Two stories, one fate:
Age-heaping and literacy in Spain, 1877-1930

Francisco J. Beltrán Tapia
Norwegian University of Science and Technology

Alfonso Díez-Minguela
Universitat de València

Julio Martinez-Galarraga
Universitat de València

Daniel A. Tirado-Fabregat
Universitat de València

[Preliminary and incomplete, please do not cite]

Abstract

This study examines the evolution of human capital during the early stages of modern economic growth. In order to do so, we have assembled a new dataset on age-heaping and literacy in Spain for both men and women between 1877 and 1930 based on six population censuses with information by age-groups for 49 provinces. Our results show that age-heaping was less prevalent during the second half of the 19th century and did not increase until the early twentieth century. By contrast, literacy increased throughout the whole period. Interestingly, age-heaping and illiteracy rates depict similar spatial patterns which confirm the stark differences in human capital across Spanish regions. Lastly, we raise critical questions as regards to sources and methods.

Keywords: Spain, age-heaping, literacy, nineteenth-century

JEL codes: I25, N9, O15, N01, I21

Address for correspondence: Departament d'Anàlisi Econòmica, Facultat d'Economia, Universitat de València, Av. dels Tarongers s/n, 46022, València. (alfonso.diez@uv.es)

1. Introduction

Economic inequality between territories is currently a widely-discussed subject in academic fields. Proof of how important the issue has become is its frequent appearance in newspaper articles and media-led debates about regional differences in income and the possible consequences[1]. In the case of Spain, regional economic inequality has also attracted the interest of economists and politicians and has consequently given rise to numerous studies (Mas *et al.*, 1994; De la Fuente, 2002; Cuadrado-Roura, 2010; Tirado *et al.*, 2016). Judging by these, there are three basic features of regional inequality in Spain.

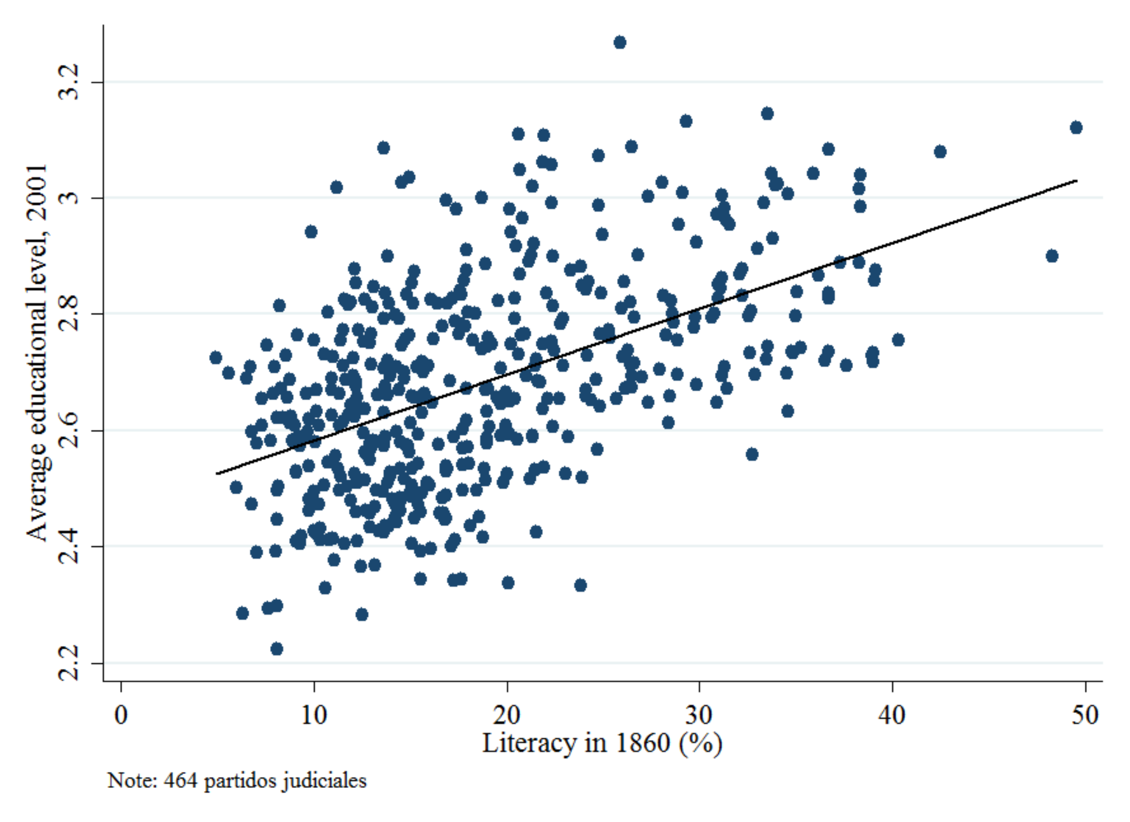
The first is its relatively high magnitude. Eurostat figures for 2015 show that the richest region in Spain, Madrid (123, with EU-28=100), has virtually twice the income per capita of Extremadura, the poorest region, with levels well below the European average (62). The second is its evolution over time. The convergence process has been interrupted and no reduction in regional differences is observed since the 1980s. If anything, regional differences have increased slightly. The third is the historical origins of regional inequality. Today's regional inequality took shape during the early stages of modern economic growth (1860-1930) when there was a significant divergence between territories (Rosés *et al.*, 2010; Martinez-Galarraga *et al.*, 2015). In other words, today's inequality is the result of different regional economic growth trajectories in the long term.

New Economic Growth theory has repeatedly pointed out that the availability of human capital is a determinant of growth over the long term (Romer, 1986; Lucas, 1988; Galor, 2011; Barro, 2013; Gennaioli *et al.*, 2013). Thus, the existence of marked differences in the regional availability of human capital in the past could be one of the main reasons for the differences in the economic trajectory of Spanish regions and therefore a key element for understanding current territorial inequality. In this respect, from the perspective of economic history, it has been pointed out that current regional differences in educational performance as stated by the PISA 2014 report are very close to the differences in literacy rates in mid-19th century. Relying on information at the district level, Figure 1 indeed shows that literacy rates in 1860 are strongly correlated with the average educational level in 2001. It is worth stressing that this correlation does not disappear if we control for the socioeconomic status

[1] Within the European Union, for example, the differences in wealth of the various territories have been and continue to be a cause of deep concern. Eurostat data for 2015 show that Inner London - West (580) was the richest NUTS2 region in the European Union (UE-28=100), while Severozapaden in Bulgaria was the poorest (27). These figures show the differences that exist between the two extremes: in terms of GDP per capita (PPS), Inner London was about 20 times richer than Severozapaden. And Luxembourg, the second wealthier region (264) had an income per capita more than 9 times that of Severozapaden.

of these districts[2]. All in all, the evidence points to the relevance of the legacy of history when it comes to determining the current regional differences in educational performance in Spain.

Figure 1.- Persistence in long-term educational attainments, Spain 1860-2001 (464 districts).



Source: Population Census (1860) and INE (2001).

Nevertheless, human capital is a broad and complex concept. It comprises health, cognitive abilities, knowledge, physical skills and even behavioural traits. Notwithstanding the fact that it is not easy to find indicators that properly measure human capital, contemporary indicators usually include information on educational attainments, school enrolment rates or years of schooling, to mention only a few. Yet, for historical periods, the availability of such information is scarce, or most frequently, absent. Population Censuses sometimes provide information on self-reported literacy rates back to the mid-nineteenth century but, in order to proxy literacy levels for previous periods, economic historians have often relied on the

[2] The regression coefficient decreases slightly (from 0.011 to 0.008) but continues to be statistically significant (p=0.000).

ability to sign official documents such as marriage registers (Cipolla, 1969; Allen, 2003; Reis, 2005)[3].

Given that these are partial indicators of a complex concept, a bulk of recent literature has proposed to broaden the number of indicators by constructing historical numeracy indexes as a complementary measure of human capital. In the end, as it has been posed, number knowledge and number discipline may be more crucial for economic growth than the ability to sign one's name in a marriage register or to self-report its ability to read or write (Crayen & Baten, 2010a). Numeracy skills can be defined as the ability to understand and work with numbers. In order to approximate these basic numerical skills -and thus human capital in a broad sense-, a growing number of works have focused on age-heaping by exploiting historical data on ages[4]. An irregular pattern in age distribution or preference for some specific digits may indicate the presence of misreported age, becoming an indicator of the aggregate numeracy skills of a society.

Following this line of work, in this paper we calculate numeracy levels in Spain on the basis of information included in the Population Censuses published in Spain since the late nineteenth century. In particular, our data set on national and provincial numeracy levels, which measure the level of age-heaping, is constructed using the information contained in several Population Censuses published between 1877 and 2010, although our analysis is mainly focused on the period 1877-1930[5]. In line with the abundant literature in the field of economic history, we first calculate the Whipple's index and a modified version of this index. Then, the Whipple index is converted into an ABCC index for practical matters. Compared to other sources used in historical age-heaping studies, Spanish Population Censuses provide high-quality information on the year-by-year age of the population for each one of the 49 provinces in successive censuses. This source also enables us to distinguish between men and women, thereby allowing us to study gender differences. Interestingly, given that population censuses also offer information on the self-reported ability to read and write, we can compare our age-heaping measures with provincial literacy rates.

Our results offer new evidence on the evolution of age-heaping in the early stages of modern economic growth in Spain. Our main finding is that age-heaping was relatively stable in the

[3] An alternative approach to measure human capital in distant periods is book production, an indicator of advanced literacy skills (Baten & Van Zanden, 2008).
[4] A recent survey of this literature can be found in Tollnek & Baten (2016).
[5] This period includes six Population Censuses: 1877, 1887, 1900, 1910, 1920, and 1930. In some instances, the Population Censuses of 1940, 1970, 1991, and 2001, are also used.

early counts, both at the national and provincial level. In fact, it did only improve significantly in the early twentieth century. This result goes in line with other works that claim that living standards only began to significantly improve in Spain in the early twentieth-century in terms of health and mortality rates (Pérez Moreda et al., 2015) or nutritional standards and heights (Martínez Carrión 2016).

Yet, this result is somewhat more difficult to reconcile with the existing evidence on literacy rates. First, contrary to age-heaping, literacy rates experienced a gradual and continuous increase throughout the period going from 1860 to 1930 (Núñez, 1992). Second, the level of age-heaping that we find, both at the national and provincial levels, was surprisingly low. ABCC indexes reach high values in a period of widespread illiteracy in Spain. Third, gender gaps in age-heaping are negligible, even in provinces with striking differences between male and female literacy rates. Taken together, these results cast some doubts on the extent to which digit preference may be capturing numeracy skills. In this sense, studying the Italian case, A'Hearn *et al.* (2016: 1) suggest that "*age-heaping is more plausibly interpreted as a broad indicator of cultural and institutional modernisation rather than a measure of cognitive skills*".

Our work also points to some additional relevant findings. From a methodological point of view, Spanish Population Censuses show that age-heaping emerges due to a preference for rounding ages in multiples of 10 (and do not exhibit a preference for ages ending in 5). Under these circumstances, the traditional Whipple index, commonly used by the literature, slightly underestimates age-heaping. In addition, using the information contained in all the Population Censuses published between 1877 and 1930, we find that age-heaping was less relevant in mid-nineteenth century Spain than what other studies suggest, thereby contradicting the clio-infra database. All in all, this paper raises critical questions as regards sources, methods and results obtained in previous work (Manzel, 2007; Crayen & Baten, 2010a; Juif & Baten, 2013 and Stolz *et al.*, 2013).

The rest of the paper is structured as follows. In section 2, we first define and discuss the concept of age-heaping. Then, the methodology and data are introduced and explained. The descriptive analysis is presented in section 3, while section 4 summarises and concludes.

2. Methodology and data

Self-reported age regularly appears in parish and military records, tax rolls, civil and legal documents, passenger lists, and population censuses. Yet, age has often been misreported either because custom and tradition, poor numeracy skills, or a badly designed and executed data collection process[6]. Digit preference for numbers ending in 0 and 5, as well as aversion for certain digits such as 4 and 13, has been extensively documented in economic and social history (A'Hearn *et al.*, 2009; Crayen & Baten, 2010a). In doing so, researchers have used a battery of tools to gauge age-heaping[7].

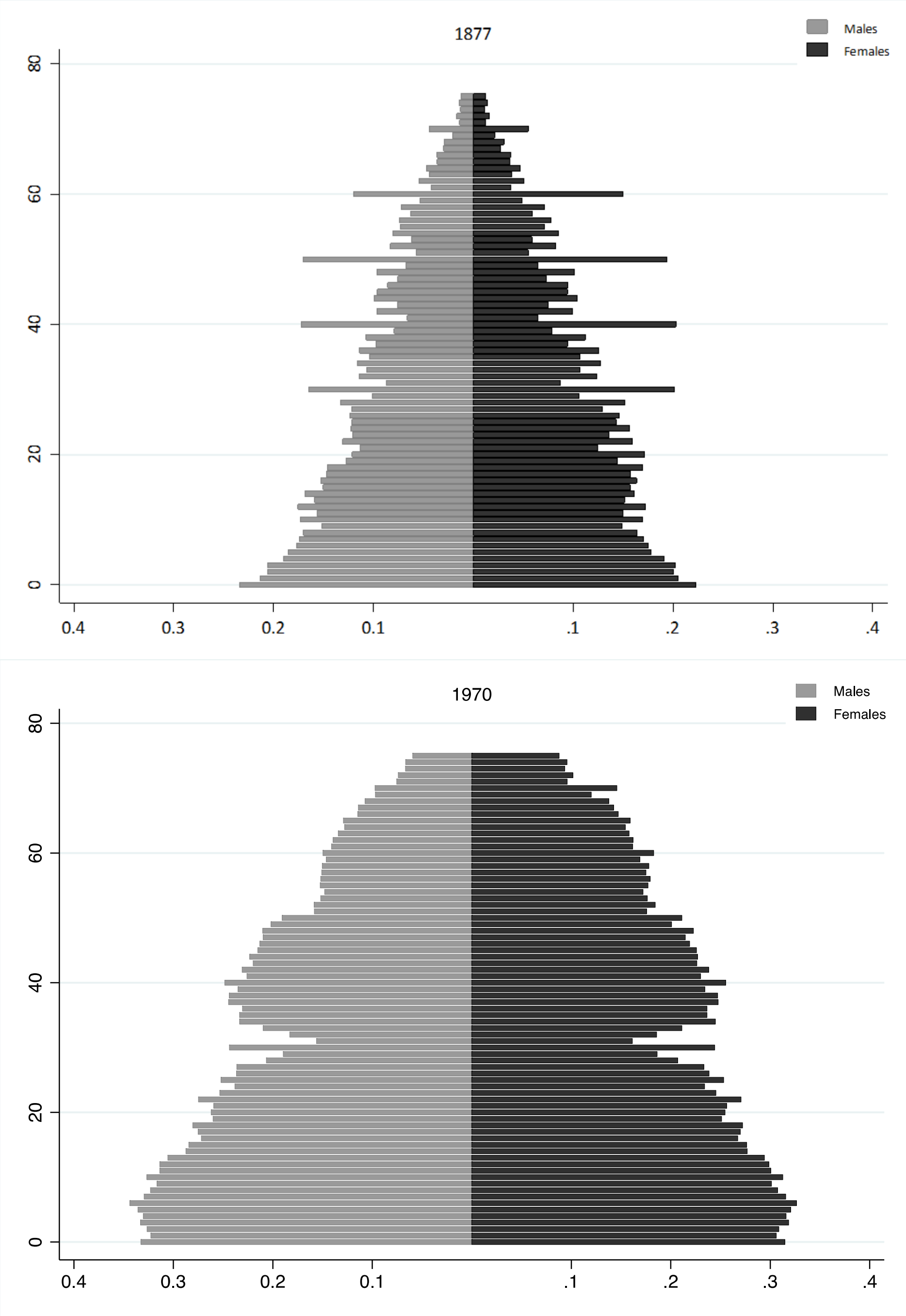
One of the simplest approaches to assess age-heaping is the Whipple index, which assumes that respondents are uniformly distributed over an age range. Suppose we have the number of individuals aged 23 to 62, where if P_{23} stands for the total number of respondents who reported an exact age of 23. If respondents were uniformly distributed, then it should not be expected any preference for ages ending in a specific digit.

$$\frac{(\mathrm{P}_{23}+\mathrm{P}_{33}+\mathrm{P}_{43}+\mathrm{P}_{53})}{1/10\,(\mathrm{P}_{23}+\cdots\mathrm{P}_{62})}=\cdots=\frac{(\mathrm{P}_{32}+\mathrm{P}_{42}+\mathrm{P}_{52}+\mathrm{P}_{62})}{1/10\,(\mathrm{P}_{23}+\cdots\mathrm{P}_{62})}=1$$

But, if digit preference exists then the above expression would not hold. This phenomenon is easily observed in the Spanish population censuses as figure 2 illustrates. The population pyramid in 1877 clearly displays an age preference for 0s which is absent in 1970. Besides, figure 2 exemplifies the changing structure of the population by age. Age-heaping is thus typically computed for a restricted age range, thereby excluding the bottom and top of the distribution. Besides, historical data sources, such as military records or marriage registers, overwhelmingly concentrated on young adults while elderly tended to exaggerate age.

[6] In traditional or preindustrial societies, unawareness of birthday or year of birth was common since records were not usually kept and numeracy skills were rather low. To eliminate or reduce age misreporting, age data are usually presented in 5-years age groups.
[7] See Shyrock & Siegel (1973) for a brief summary.

Figure 2. Population pyramid by age (0-75 years old) and gender in 1877 and 1970.



Source: INE

Bearing this in mind, the Whipple index is possibly the simplest and most popular indicator to measure preference for ages ending in 0 and 5 (Spoorenberg & Dutreuilh, 2007). It can be computed as the ratio of the number of respondents who reported an age ending in 0 or 5 to a fifth of the sample population. As previously stated, this approach assumes that respondents are uniformly distributed, consequently the oldest and youngest individuals are excluded. Conventionally, the Whipple index is computed for specific age cohorts (23-32, 33-42…) or for a whole range (23-62; 23-72). In both cases, the intervals must include an equal frequency of ending digits. Then, if P_{25} denotes the total number of respondents who reported an age of 25, the Whipple index can be computed as follows,

$$W = \frac{(P_{25} + P_{30} + P_{35} + P_{40} + P_{45} + P_{50} + P_{55} + P_{60})}{1/5\,(P_{23} + P_{24} \ldots + P_{61} + P_{62})}$$

By definition, the Whipple index varies between 1, which indicates no preference for ages ending in 0 or 5, and 5, which implies *perfect heaping*. Nevertheless, this indicator only detects the preference for multiples of 5 over those aged 23 to 62. It could be that age preference occurs only for ages ending in 0, as figure 2 illustrates, or for other terminal digits. Then, we may compute the preference for each digit[8]:

$$W_0 = \frac{(P_{30} + P_{40} + P_{50} + P_{60})}{1/10\,(P_{23} + P_{24} \ldots + P_{61} + P_{62})}$$

$$W_5 = \frac{(P_{25} + P_{35} + P_{45} + P_{55})}{1/10\,(P_{23} + P_{24} \ldots + P_{61} + P_{62})}$$

Still, this approach assumes linearity over 10 years, which seems a less plausible assumption (Spoorenberg & Dutreuilh, 2007). In order to account for preference and avoidance of all digits, Noumbissi (1992) proposed a modified version of the Whipple index:

$$\widetilde{W} = \sum_{i=0}^{9} (|W_i - 1|)$$

[8] The Whipple index can also be computed as follows: $W = (W_0 + W_5)/2$.

W_i measures the age preference for each terminal digit and is computed as,

$$W_0 = \frac{(P_{30} + P_{40} + P_{50} + P_{60})}{1/5\,(P_{30}^5 + P_{40}^5 + P_{50}^5 + P_{60}^5)}$$

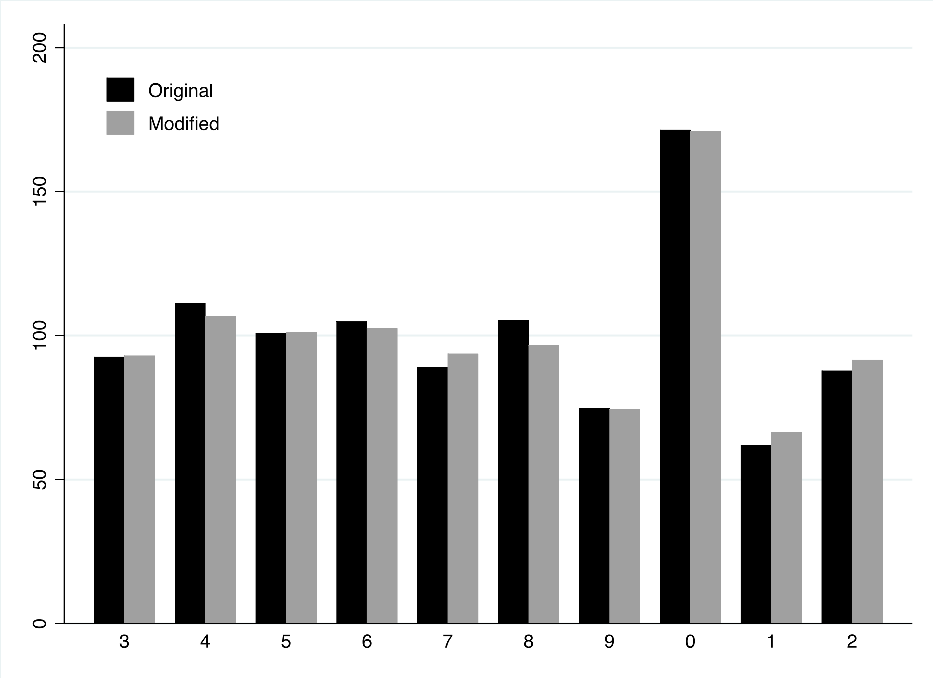
$$\dots$$

$$W_9 = \frac{(P_{29} + P_{39} + P_{49} + P_{59})}{1/5\,(P_{29}^5 + P_{39}^5 + P_{49}^5 + P_{59}^5)}$$

where P_{23} stands for the number of respondents who reported an age of 23 years old while P_{23}^5 is the population of those aged 21 to 25. In this case, a value of 0 would imply no heaping, while absolute preference for a specific digit would deliver a maximum value of 16. This modified version of the original Whipple index is also easy to compute, and more importantly, uses all information thereby providing further evidence and robustness. Anyhow, using the original and the modified approaches provide similar results, as figure 3 shows, and confirm that there is no preference for 5s[9].

[9] Digit preference for 5s is practically insignificant in the Spanish censuses, ranging from a maximum of 113.4 in 1920 to a minimum of 100 in 2001.

Figure 3. Digit preference (23-62 years) by method in the census of 1877.



Source: INE and authors' calculations.

For the sake of clarity and to facilitate the analysis, we follow A'Hearn *et al.* (2009) and convert the original Whipple index into a new measure that ranges from 0 to 100:

$$\mathrm{ABCC} = \left\{1 - \frac{(\mathrm{W}-1)}{4}\right\} \times 100 \ \text{ for W} \geq 1$$

$$\mathrm{ABCC} = \ 100 \ \text{ elsewhere}$$

This ABCC index indicates the share of respondents that report age correctly and varies between 0 and 100. Equally, the modified version can be easily transformed[10]. As regards data, we have the number of males and females aged 23 to 72 by exact age at the national level for all of the Spanish population censuses, except for the 1860 count, and provincial information for the following census years: 1877, 1887, 1900, 1910, 1920, 1930, 1940, 1970, 1991, 2001. All this information permits an in-depth study of age-heaping in Spain over time and across space since mid-nineteenth century.

[10] The modified Whipple index can be transformed as follows: $\widetilde{\mathrm{ABCC}} = \left\{1 - \frac{\widetilde{\mathrm{W}}}{16}\right\} \times 100$

Table 1. Age-heaping in Spain by population census.

Census	23-62 yrs.			23-72 yrs.		
Year	**ABCC**	**ABCC[1]**	**Obs.**	**ABCC**	**ABCC[1]**	**Obs.**
1877	91.0	89.6	8,016,415	91.0	89.5	8,655,253
1887	91.0	89.7	8,267,758	91.0	89.4	9,104,040
1900	90.1	89.4	8,704,427	89.9	89.0	9,528,154
1910	90.9	90.3	9,100,148	90.6	89.8	10,055,012
1920	91.5	91.3	9,775,251	91.2	90.9	10,804,622
1930	93.9	93.4	10,922,214	93.6	93.0	12,104,926
1940	95.6	95.3	12,221,886	95.3	95.0	13,580,667
1950	97.3	97.0	13,884,787	97.2	96.9	15,454,632
1960	98.6	98.3	15,387,112	98.6	98.2	17,289,637
1970	98.6	98.2	16,428,688	98.5	98.1	18,862,259
1981	99.4	98.8	18,087,752	99.4	98.9	20,898,435
1991	99.5	99.0	19,882,602	99.5	99.2	23,377,546
2001	100.0	99.1	22,925,766	100.0	99.3	26,916,613
2011	100.0	99.7	26,982,824	100.0	99.4	31,195,495

Note: ABCC[1] computed with the modified Whipple index or $\widetilde{W}$.
Source: INE and authors' calculations.

Table 1 summarises both the original and adjusted ABCC index for all population censuses. Unsurprisingly, using the traditional approach tends to slightly underestimate age-heaping. This is partly due to the fact that there seems to be no preference whatsoever for the digit 5 in the Spanish population censuses, as figure 2 shows. Moreover, although ABCC indices are greater when using the 23-62 age range, these differences are negligible. For the sake of comparability and robustness, we shall concentrate on the traditional ABCC index and the 23-62 age range in the remainder of this paper.

Having said that, it is also worth stressing that age-heaping has been generally associated with ignorance or low numeracy skills of the respondents, but it may also result from the enumerator (A'Hearn *et al.* 2009). For example, the capacity of an administration to carry out such a vast endeavour needs to be taken into account. A'Hearn *et al.* (2016) has recently suggested that age-heaping is not directly measuring numeracy skills, at least in nineteenth century Italy, but a "*broader mix of contextual factors*" related to the process of economic development. Anyhow, the difficulty of disentangling whether age heaping results from low numeracy skills of the respondents or a poorly designed and implemented data collection process recommends caution in the descriptive analysis that follows.

3. Descriptive analysis

The descriptive analysis begins with the population census of 1877, which was the earliest count reporting age-specific information in Spain, and a comparison of our findings with previous studies. In this regard, the clio-infra project provides historical estimates of age-heaping, measured with an ABCC index by birth decade for a large selection of countries[11]. Table 2 displays age-heaping estimates in the 1850s for a selection of European countries. In short, there appears to be a divide between Spain, Russia and the rest.

Table 2. Age-heaping in Europe during the 1850s.

Country	ABCC index (%)
Belgium	100.0
Finland	100.0
France	100.0
Sweden	100.0
Switzerland	100.0
Germany	99.7
Italy	99.4
Denmark	99.2
Netherlands	99.0
Norway	98.4
Austria	98.0
UK	97.1
Spain	87.4
Russia	83.6

Note: ABCC index (%) illustrated above is a birth decadal average.
Source: clio-infra (https://www.clio-infra.eu)

Nevertheless, the above estimates must be taken with caution. As regards Spain, the ABCC index presented in table 2 refers to the 43-52 age-group of the population census of 1900 (Manzel, 2007; Crayen & Baten, 2010a)[12]. Since age-heaping varies with age, it is important to check whether there is an age effect[13]. Crayen & Baten (2010b), for instance, found that

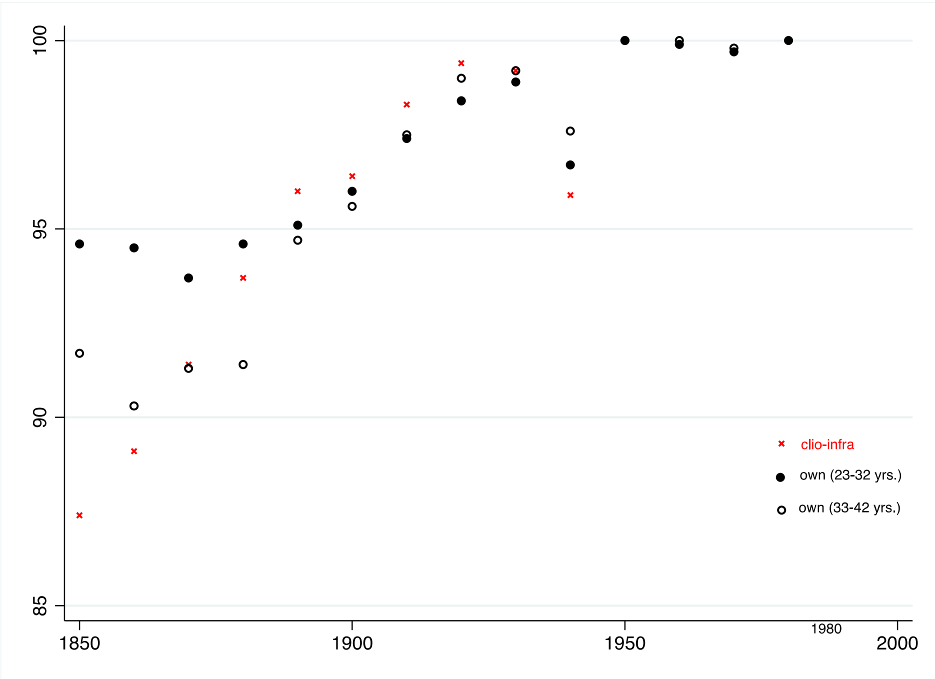
[11] Although the clio-infra historical estimates are based on several studies, the main references are A'Hearn *et al.* (2009), Crayen & Baten (2010a), and Prayon & Baten (2013). For a detailed list of sources and references see: https://www.clio-infra.eu/Indicators/NumeracyTotal.html. As regards Spain, the main references will be Manzel (2007); Crayen & Baten (2010a); Juif & Baten (2013) and Stolz *et al.* (2013).

[12] Similarly, the ABCC index reported for the 1830s and 1840s were constructed with the population census of 1900 (Manzel, 2007: 28). In this case, the age groups used were 63-72 and 53-62 years old respectively.

[13] See Crayen & Baten (2010b: 93-96).

the stronger age-heaping is the larger the age-effect[14]. In our case, data availability permits the comparison of distinct age-groups over time. In figure 4, we compare the age-heaping of those aged 23-32 and 33-42 in each population census with the clio-infra estimates. All values are presented by birth decade. Using the clio-infra ABCC indices, there appears to be a rapid improvement in age-heaping since 1850. If, however, we restrict to specific age-groups the ABCC indices remained relatively stable in the late nineteenth-century.

Figure 4. ABCC index (%) in Spain by birth decade, 1850-1980.



Sources: clio-infra (red coloured small-x); INE and authors' calculations for black-coloured circles.

This preliminary evidence thus casts doubts on the use of different age-groups from diverse data sources to evaluate the long-run dynamics of age-heaping. In fact, there are significant differences between the levels of age-heaping in each age-group until the early twentieth-century. Moreover, if ABCC indices for the whole 23-62 range and by gender are displayed (see figure A.1 in the appendix), age-heaping remains steady until the population census of 1920.

---

[14] In particular, Crayen & Baten (2010) suggest an adjustment to correct for the age-effect of the 23-32 group that consists of adding 0.2 Whipple units for every Whipple unit above 100 of the 33-42 group.

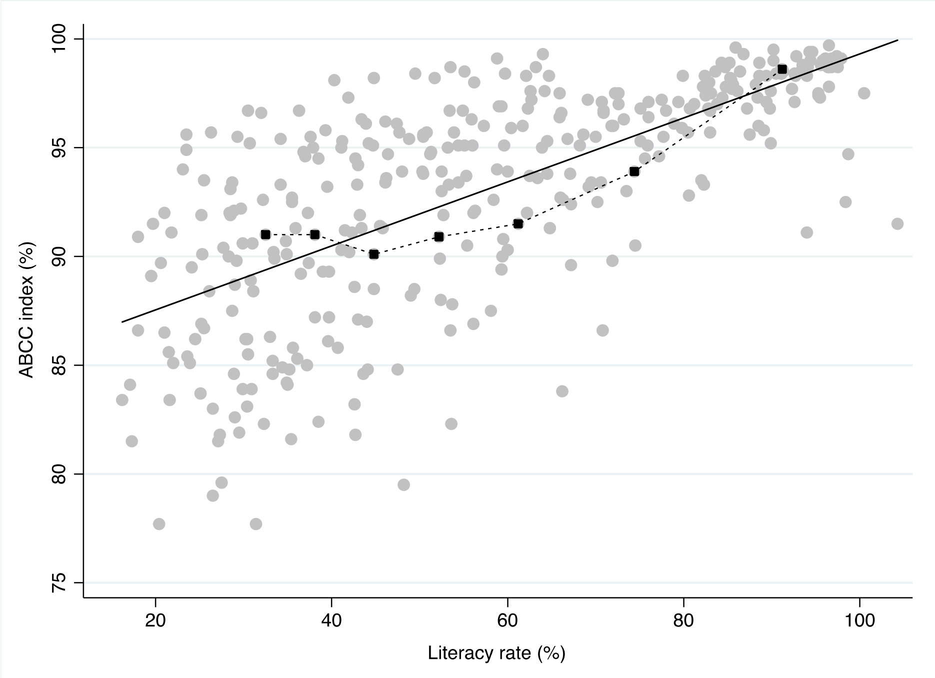
Our findings are also easy to reconcile with other historical studies. Pérez Moreda *et al.* (2015) claim that mortality rates worsened between the 1850s and the 1880s, improving slowly thereafter, especially in the early twentieth-century. Martínez Carrión (2016), on the other hand, examines the average height of military conscripts and argues that nutritional standards increased from the late nineteenth century to the Great War. Still more, A'Hearn *et al.* (2016), using the census of 1871, provide Whipple indices for Italy. When comparing our estimates with theirs, it appears that age-heaping was less prominent in Spain, thereby contradicting the clio-infra database[15]. In sum, our findings present a different story, where age-heaping was less significant and, above all, it did not improve until the early twentieth-century.
The above findings also raise a question on the extent to which age-heaping correlates with other proxies of human capital. In this regard, Núñez (1992) found that literacy improved during the second half of the nineteenth-century, which would challenge the early scenario. Figure 5 thus presents the ABCC index and literacy rate for each of the 49 provinces and several population censuses. Ideally, comparisons between age-heaping and literacy should be made using an identical sample, but Spanish censuses did not publish literacy by age. That said, it is worth remembering that the ABCC index is computed using information for individuals aged 23-62, while literacy is the share of the population aged 10 years or more capable of reading and writing. All in all, figure 5 displays 343 observations from 1877 to 1970.

At first glance, there appears to be a strong correlation between age-heaping and literacy[16]. However, when national averages are plotted (black-squares) it seems that literacy increased while age-heaping remained constant until the early twentieth-century[17]. This indeed calls for further analysis and discussion. As previously said, the census of 1877 was the first count reporting age-specific information, and it did so at the provincial level too[18]. Besides, the ABCC index relies on information for individuals aged 23 to 62 years, thereby implying that age-heaping refers to those born between 1815 and 1854. Therefore, the information

[15] See tables A.1 and A.2 in the appendix.
[16] Figure A.2 in the appendix displays the national ABCC indices and literacy rates for all population censuses. The correlation between age-heaping and literacy holds for each census too.
[17] In the appendix, figure A.3 further illustrates the relationship between age-heaping and literacy.
[18] We collected information of the population aged 23 to 62 years for each of the 49 provinces. This gives a total of 1,960 observations per census, and since we work with 7 censuses, our data set has been constructed on the basis of 13,720 observations. In addition, we also collected data by gender (which increases our total sample to 41,160 observations). Nonetheless, we leave this information for future analyses.

extracted from the census of 1877 resulted from an educational system rooted in the *Ancien Régime*[19].

Figure 5. ABCC index (%) and literacy rate in Spain by province for a selection of censuses.



Notes: Data for the following population censuses: 1877, 1887, 1900, 1910, 1920, 1930 and 1970. The solid line represents the regression line, while the black-squares are the national average.
Source: INE and authors' calculations.

That said, A'Hearn *et al.* (2009), using nineteenth-century United States censuses, examine the "numeracy-literacy" relationship with individual-level data[20]. For this, data were first aggregated by region, ethnic group and census. Then, the slope coefficient was estimated with Ordinary Least Squares (OLS). In general, a strong and positive correlation is found, but also substantial noise. When we regress age-heaping on literacy, the R^2 are on the whole smaller than in their case, ranging from 34.4 to 50.2, instead of the 77-83 range for the 1850,

[19] The first national education law (*Ley de Instrucción Pública*, also known as *Ley Moyano)* was passed in 1857. Thus, with the only exception of the cohort of 23-32 years, the remaining part of the population went to primary school under the *Ancien Régime* education system. By 1910, all the population considered went under the education system established by the *Ley Moyano*.

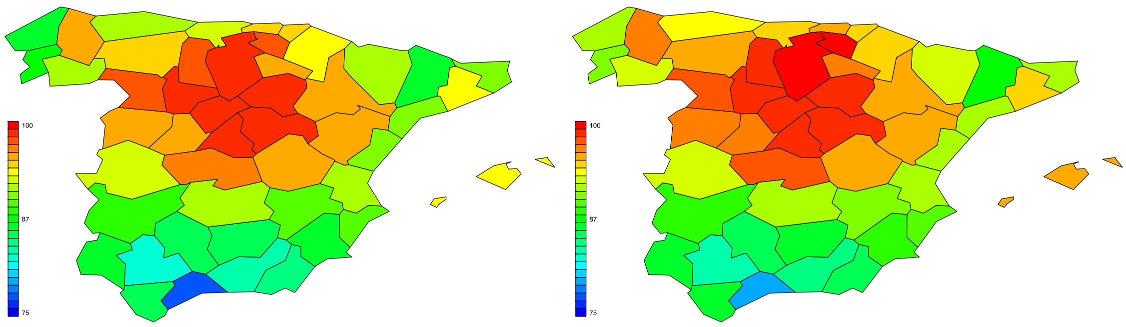
[20] A'Hearn *et al.* (2009) use information on men and women aged 20-69 for 1850, 1870, and 1900, from the Integrated Public Use Micro Samples (IPUMS).

1870, and 1900 U.S. censuses[21]. Moreover, the average literacy is substantially lower in Spain, while there appears to be less age-heaping.

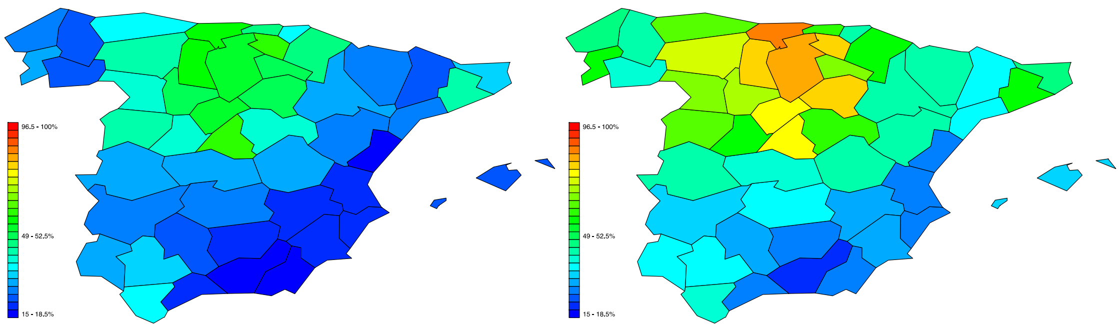
The above "anomaly", following A'Hearn *et al.* (2016), seems difficult to reconcile with the research that relates age-heaping with numeracy skills. For this reason, we also investigate the spatial dimension. Map 1 describes the spatial distribution of age-heaping and literacy in 1877. The distribution in each case has been structured into 25 equal-size categories, where dark-red provinces were near 100%, while those in dark-blue were at the other end of the distribution[22].

Map 1. Age-heaping and literacy in Spain in 1877.

(a) ABCC index (%), total (left) and male (right)



(b) Literacy rate (%), total (left) and male (right)



Notes: Provincial ABCC index (%) and literacy rates (%) classified into 25 equivalent categories for all censuses ranging from the minimum value (dark-blue) to the maximum or 100 (dark-red).
Source: INE and authors' calculations.

The spatial distribution of both measures presents certain similarities. There appears to be a core, namely the centre and centre-north, where age-heaping was practically inexistent. In

---

[21] Table A.3 in the appendix shows the coefficient and R^2 of regressing age-heaping on literacy for population censuses from 1877 to 1930.
[22] Map A.1 in the appendix presents the evolution of the ABCC indices and literacy rates from 1877 to 1970.

Burgos, Guadalajara, Madrid, Segovia, Soria, and Valladolid, the ABCC index was above 98%, thereby implying that less than 2% of individuals reported their age incorrectly[23]. To our surprise, this level of accuracy would place these provinces with the most advanced societies of the mid-nineteenth century[24]. Moreover, if we only consider those aged 23-32, the ABCC index is around 99%. Literacy rates, however, ranged from 40.3% in Guadalajara to 62.1% in Madrid, which would be rather modest[25].

This "anomaly" between age-heaping and literacy is even more acute when looking at both terms by gender. Whereas the difference in age-heaping was unimportant, female and male literacy rates differed widely. In Burgos, the ABCC index for men and women were 99.2% and 97.5%, but about a quarter of all women could read and write[26]. If we examine age-heaping for those aged 23-32, the gender gap fades. If, however, we explore the correlation between age-heaping and literacy by gender the relationship weakens once data for females are used, especially in the early censuses. All these findings call, once again, for a careful use of age-heaping as proxy of numeracy. In particular, there appears to be two major issues. First, the lack of correlation between gender gaps questions whether the information was self-reported. For the remainder of the paper, we shall use male information to overcome this problem. Second, age-heaping is abnormally low, given the widespread illiteracy, which casts doubt on the data-collection process.

As regards other territories, age-heaping was notable in Andalusia, Canary Islands, Galicia and the eastern coast where it ranged from 75% to 90%, while literacy rates were below the national average (32.5%). Malaga, for instance, exhibited the lowest ABCC index, 77.7%[27]. Asturias, Cantabria, Basque country, Navarra, Aragon, and Barcelona, on the other hand, were in-between of the distribution. Therefore, there was a divide within Spain, as map 1 illustrates and previous studies suggested (Núñez, 1992). This divide is particularly evident on the eve of the Spanish Civil War (1936-1939), when most northern provinces, apart from

---

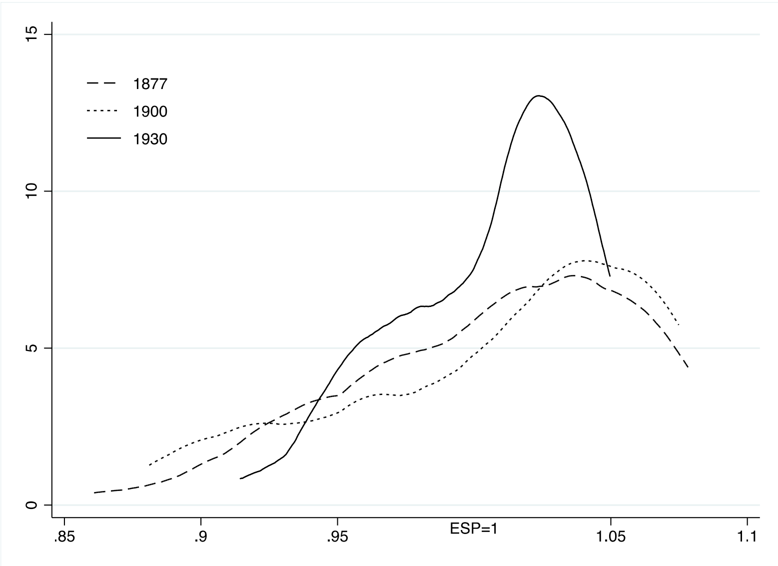
23 For the United Nations Statistics Division (UNSD) a Whipple index smaller than 105 (or an ABCC greater than 98.8%) is considered "highly accurate data".
24 Using clio-infra historical estimates, age-heaping in the north and centre-north provinces of Spain will be on similar levels than those computed for Norway or the Netherlands, see table 1.
25 In Spain, around 32.5% of those aged 10 or more years were capable of reading and writing in 1877.
26 In Soria, for instance, the male and female ABCC indices were 98.8% and 98.0% while the male and female literacy rates were 82.1% and 20.4%. Madrid, the male and female literacy rates were 76.3% and 48.3%. Figure A.4 displays the correlation between the gender gap in age-heaping and literacy from 1870 to 1930.
27 The United Nations Statistics Division (UNSD) labelled a Whipple index above 175 (or ABCC<81.3%) as "very rough data".

the Galician ones, had almost completed the transition to universal literacy, while illiteracy was still rampant in the south.

In order to shed further light on the subject, we also evaluate the long-run dynamics of age-heaping and literacy. Figure 6 presents the kernel densities of the provincial male ABCC index and literacy rates for three censuses: 1877, 1900, and 1930. In short, the initial variation in age-heaping within Spain eventually turned into a more "normal distribution". Interestingly, this convergence mainly occurred between 1900 and 1930, thereby reinforcing previous findings. Literacy, however, evolved into a "bimodal distribution" [28]. Again, the contrasting developments point to the nature of age-heaping.

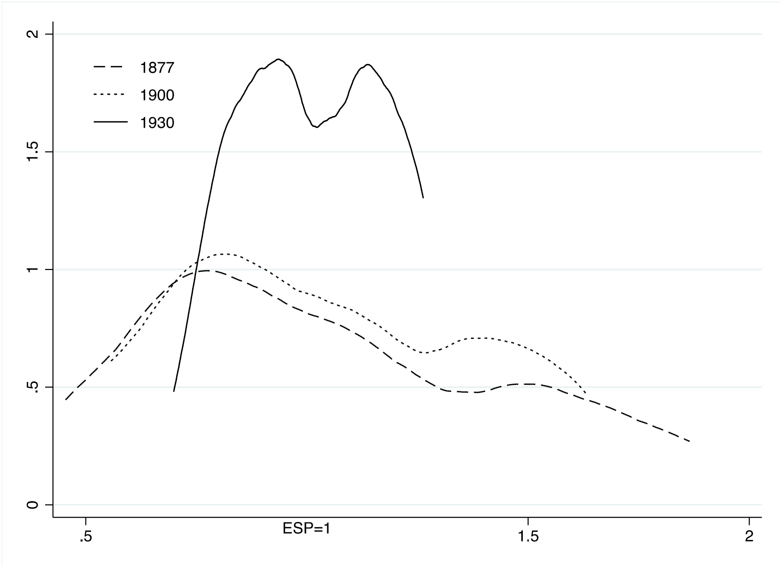
Figure 6. Male age-heaping and literacy kernel densities in Spain in 1877, 1900, and 1930.

(a) ABCC index (ESP=1)



[28] Map A.2 in the appendix illustrates the spatial distribution of age-heaping and literacy in Spain in 1930.

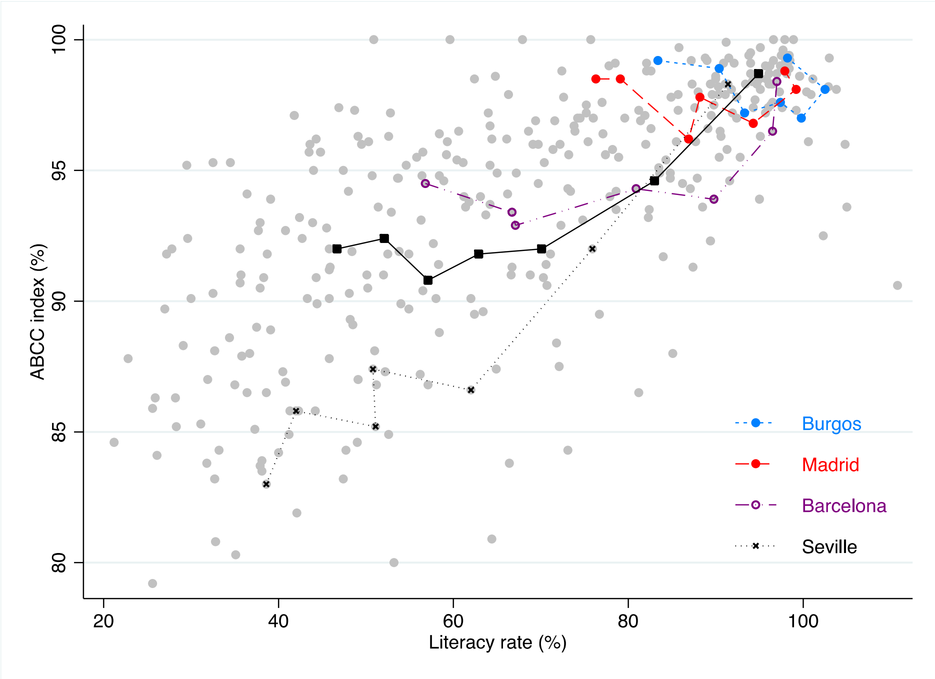
(b) Literacy rate (ESP=1)



Notes: Provincial ABCC index (%) and literacy rates (%) ranging from the minimum value to the maximum or 100.
Source: INE and authors' calculations.

To conclude, figure 7 displays the trajectories followed by four selected provinces: Burgos, Madrid, Barcelona, and Seville. Burgos and Seville, in terms of age-heaping, represent provinces at the top and bottom groups of the distribution, while Barcelona and Madrid contained the two largest cities. All things considered, figure 7 summarises previous findings. On the one hand, data for Burgos, and to a certain extent for Madrid, are apparently accurate and reliable. That is to say, age-heaping was practically inexistent. Likewise, literacy was around 80% in the earliest count, improving thereafter. In Barcelona, however, literacy rates were initially far from impressive but increased up to 1920 while the ABCC index shows a stagnant trend in that period. Finally, Seville eventually converged to a trivial age-heaping and almost complete male literacy by 1970.

Figure 7. Male age-heaping and literacy in Spain by province for a selection of censuses.



Source: INE and authors' calculations.

An important feature absent in our analysis are internal migrations, which could affect the stock of human capital[29]. In this line of research, Beltrán-Tapia and De Miguel Salanova (2017) found that the migrants moving to Madrid from 1880 to 1930 were on average more literate than the ones that remained in their provinces of origin. This is certainly relevant, since a positive selection would benefit receiving provinces while a negative one would, on the other hand, drive the stock of literates down. We acknowledge the relevance of internal migrations, but this early approach just aims at describing age-heaping in Spain.

4. Conclusions

This paper explores age-heaping in Spain since mid-nineteenth century. In doing so, we use the Spanish Population Censuses to compute a Whipple index and a modified version for each census year, by gender and by province. Then, the Whipple index is converted into an ABCC index for practical matters. Overall, our main findings can be summarised as follows. First, Spanish Population Censuses do not exhibit a preference for ages ending in 5. Consequently, age-heaping thus emerges due to a preference for rounding ages in multiples

[29] Although internal migrations have been recurrent in Spanish history, the flows increased rapidly at the turn of the twentieth-century (Silvestre, 2001, 2005).

of 10. Then, the traditional Whipple index slightly underestimates age-heaping. Similarly, there appears to be minor differences between age ranges (23-62; 23-72). Anyhow, we follow the literature and use the traditional approach, for the sake of comparability, in the descriptive analysis.

Second, age-heaping, either for an age-group (23-32) or over an age range (23-62), was less relevant in the mid-nineteenth century than previous studies suggest. Furthermore, there seems that age-heaping was relatively stable in the early counts. In fact, it did only improve significantly in the twentieth-century. But, if age-heaping is a proxy of numeracy this result is difficult to reconcile with the gradual improvement of literacy in Spain since the mid-nineteenth century. Third, the level of age-heaping at the national and provincial levels were surprisingly low, especially as regards the widespread illiteracy. In this regard, Italy and Spain exhibited similar levels of literacy in the 1870s, but age-heaping was less prominent in the Spanish censuses. Finally, gender gaps in age-heaping were negligible, even more in provinces where differences between male and female literacy rates were noteworthy.

In sum, the study of age-heaping in Spain provides further insight into the subject, but also casts doubts on the extent to which digit preference proxies numeracy skills, and henceforth human capital. If the enumerator succeeds to accurately gather self-reported information, then age-heaping results from poor numeracy skills, ignorance, or simply the "reluctance to make the effort to report age accurately" (A´Hearn *et al.*, 2016: 24). However, the capacity of the public administration to carry out a population census could somehow affect the effectiveness of the data collection process. In this regard, our findings indicate that in the centre, capital city of Madrid, and north-centre provinces age-heaping was trivial during the period of study which, again, is difficult to reconcile with the traditional backward view of nineteenth-century Spain. For this, further research and analysis is desired because the age-heaping and literacy stories should in the end converge towards the same fate.

References

A'Hearn, B., Baten, J., and Crayen, D. 2009. "Quantifying quantitative literacy: Age heaping and the history of human capital", *Journal of Economic History*, 69 (3), 783-808.

A'Hearn, B., Delfino, A., and Nuvolari, A. 2016. "Rethinking age-heaping. A cautionary tale from nineteenth century Italy", *Discussion Papers in Economic and Social History, University of Oxford*, Number 148.

Allen, R.C. Progress and poverty in early modern Europe", *Economic History Review*, 56 (3), 403-443.

Barro, R.J. 2013. "Education and economic growth", *Annals of Economics and Finance*, 14 (2), 301-328.

Baten, J., Crayen, D., and Voth, H.-J. 2014. "Numeracy and the impact of high food prices in industrializing Britain, 1780-1850", *Review of Economics and Statistics*, 96 (3), 418-430.

Baten, J. and Van Zanden, J.L. 2008. "Book production and the onset of modern economic growth", *Journal of Economic Growth*, 13 (3), 217-235.

Beltrán-Tapia, F.J., and De Miguel Salanova, S. 2017. "Migrants' self-selection in the early stage of modern economic growth, Spain (1880-1930)", *Economic History Review*, 70 (1), 101-121.

Budd, J.W., and Guinnane, T. 1991. "Intentional age-misreporting, age-heaping, and the 1908 Old Age Pensions Act in Ireland", *Population Studies*, 45 (3), 497-518.

Cipolla, C.M. 1969. *Literacy and development in the West*, Harmondsworth: Penguin.

Crayen, D., and Baten, J. 2010a. "Global trends in numeracy 1820-1949 and its implications for long-term growth", *Explorations in Economic History*, 47 (1), 82-99.

Crayen, D., and Baten, J. 2010b. "New evidence and new methods to measure human capital inequality before and during the industrial revolution: France and the US in the seventeenth to nineteenth centuries", *Economic History Review*, 63 (2), 452-478.

Cuadrado-Roura, J.R. 2010. *Regional policy, economic growth and convergence. Lessons from the Spanish case*, Heidelberg: Springer.

de la Fuente, A. 2002. "On the sources of convergence: a closer look at the Spanish regions", *European Economic Review*, 46(3), 569-599.

de Moor, T., and van Zanden, J.L. 2010. "'*Every Woman Counts*': A gender-analysis of numeracy in the Low Countries during the early modern period", *Journal of Interdisciplinary History*, XLI (2), 179-208.

Galor, O. 2011. "Inequality, human capital formation and the process of development". In E.A. Hanushek, S. Machin and L. Woessmann (eds.), *Handbook of the Economics of Education*, vol. 4, Amsterdam: Elsevier, 441-493.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., and Shleifer, A. 2013. "Human capital and regional development", *Quarterly Journal of Economics*, 128 (1), 105-164.

Guereña, J.-L., and Viñao Frago, A. 1996. *Estadística escolar, proceso de escolarización y sistema educativo nacional en España (1750-1850)*, Barcelona: Historia de la Educación (EUB).

Juif, D.T., and Baten, J. 2013. "On the human capital of 'Inca' Indios before and after the Spanish conquest. Was there a 'pre-colonial legacy'?", *Explorations in Economic History*, 50 (2), 227-241.

Lucas, R. 1988. "On the mechanics of economic development", *Journal of Monetary Economics*, 22 (1), 3-42.

Manzel, K. 2007. "Human capital development and the impact of the *Hungry Forties* in Spain", *mimeo*.

Martínez Carrión, J.M. 2016. "Living standards, nutrition and inequality in the Spanish industrialisation. An anthropometric view", *Revista de Historia Industrial*, 64, 11-50.

Martinez-Galarraga, J., Rosés, J.R. and Tirado, D.A. 2015. "The long-term patterns of regional income inequality in Spain, 1860-2000", *Regional Studies*, 49 (4), 502-517.

Mas, M., Maudos, J., Pérez, F. and Uriel, E. 1994. "Disparidades regionales y convergencia en las comunidades autónomas", *Revista de Economía Aplicada*, 4, 129-148.

Myers, R.J. 1954. "Accuracy of age reporting in the 1950 United States Cemsus", *Journal of the American Statistical Association*, 49 (268), 826-831.

Noumbissi, A. 1992. "L'indice de Whipple modifié: une application aux données du Cameroun, de la Suède et de la Belgique", *Population*, 47 (4), 1038-1041.

Núñez, C.E. 1992. *La fuente de la riqueza: educación y desarrollo económico en la España contemporánea*, Madrid: Alianza.

Núñez, C. E. and Tortella, G. 1993. *La maldición divina: ignorancia y atraso económico en perspectiva histórica*, Madrid: Alianza.

Pérez Moreda, V. Reher, D.-S., and Sanz Gimeno, D. 2015. *La conquista de la salud: mortalidad y modernización en la España contemporánea*, Madrid: Marcial Pons.

Prados de la Escosura, L. 2017. *Spanish economic growth, 1850-2015*, Cham: Palgrave-Macmillan.

Prayon, V., and Baten, J. 2013. "Human capital, institutions, settler mortality, and economic growth in Africa, Asia and the Americas", *Working Paper University Tuebingen*.

Reis, J. 2005. "Economic growth, human capital formation and consumption in Western Europe before 1800". In R.C. Allen, T. Bengtsson, and M. Dribe (eds.), *Living standards in the past. New perspectives on well-being in Asia and Europe*, Oxford: Oxford University Press, 195-225.

Romer, P.M. 1986. "Increasing returns and long-run growth", *Journal of Political Economy*, 94 (5), 1002-1037.

Rosés, J.R., Martinez-Galarraga, J. and Tirado, D.A. 2010. "The upswing of regional income inequality in Spain (1860-1930)", *Explorations in Economic History*, 47 (2), 244-257.

Shyrock, H.S., Siegel, J. and Associates 1973. *The Methods and Materials of Demography*, US Bureau of the Census.

Silvestre, J. 2001. "Viajes de corta distancia: una visión espacial de las migraciones interiores en España, 1877-1930", *Revista de Historia Económica*, 19, 247-286.

Silvestre, J. 2005. "Internal migrations in Spain, 1877-1930", *European Review of Economic History*, 9 (2), 233-265.

Spoorenberg, T., and Dutreuilh, C. 2007. "Quality of age reporting and application of the modified Whipple's index", *Population*, 62 (4), 729-741.

Stolz, Y., Baten, J., and Reis, J. 2013. "Portuguese living standards 1720-1980 in European comparison: heights, income and human capital", *Economic History Review*, 66 (2), 545-578.

Tirado, D.A., Díez-Minguela, A., and Martinez-Galarraga, J. 2016. "Regional inequality and economic development in Spain, 1860-2010", *Journal of Historical Geography*, 54 (4), 87-98.

Tollnek, F., and Baten, J. 2016. Age-heaping-based human capital estimates. In C. Diebolt and M. Haupert (eds.), *Handbook of Cliometrics*, London: Springer, 131-154.

APPENDIX: TABLES, FIGURES, and MAPS.

Table A.1 Age-heaping in Spain by age-group and gender in the population census of 1877.

	Total		Men		Women	
age-group	**W**	**Obs.**	**W**	**Obs.**	**W**	**Obs.**
23-32	122	2,589	118	1,210	125	1,379
33-42	135	2,178	130	1,060	139	1,118
43-52	150	1,846	147	908	153	938
53-62	147	1,403	139	694	155	709
63-72	134	639	126	317	142	321
average	138		132		143	
23-72	136		132		140	

Notes: Number of observations in thousands
Source: INE and authors' calculations.

Table A.2 Age-heaping in Italy by age-group and gender in the population census of 1871.

	Total		Men		Women	
age-group	**W**	**Obs.**	**W**	**Obs.**	**W**	**Obs.**
23-32	125	4,206	118	2,076	131	2,130
33-42	148	3,505	142	1,758	154	1,747
43-52	161	2,903	155	1,467	167	1,436
53-62	167	1,968	157	995	177	973
63-72	158	1,161	148	596	169	565
average	152		144		159	
23-72	147		140		154	

Notes: Number of observations in thousands
Source: A'Hearn *et al.* (2016: Table 2.1)

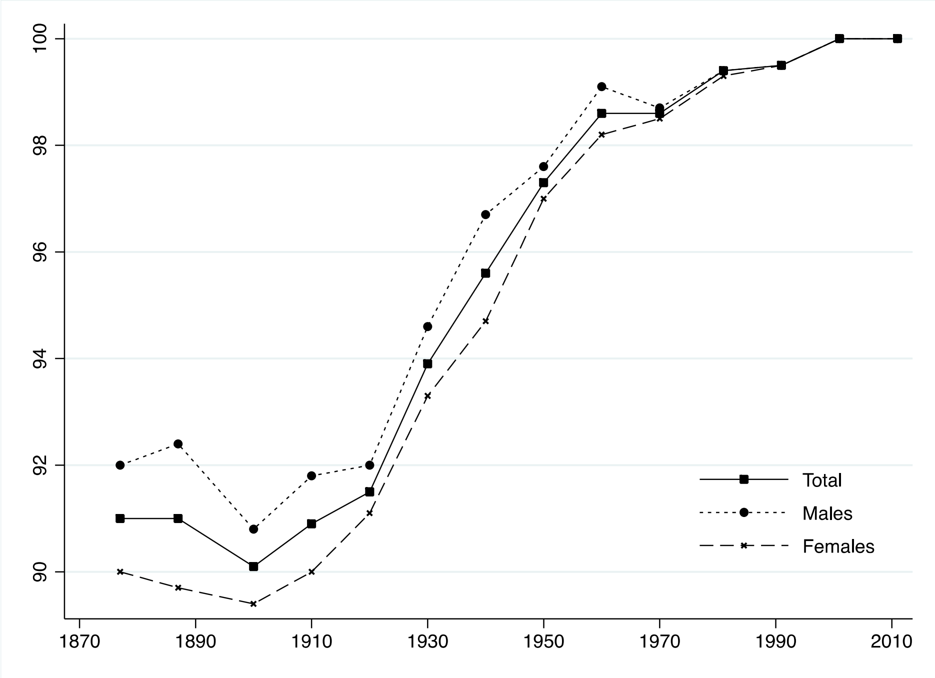
Table A.3 Provincial age-heaping and literacy in Spain.

Census year	Gender	Coefficient	R-squared	Obs.	Literacy average	Age-heaping average
1877	T	1.750 (0.275)	46.3	49	32.5	91.0
	M	2.987 (0.374)	58.1	49	46.7	92.0
	F	0.735 (0.236)	17.1	49	19.3	90.0
1887	T	1.975 (0.251)	46.8	49	38.1	91.0
	M	2.909 (0.402)	46.8	49	52.1	92.4
	F	1.291 (0.205)	36.2	49	25.0	89.7
1900	T	2.181 (0.269)	51.7	49	44.8	90.1
	M	2.766 (0.357)	53.7	49	57.1	90.8
	F	1.678 (0.242)	40.6	49	33.3	89.4
1910	T	2.568 (0.309)	44.0	49	52.2	90.9
	M	2.940 (0.425)	42.4	49	62.9	91.8
	F	2.222 (0.280)	39.2	49	42.4	90.0
1920	T	2.722 (0.361)	50.2	49	61.2	91.5
	M	2.824 (0.420)	49.8	49	70.1	92.0
	F	2.567 (0.330)	45.2	49	52.9	91.1
1930	T	2.437 (0.497)	34.4	49	74.4	93.9
	M	2.503 (0.563)	34.6	49	83.0	94.6
	F	2.337 (0.471)	29.8	49	66.3	93.3

Note: Age-heaping measured with ABCC index. Robust standard errors are in parentheses.

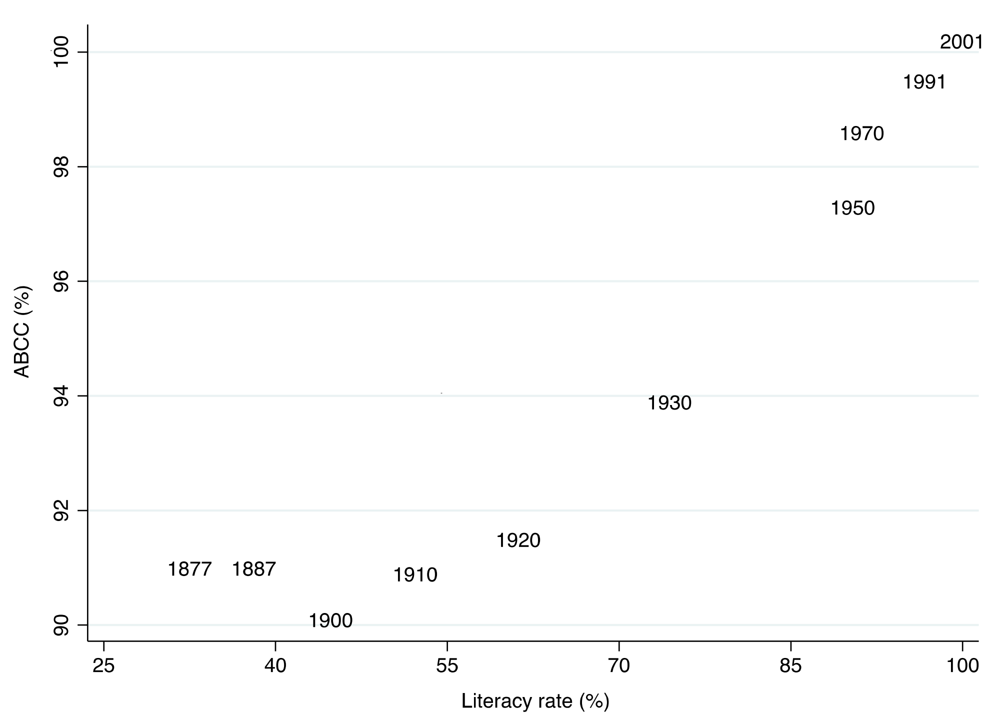
Source: INE and authors' calculations.

Figure A.1 ABCC index (%) in Spain by gender and population census, 1877-2001.



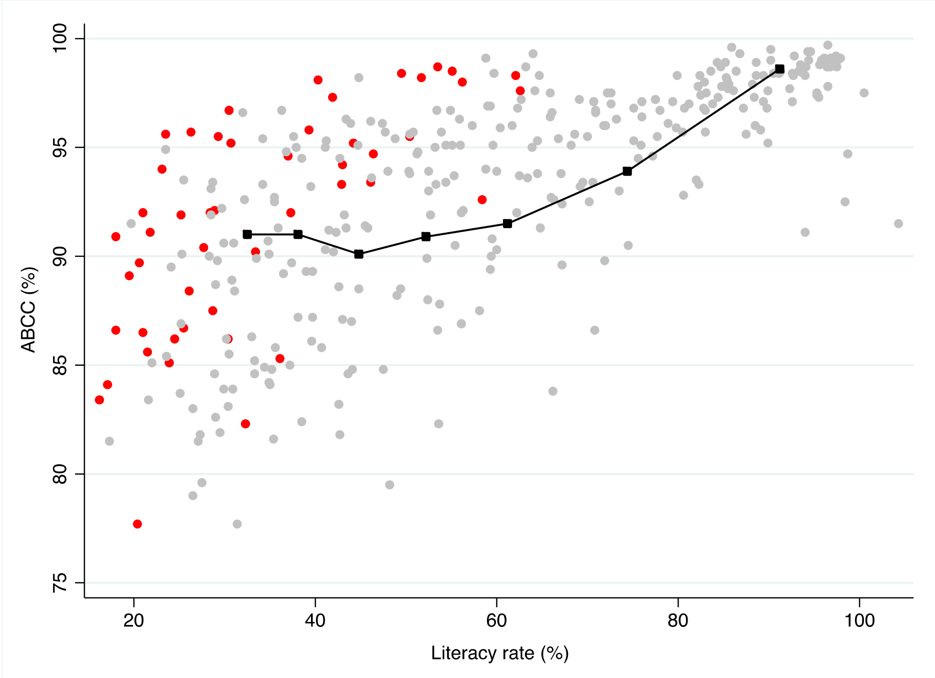
Note: ABCC index (%) constructed for those aged 23-62 yrs.
Source: INE and authors' calculations.

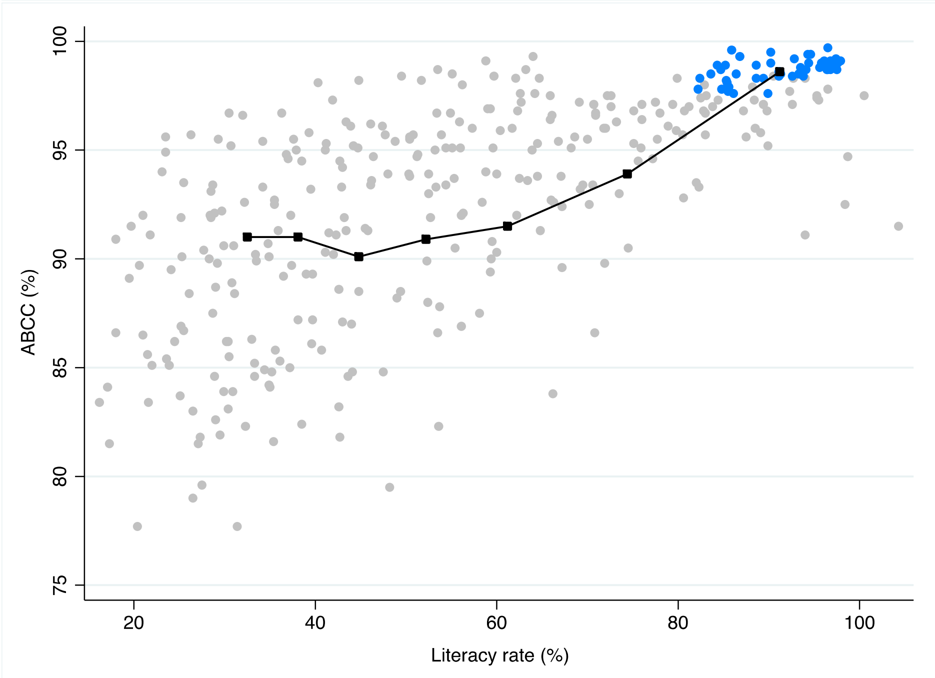
Figure A.2 ABCC index (%) and literacy rate (%) in Spain by population census, 1877-2001.



Note: ABCC index (%) constructed for those aged 23-62 yrs.
Source: Núñez (1992, 1993), INE and authors' calculations.

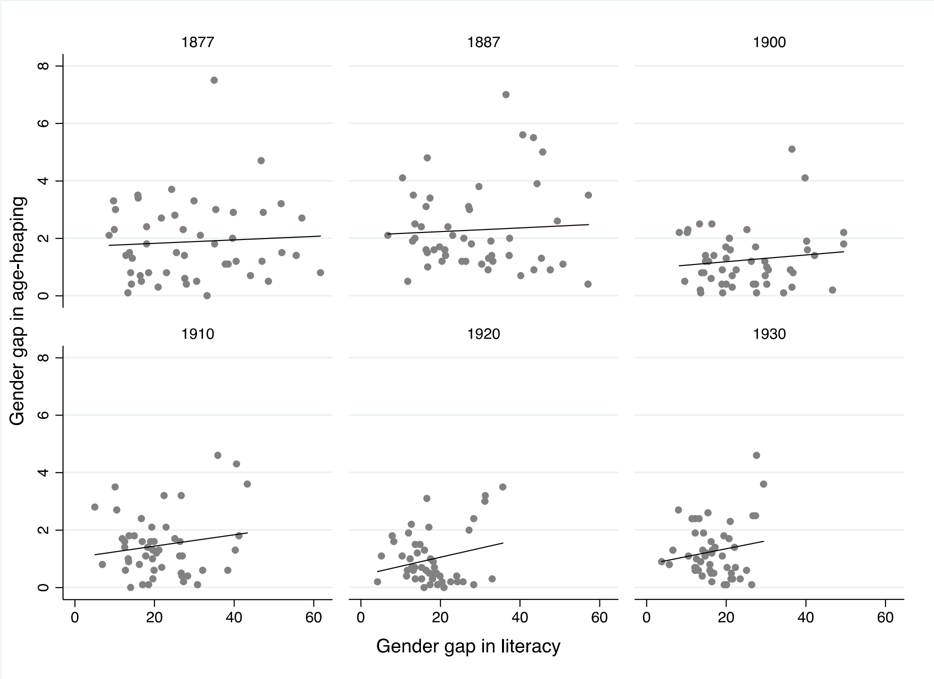
Figure A.3 ABCC index (%) and literacy rate (%) in Spain for a selection of censuses.





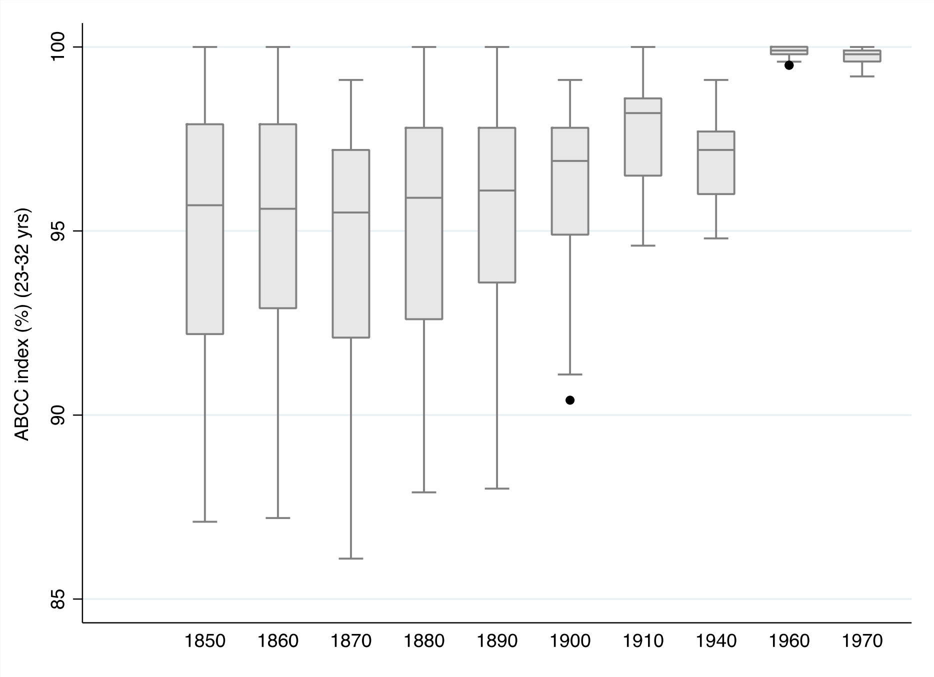
Notes: Data for the following population censuses: 1877, 1887, 1900, 1910, 1920, 1930 and 1970. The solid line represents the national average, while the red and blue dots illustrate data for 1877 and 1970 censuses respectively.
Source: INE and authors' calculations.

Figure A.4 Gender gap in age-heaping and literacy for a selection of censuses.



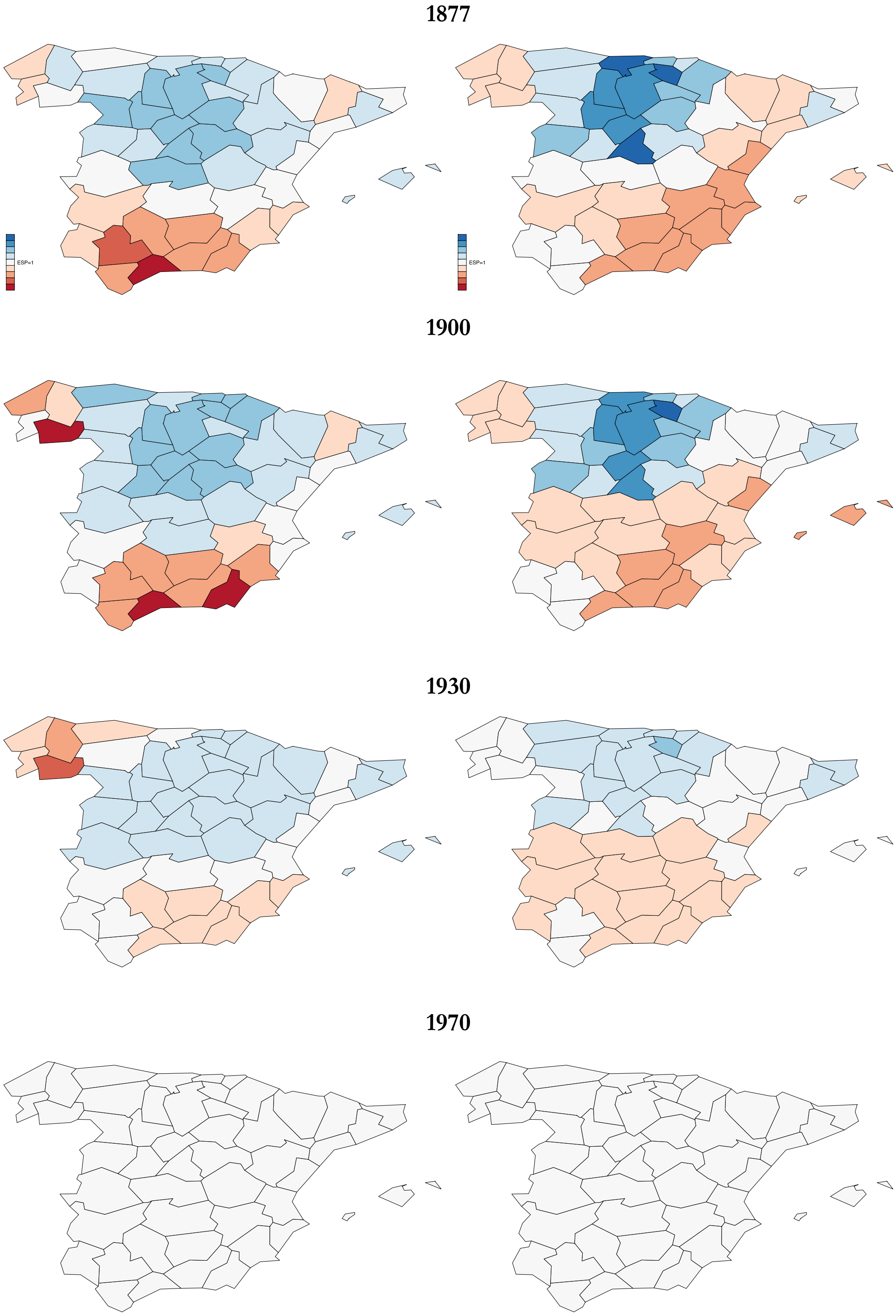
Source: INE and authors' calculations.

Figure A.5 Age-heaping for those aged 23-32 in Spain by province, birth decade, and census.



Source: INE and authors' calculations.

Map A.1 ABCC index (left) and literacy rate (right) in Spain in 1877, 1900, 1930 and 1970.

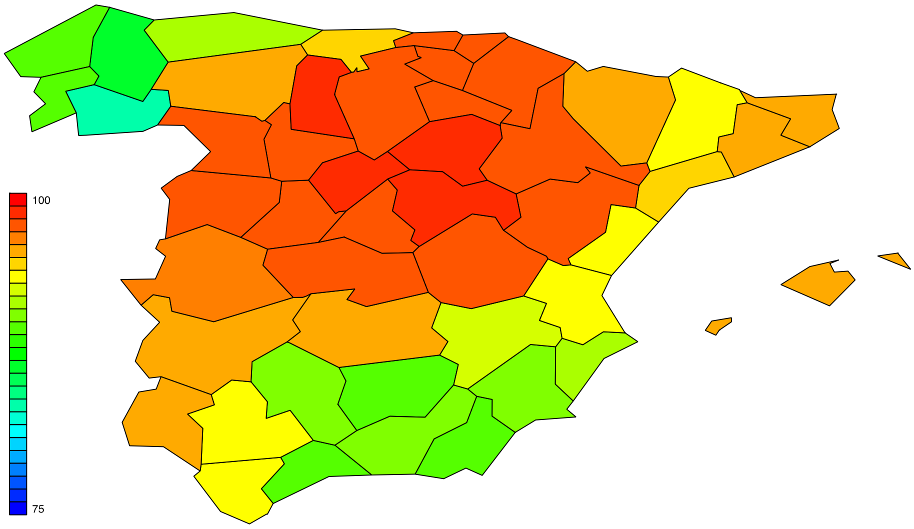


Notes: Provincial ABCC index (0-100) and literacy rates (%) classified into 9 equivalent categories for all censuses with Spain in white (ESP=1) and those above the national average in blue and below in red.

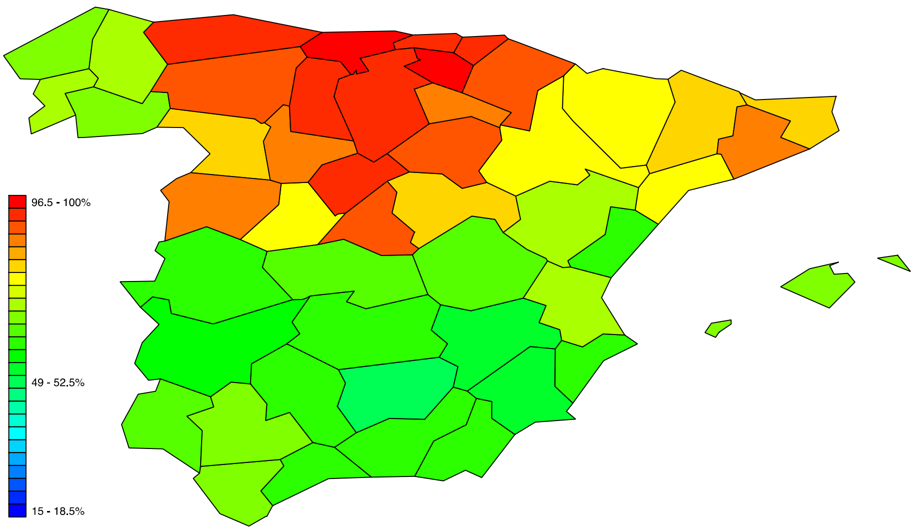
Source: INE and authors' calculations.

Map A.2 Age-heaping and literacy in Spain in 1930.

(a) ABCC index (%)



(b) Literacy rate (%)



Notes: Provincial ABCC index (%) and literacy rates (%) classified into 10 equivalent categories for all censuses ranging from the minimum value (light-yellow) to the maximum or 100 (dark-red).

Source: INE and authors' calculations.