): *Literacy and Development in the West*. Penguin Books. Harmondsworth.
- Comín, F. (1996): *Historia de la Hacienda Pública, II. España (1808–1995)*. Crítica. Barcelona.
- Cusidó i Vallverdú, T.A.; Gil-Alonso, F. (2012): “Los censos en España: entre continuidad y cambio (1857–1970)”. *Revista de Demografía Histórica*, 30-1, p. 29–67.

- Diario de las sesiones de Cortes (1896): *Congreso de los Diputados. Sesión del viernes 19 de agosto de 1896*. Madrid.
- Diario de las sesiones de Cortes (1902): *Congreso de los Diputados. Sesión del lunes 24 de noviembre de 1902*. Madrid.
- Diebolt, C.; Jaoul, M.; San Martino, G. (2005): "Le mythe de Ferry: une analyse cliométrique". *Revue d'économie politique*, 115-4, p. 471–497.
- Flora, P.; Alber, J.; Eichenberg, R.; Kohl, J.; Kraus, F.; Pfenning, W.; Seebohm, K. (1983): *State, Economy, and Society in Western Europe 1815–1975. A Data Handbook in Two Volumes. Volume I*. Campus Verlag. Frankfurt.
- Furet, F.; Ozouf, J. (1977): *Lire et écrire*. Les Éditions de Minuit. Paris.
- Furet, F.; Sachs, W. (1974): "La croissance de l'alphabétisation en France (XVIIIe–XIXe siècle)". *Annales. Économies, Sociétés, Civilisations*, 29-3, p. 714–737.
- Gabriel, N. de (1997): "Alfabetización y escolarización en España (1897–1950)". *Revista de Educación*, 314, p. 217–243.
- Gabriel, N. de (1998): "Literacy, Age, Period and Cohort in Spain (1900–1950)". *Paedagogica Historica*, 34-1, p. 29–62.
- Gabriel, N. de (2019): "Nacionalismos y educación en España". *Historia de la Educación*, 37, p. 115–144.
- Galor, O. (2011): *Unified Growth Theory*. Princeton University Press. Princeton.
- Gomes, P.; Machado, M.P. (2020): "Literacy and primary school expansion in Portugal: 1940–62". *Revista de Historia Económica - Journal of Iberian and Latin American Economic History*, 38, p. 111–145.
- González Ollé, F. (1985): "El uso de las lenguas regionales en la enseñanza primaria ante el Congreso de los Diputados (1896)". *Boletín de la Real Academia Española*, 65, p. 345–355.
- Gozálvez Pérez, V.; Martín Serrano Rodríguez, G. (2016): "El censo de la población de España de 1860. Problemas metodológicos. Inicio de la aportación social en los censos". *Boletín de la Asociación de Geógrafos Españoles*, 70, p. 329–370.
- Gutiérrez, J.M. (2024): "Census-based comparability of data on literacy processes in Western Europe". *Spanish Journal of Statistics*, 6, p. 39–59.
- Gutiérrez, J.M. (2025): "A proxy variable built from a Kulturkampf". *Boletín de Estadística e Investigación Operativa*, 41, p. 47–53.
- Hanushek, E.A.; Woessmann, L. (2010): "Education and economic growth". In: Baker, E.; McGaw, B. and Peterson, P. (eds.), *International Encyclopedia of Education*, p. 245–252. Elsevier. Oxford.
- Johansson, E. (1977): *The History of Literacy in Sweden, in comparison with some other countries*. Umeå: Umeå University and School of Education, Educational Reports Umeå, 12. Partially reprinted as "The History of Literacy in Sweden", in H.J. Graff, A. Mackinnon, B. Sandin and I. Winchester (eds.), *Understanding Literacy in its Historical Contexts*, p. 28–57. Nordic Academic Press. Lund (2009).

- Johnson, R.A.; Wichern, D.W. (1998): *Applied Multivariate Statistical Analysis*, 4th edition. Prentice Hall. Upper Saddle River, NJ.
- Juge, M.L. (2007): "Catalan's Place in Romance Revisited". *Catalan Review*, 21, p. 257–277.
- Juif, D.; Quiroga, G. (2019): "Do you have to be tall and educated to be a migrant? Evidence from Spanish recruitment records, 1890–1950". *Economics and Human Biology*, 34, p. 115–124.
- Larquié, C. (1981): "L'alphabétisation à Madrid en 1650". *Revue d'Histoire Moderne et Contemporaine*, 28-1, p. 132–157.
- Marín Gutiérrez, A. (2015): *La desamortización forestal en la provincia de Soria. La génesis de los "montes de socios"*. Diputación Provincial de Soria. Soria.
- Melón, A. (1951): "Los censos de la población en España (1857–1940)". *Estudios Geográficos*, Vol. 12, No. 43, p. 203–281.
- Mitch, D. (1993): "Educación y crecimiento económico: ¿Otro axioma de indispensabilidad? Del capital humano a las capacidades humanas". In: Núñez, C.E. and Tortella, G. (eds.), *La maldición divina. Ignorancia y atraso en perspectiva histórica*, p. 41–60. Alianza Editorial. Madrid.
- Mitch, D. (2013): "The economic history of education". In: Parker, R.E. and Whaples, R.M. (eds.), *Routledge handbook of modern economic history*, p. 247–264. Routledge. London and New York.
- Moral Ruiz, J. del (1984): *Hacienda central y haciendas locales en España, 1845-1905*. Instituto de Estudios de Administración Local. Madrid.
- Myllyntaus, T. (1990): "Education in the Making of Modern Finland". In: Tortella, G. (ed.), *Education and Economic Development Since the Industrial Revolution*, p. 153–171. Generalitat Valenciana. Valencia.
- Nicolau, R. (2005): "Población, salud y actividad". In: Carreras, A. and Tafunell, X. (eds.), *Estadísticas Históricas de España: Siglos XIX–XX*, 2nd edition, p. 77–154. Fundación BBVA. Bilbao.
- Noble, F. (1965): "Istruzione scolastica". In: *Sviluppo della popolazione italiana dal 1861 al 1961*, Annali di Statistica, Serie VIII, Vol. 17, p. 295–317. Istituto Centrale di Statistica. Roma.
- Nunes, A.B. (2003): "Government Expenditure on Education, Economic Growth and Long Waves: The Case of Portugal". *Paedagogica Historica*, 39-5, p. 559–581.
- Núñez, C.E. (1992): *La fuente de la riqueza. Educación y desarrollo económico en la España contemporánea*. Alianza Editorial. Madrid.
- Núñez, C.E. (1993): "Alfabetización y desarrollo económico en España: una visión a largo plazo". In: Núñez, C.E. and Tortella, G. (eds.), *La maldición divina. Ignorancia y atraso en perspectiva histórica*, p. 223–236. Alianza Editorial. Madrid.
- Núñez, C.E. (1997): "La educación como fuente de crecimiento". *Papeles de Economía Española*, 73, p. 213–242.
- Núñez, C.E. (2003a): "Literacy, Schooling and Economic Modernization: A Historian Approach". *Paedagogica Historica*, 39-5, p. 535–558.

- Núñez, C.E. (2003b): "Within the European Periphery: education and labor mobility in twentieth-century Spain". *Paedagogica Historica*, 39-5, p. 621–649.
- Núñez, C.E. (2010): "Sobre la escasez de capital social fijo y humano en la España contemporánea". In: Morilla, J. et al. (eds.), *Las claves del desarrollo económico y social. Ensayos en homenaje a Gabriel Tortella*, p. 241–270. LID Editorial - Universidad de Alcalá. Madrid.
- Pérez Moreda, V. (1997): "El proceso de alfabetización y la formación de capital humano en España". *Papeles de Economía Española*, 73, p. 243–253.
- Pleijt, A. de; Nuvolari, A.; Weisdorf, J. (2020): "Human capital formation during the first industrial revolution: evidence from the use of steam engines". *Journal of the European Economic Association*, 18-2, p. 829–889.
- Quiroga, G. (1999): *Yo aprendí a leer en la Mili. El papel alfabetizador del ejército de tierra español, 1893-1954*. Ministerio de Defensa. Madrid.
- Ramallo, F. (2007): "Sociolinguistics of Spanish in Galicia". *International Journal of the Sociology of Language*, 184, p. 21–36.
- Regueira, X.L. (1999): "Estándar oral e variación social da lingua galega". In: Álvarez, R. and Vilavedra, D. (eds.), *Cinguidos por unha arela común. Homenaxe ó profesor Xesús Alonso Montero*, 855–875. Universidade de Santiago. Santiago de Compostela.
- Reher, D. (1997): "La teoría del capital humano y las realidades de la Historia". *Papeles de Economía Española*, 73, p. 254–261.
- Reher, D.S. (2023): "Patterns of literacy in historic Spain: An interpretation". *Revista de Historia Económica / Journal of Iberian and Latin American Economic History*, 41-1, p. 83–117.
- Reis, J. (1993): "El analfabetismo en Portugal en el siglo XIX: una interpretación". In: Núñez, C.E. and Tortella, G. (eds.), *La maldición divina. Ignorancia y atraso en perspectiva histórica*, p.237–269. Alianza Editorial. Madrid.
- Rodríguez, M.C.; Bennassar, B. (1978): "Signatures et niveau culturel des témoins et accusés dans le procès d'Inquisition du ressort du Tribunal de Tolède (1525-1817) et du ressort du Tribunal de Cordoue (1595-1632)". *Cahiers du monde hispanique et lusobrasílienne*, 31, p. 17–46.
- Sandberg, L.G. (1982): "Ignorance, Poverty and Economic Backwardness in the Early Stages of European Industrialization: Variations on Alexander Gerschenkron's Grand Theme". *Journal of European Economic History*, 11, p. 675–698.
- Sarasúa, C. (2002): "El acceso de niñas y niños a los recursos educativos en la España rural del siglo XIX". In: Martínez Carrión, J.M. (ed.), *El nivel de vida en la España rural, siglos XVIII-XX*, p. 549–609. Universidad de Alicante. Alicante.
- Sarasúa, C. (2019): "Women's work and structural change: occupational structure in eighteenth-century Spain". *The Economic History Review*, 72-2, p. 481–509.
- Schofield, R. (1973): "Some dimensions of illiteracy, 1750-1850". *Explorations in Economic History*, 10-4, p. 437–454.

- Smith, A. (1976 [1776]): *An Inquire into the Nature and Causes of the Wealth of Nations*, (2 vol.). Clarendon Press. Oxford.
- Terrón Bañuelos, A. (1997): "La modernización de la educación en España (1900-1939)". In: Escolano Benito, A. and Fernandes, R. (eds.), *Los caminos hacia la modernidad educativa en España y Portugal (1800-1975)*, p. 101–121. Fundación Rei Afonso Henriques. Zamora.
- Tortella, G. (1994): *El desarrollo de la España contemporánea. Historia económica de los siglos XIX y XX*. Alianza Editorial. Madrid.
- Tveit, K. (1991): "The Development of Popular Literacy in the Nordic Countries. A Comparative Historical Study". *Scandinavian Journal of Educational Research*, 35-4, p. 241–252.
- UNESCO (1953): *Progress of Literacy in Various Countries*. United Nations Educational, Scientific and Cultural Organization. Paris.
- UNESCO (1957): *World Illiteracy at Mid-Century*. United Nations Educational, Scientific and Cultural Organization. Paris.
- Urbanitsch, P. (2021): "Bildung und Bildungsinstitutionen zwischen Kulturförderung und Politik in Cisleithanien". In: Gottsmann, A. (ed.), *Die Habsburgermonarchie 1848-1918. Band X. Das kulturelle Leben. Akteure – Tendenzen – Ausprägungen*, p. 207–284. Verlag der Österreichischen Akademie der Wissenschaften. Wien.
- Vilanova Ribas, M.; Moreno Julià, X. (1992): *Atlas de la evolución del analfabetismo en España de 1887 a 1981*. Ministerio de Educación y Ciencia. Madrid.
- West, E.G. (1978): "Literacy and the Industrial Revolution". *The Economic History Review*, 31-3, p. 369–383.
- Zamagni, V. (1993): "Instrucción y desarrollo económico en Italia, 1861-1913". In: Núñez, C.E. and Tortella, G. (eds.), *La maldición divina. Ignorancia y atraso en perspectiva histórica*, p. 181–222. Alianza Editorial. Madrid.
- Zilch, R. (2014): *Finanzierung des Kulturstaats in Preußen seit 1800*, Acta Borussica, Reihe 2: Preußen als Kulturstaat. Abteilung II: Der preußische Kulturstaat in der politischen und sozialen Wirklichkeit. De Gruyter. Berlin.

## Appendix 1 Literacy rates (male, female and total aged 10 years old and over) in Spain (1860, 1877, 1887, 1900 and 1910)

	1860	1877	1887	1900	1910	1860	1877	1887	1900	1910	1860	1877	1887	1900	1910
	Male $T_{10}$ (estimation)	Male $T_{10}$ (estimation)	Male $T_{10}$	Male $T_{10}$	Male $T_{10}$	Female $T_{10}$ (estimation)	Female $T_{10}$ (estimation)	Female $T_{10}$	Female $T_{10}$	Female $T_{10}$	Total $T_{10}$ (estimation)	Total $T_{10}$ (estimation)	Total $T_{10}$	Total $T_{10}$	Total $T_{10}$
Álava	69.54	76.27	80.19	84.36	87.20	26.28	38.82	50.22	62.95	70.53	48.74	57.93	65.32	73.71	78.79
Albacete	25.66	28.37	33.44	36.56	40.08	7.71	10.72	14.23	17.78	21.99	16.60	19.42	23.80	27.15	31.05
Alicante	20.70	25.40	30.57	35.36	42.25	7.11	11.71	15.13	20.30	27.32	13.76	18.31	22.64	27.61	34.48
Almería	21.05	24.13	24.55	32.33	35.17	5.70	9.04	8.83	17.16	19.91	12.99	16.09	16.31	24.41	26.85
Ávila	45.09	54.17	59.03	63.57	68.08	9.73	18.98	27.70	35.79	43.50	27.54	36.42	43.17	49.39	55.44
Badajoz	28.08	32.38	35.52	35.32	42.55	10.85	16.39	19.72	22.33	29.24	19.74	24.48	27.74	28.87	35.90
Baleares	25.38	30.57	31.88	33.81	44.89	7.64	13.56	15.72	19.92	31.24	16.32	21.66	23.54	26.57	37.69
Barcelona	44.52	52.92	60.76	62.25	73.19	14.98	27.24	35.76	42.53	54.70	29.80	39.87	47.91	52.03	63.54
Burgos	68.80	77.48	81.80	84.84	88.05	13.28	25.69	34.84	51.30	62.87	41.08	51.13	58.12	67.94	75.28
Cáceres	36.35	40.13	44.16	49.54	52.07	9.14	14.09	18.27	24.26	31.57	22.99	27.16	31.30	36.92	41.68
Cádiz	36.07	38.35	43.03	44.53	50.57	23.86	28.49	32.81	34.97	40.83	30.44	33.54	38.02	39.73	45.75
Canarias	18.53	21.37	23.57	31.51	32.13	8.94	12.95	17.66	24.39	28.10	13.09	16.71	20.19	27.58	29.86
Castellón	21.39	25.64	27.95	35.11	42.53	4.58	8.62	9.16	17.26	23.94	12.94	16.95	18.55	26.12	33.16
Ciudad Real	30.28	33.56	36.48	39.32	42.63	8.36	13.73	16.65	20.68	23.88	19.42	23.64	26.53	29.88	33.17
Córdoba	25.25	29.12	34.85	35.46	38.58	10.84	15.56	19.07	21.83	26.70	18.01	22.29	26.94	28.65	32.66
Coruña	37.78	41.94	47.15	51.50	58.37	6.52	10.15	13.18	17.92	26.15	20.00	23.56	27.61	32.04	39.41
Cuenca	38.00	41.95	46.10	47.46	50.90	7.97	13.27	17.53	21.41	25.66	22.92	27.41	31.73	34.37	38.23
Gerona	38.10	45.13	51.22	59.79	66.45	9.82	17.11	24.13	34.38	46.75	23.94	31.09	37.72	47.00	56.58
Granada	23.87	20.33	26.50	31.23	38.19	9.21	10.52	13.97	18.69	25.44	16.48	15.35	20.16	24.88	31.71
Guadalajara	52.99	58.01	59.44	66.24	69.25	9.67	16.90	22.04	31.75	39.12	32.01	37.40	40.87	49.04	54.31
Guipúzcoa	31.51	40.72	47.70	58.16	67.41	15.89	27.20	35.85	49.62	61.48	23.64	33.93	41.64	53.75	64.34
Huelva	30.92	36.05	33.48	47.01	48.96	12.90	20.34	20.50	32.40	36.32	22.02	28.24	27.24	39.66	42.64
Huesca	34.47	41.09	46.82	55.70	59.99	5.76	11.61	16.79	27.32	35.38	20.54	26.58	32.22	41.77	47.94
Jaén	23.85	26.69	30.36	29.91	34.60	9.81	13.42	15.95	17.03	22.12	17.01	20.18	23.23	23.53	28.45
León	60.10	67.44	73.05	76.53	81.31	8.91	15.08	19.42	30.30	43.02	33.27	39.71	44.78	52.08	60.71
Lérida	29.48	34.44	40.92	50.03	57.65	5.03	11.55	15.75	26.54	36.80	17.49	23.01	28.56	38.58	47.49
Logroño	59.14	63.62	67.43	70.73	74.58	22.35	30.95	37.84	46.49	56.03	40.23	46.73	52.23	58.28	64.87
Lugo	41.40	41.36	49.39	53.40	61.00	3.96	5.60	8.42	15.64	21.10	21.09	22.09	27.61	33.26	39.31
Madrid	61.53	72.25	74.49	80.51	81.49	32.33	45.28	49.99	61.31	63.79	47.58	58.59	61.95	70.37	72.14
Málaga	23.27	24.27	27.36	29.94	30.64	11.04	14.25	16.33	20.31	20.68	17.14	19.11	21.69	25.00	25.54
Murcia	24.12	27.26	29.97	35.12	39.57	8.38	12.25	14.45	21.20	22.78	16.22	19.64	22.14	28.08	31.04
Navarra	48.66	54.45	59.88	68.02	72.19	19.30	30.51	38.72	52.48	61.00	34.00	42.69	49.16	60.11	66.47
Orense	36.02	38.10	45.85	49.50	48.21	3.11	5.28	7.50	12.15	26.75	18.51	20.48	25.29	29.10	36.43
Oviedo	51.25	58.28	64.90	69.63	79.78	9.67	15.60	22.83	33.12	56.32	27.94	34.10	41.26	49.70	66.76
Palencia	68.27	77.19	80.45	83.12	86.92	15.26	28.40	36.14	48.57	61.12	41.92	52.46	58.12	65.62	73.71
Pontevedra	47.86	51.69	53.95	57.28	65.13	4.34	9.16	11.75	18.19	28.72	22.49	26.55	28.67	33.71	43.22
Salamanca	50.67	59.72	65.53	71.71	76.39	12.11	22.88	30.96	44.14	51.47	31.39	40.93	48.03	57.64	63.40
Santander	72.81	78.97	83.18	80.87	91.37	19.89	32.27	43.34	57.06	83.92	44.00	53.21	61.29	68.05	87.34
Segovia	65.27	72.10	78.18	82.37	87.26	14.92	26.81	37.68	50.07	62.68	40.55	49.44	57.95	66.19	74.83
Sevilla	31.29	36.43	39.39	47.76	47.65	16.93	23.84	26.47	34.65	34.88	24.33	30.15	32.92	41.13	41.20
Soria	68.67	75.50	78.74	82.32	83.94	10.80	18.86	27.03	39.02	50.60	38.75	45.60	51.65	59.73	66.53
Tarragona	31.64	36.20	42.24	47.82	54.37	8.56	15.76	22.02	29.31	38.49	19.99	25.84	32.14	38.49	46.36
Teruel	36.74	40.55	44.73	49.66	54.42	5.38	9.66	14.49	21.61	25.82	20.69	24.60	29.37	35.55	40.12
Toledo	36.23	39.37	43.81	44.57	49.40	10.97	17.50	23.58	26.29	31.82	23.96	28.58	33.80	35.45	40.61
Valencia	26.81	26.26	35.41	41.30	44.91	8.90	13.35	18.12	23.62	28.95	17.77	19.71	26.70	32.37	36.86
Valladolid	60.39	67.85	72.08	74.87	78.87	18.76	29.73	37.50	46.72	55.71	39.97	48.33	54.62	60.38	66.84
Vizcaya	46.89	55.64	63.75	72.64	80.59	19.77	30.00	39.33	52.37	63.84	32.66	42.59	51.24	62.51	71.87
Zamora	54.76	61.64	70.28	76.94	79.29	11.03	18.02	24.20	32.30	41.89	32.41	39.03	46.38	53.57	59.22
Zaragoza	33.02	40.99	46.33	49.84	55.50	9.39	16.80	22.77	29.77	38.09	21.40	28.74	34.53	39.64	46.59
Melilla y P.S.	40.28	49.42	57.46	53.79	64.12	44.95	45.81	50.42	50.31	46.07	40.74	48.72	56.22	52.77	60.15
ESPAÑA	38.90	43.51	48.18	52.69	57.55	11.14	17.86	22.84	30.54	38.55	24.78	30.27	35.14	41.24	47.68

## Appendix 2 Literacy rates (male, female aged 11-15) in Spain (1887, 1900 and 1910)

	1887	1900	1910	1887	1900	1910
	Male $T_{11-15}$	Male $T_{11-15}$	Male $T_{11-15}$	Female $T_{11-15}$	Female $T_{11-15}$	Female $T_{11-15}$
Álava	83.94	83.88	88.63	69.76	77.80	84.15
Albacete	27.21	27.99	31.06	16.84	19.85	25.04
Alicante	23.81	35.75	34.32	16.31	26.63	28.73
Almería	15.49	26.61	26.83	8.19	19.61	21.47
Ávila	58.94	60.61	65.25	45.36	50.55	59.61
Badajoz	32.15	27.90	42.67	25.21	23.26	40.49
Baleares	26.10	30.17	45.33	18.57	24.62	40.21
Barcelona	61.28	61.27	73.81	46.14	52.35	66.01
Burgos	84.57	83.40	87.18	61.26	71.76	80.05
Cáceres	39.88	40.81	43.67	23.24	29.47	34.43
Cádiz	35.03	37.94	37.87	31.94	34.54	40.46
Canarias	20.85	28.20	28.39	22.23	28.81	31.05
Castellón	21.05	32.41	39.91	10.67	24.19	32.65
Ciudad Real	30.87	34.52	31.43	19.13	23.96	23.55
Córdoba	29.82	26.91	30.94	22.71	22.90	29.22
Coruña	42.06	47.20	56.23	20.43	28.54	39.73
Cuenca	40.11	37.71	42.65	23.62	24.49	29.18
Gerona	55.71	63.96	71.26	34.69	49.12	63.15
Granada	19.05	24.61	32.56	14.93	20.37	26.37
Guadalajara	55.87	61.67	64.79	31.39	43.38	51.39
Guipúzcoa	54.47	67.49	75.87	48.87	66.72	77.11
Huelva	25.74	38.00	44.97	22.08	33.47	40.90
Huesca	50.40	65.25	64.26	27.15	49.24	54.79
Jaén	27.25	22.66	23.92	20.28	18.23	12.54
León	66.08	69.86	77.75	28.09	41.54	60.46
Lérida	43.05	57.50	62.96	25.05	45.91	55.03
Logroño	67.27	69.02	73.62	54.04	59.99	69.00
Lugo	39.93	43.11	51.55	12.27	17.45	30.25
Madrid	72.48	77.82	77.06	58.01	72.01	69.78
Málaga	22.86	24.54	21.28	17.00	22.57	19.48
Murcia	23.92	28.06	28.51	15.85	21.33	21.72
Navarra	66.15	76.42	76.00	55.99	72.47	76.38
Orense	37.94	39.65	43.66	12.13	18.34	32.01
Oviedo	62.60	66.13	79.92	32.42	36.64	68.67
Palencia	80.86	80.56	85.47	53.54	63.26	75.69
Pontevedra	44.25	49.55	59.94	17.29	26.78	43.95
Salamanca	68.40	72.40	76.22	49.87	60.80	68.53
Santander	85.36	82.91	93.88	59.89	66.12	91.56
Segovia	80.74	81.19	86.28	57.21	66.35	77.73
Sevilla	32.96	46.40	38.03	28.82	39.67	33.62
Soria	80.01	82.16	85.38	47.96	60.39	77.21
Tarragona	44.53	47.67	59.39	31.26	39.45	53.54
Teruel	45.81	49.45	51.91	23.93	34.91	36.09
Toledo	38.21	35.36	38.06	27.79	27.90	32.46
Valencia	29.35	35.94	38.36	19.41	26.71	29.58
Valladolid	72.28	71.32	75.66	52.54	57.12	65.02
Vizcaya	67.76	76.96	85.80	52.48	68.61	78.41
Zamora	67.74	75.03	75.38	37.14	47.83	59.42
Zaragoza	49.87	50.02	56.71	33.51	39.24	50.35
Melilla y P.S.	63.00	62.93	60.05	65.69	57.61	49.46
ESPAÑA	44.33	49.11	53.43	29.58	38.07	45.90