European
Historical
Economics
Society

EHES WORKING PAPERS IN ECONOMIC HISTORY | NO. 139

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

OKTOBER 2018

EHES Working Paper | No. 139 | October 2018

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

Abstract

This study looks at human capital in Spain during the early stages of modern economic growth. In order to do so, we have assembled a new dataset on ageheaping and literacy in Spain for both men and women between 1877 and 1930 based on six population censuses with information for 49 provinces. Our results show that age-heaping was less prevalent during the second half of the 19th century than previously thought and did not decrease 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 within Spain. Lastly, we raise critical questions as regards sources, methods, and the interpretation of age-heaping.

JEL Codes: I25, N9, O15, N01, I21.

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

Acknowledgements:
Corresponding author: Alfonso Díez-Minguela (alfonso.diez@uv.es). Departament d'Anàlisi Econòmica, Facultat d'Economia, Universitat de València, Av. dels Tarongers s/n, 46022, València.

Notice
The material presented in the EHES Working Paper Series is property of the author(s) and should be quoted as such. The views expressed in this Paper are those of the author(s) and do not necessarily represent the views of the EHES or its members

1. Introduction

Economic and social progress is closely related to advances in human capital. In economics, numerous theoretical and empirical studies have stressed its relevance for economic growth (Romer, 1986; Lucas, 1988; Mankiw et al., 1992; Barro, 2013; Gennaioli *et al.*, 2013). In economic history, however, there is an intense debate about human capital. While some authors have claimed that human capital (and upper-tail knowledge) had a significant contribution to economic growth during the early stages of industrialization (Mokyr, 2009; Mokyr and Voth, 2010; Galor, 2011) other works question the existence of a large impact (Mitch, 1993, 1999, 2004; Allen, 2003; McCloskey 2010).

Human capital however constitutes a broad and complex concept (Goldin, 2015). 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 ability to sign official documents such as marriage registers (Cipolla, 1969; Allen, 2003; Reis, 2005)[1].

Given that these are partial indicators, 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 and 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[2]. 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.

[1] An alternative approach to measure human capital in distant periods is book production, an indicator of advanced literacy skills (Baten and Van Zanden, 2008).
[2] A recent survey of this literature can be found in Tollnek and Baten (2016).

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[3]. In line with the abundant literature on this topic, 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 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 first decades of the 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. First, contrary to age-heaping, literacy rates experienced a gradual and continuous increase between 1860 and 1930 (Núñez, 1992). Second, the level of age-heaping that we find, both at the national and provincial levels, is surprisingly low. ABCC indexes reach high values in a period of rampant illiteracy. 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 might 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*".

[3] 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.

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 and Baten, 2010a; Juif and Baten, 2013 and Stolz *et al.*, 2013). The rest of the paper is structured as follows. In section 2, a briefly historical background is presented. Next, the concept of age-heaping is defined and discussed in section 3. There, the methodology and data are introduced and explained. The descriptive analysis is presented in section 4, while section 5 summarises and concludes.

2. Historical background

The transition from the *Ancient Régime* to a liberal state in Spain was plagued with difficulties both in terms of recurrent political conflict and social unrest (Calatayud et al., 2016). In addition, from an economic perspective and compared to its European counterparts, Spain underwent a rather slow industrialization process (Prados de la Escosura, 2017). As a result, by 1910, Spain was still an agrarian economy, where roughly 66% of the active population was enrolled in agriculture (Nicolau, 2005). In this general context of economic backwardness and socio-political instability, educational matters were a source of concern for the successive liberal governments throughout the nineteenth century. The first national law of education (*Ley de Instrucción Pública*, also known as the *Moyano Act*) was passed in 1857. The *Moyano Act* established compulsory schooling for children aged 6-9 years, but funding relied upon local councils. Primary schooling thus depended upon the budget and wishes of each community[4]. The disappointing results of the *Moyano Act*, lack or slow progress in terms of literacy, witnessed during the second half of the nineteenth-century, led to the creation of the *Ministerio de Instrucción Pública y Bellas Artes* (Núñez, 1991).

[4] For a summary of the educational system before the *Moyano Act* see Guereña and Viñao-Frago (1996).

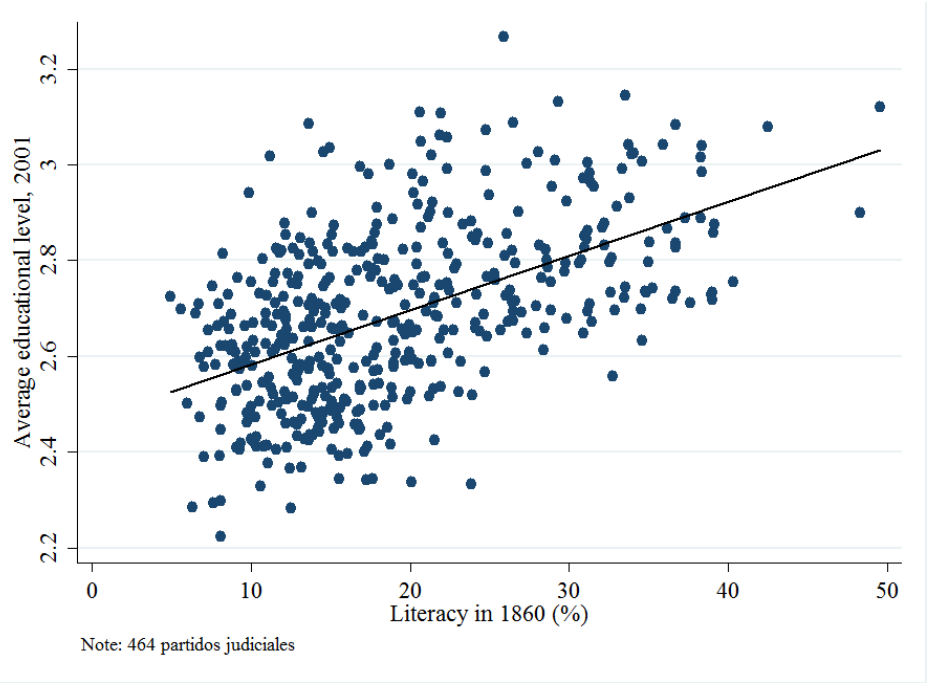
Accordingly, literacy rates in 1870 were indeed rather modest. While in Germany or United Kingdom, around 80% of the adult population could read and write, in Spain (as in Italy), literacy rates were around 30% (Crafts, 1997; Pamuk and Van Zanden, 2010). This figure is slightly lower if we consider the more precise information provided ten years before by the Population Census of 1860. In that year, only 26% of the population above 10 years reported that they were able to read and write (Núñez, 2005)[5]. Things were not much better regarding years of schooling. In 1870, the average years of schooling in Spain was 1.5 years, well behind other Western European countries such as Switzerland (6.1 years), Germany (5.4), France (4.1) or the UK (3.6)[6].

This situation of backwardness hides, nonetheless, large regional differences. A dual structure existed with the Northern provinces reaching higher rates of literacy than those in the South of the country (Núñez, 1992). In 1860, literacy rates in provinces like Almería were only 14%, but in Álava the population above 10 years able to read and write was 53%. It has been argued that 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 their diverse long-term economic success (Núñez, 1992; Núñez and Tortella, 1993)[7].

In this respect, from the perspective of economic history, it could be 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.[8] 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 and thus calls for a better understanding of the historical trajectory of human capital in Spain.

[5] Large gender differences were also present: While male literacy rates were 40%, female rates were only 12%.
[6] Spain ranked in number 20, just below Bulgaria and Uruguay. Other southern European countries, nonetheless, had a more disappointing performance like Greece (1.4), Italy (0.8) or Portugal (0.5). Data come from the Clioinfra project.
[7] On the long-term evolution of regional economies, see Martinez-Galarraga et al. (2015) and Tirado et al. (2016).
[8] It is worth stressing that this correlation does not disappear if we control for the socioeconomic status of these districts. The regression coefficient decreases slightly (from 0.011 to 0.008) but continues to be statistically significant (p=0.000).

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



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

3. 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[9]. 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 and Baten, 2010a). In doing so, researchers have used a battery of tools to gauge age-heaping[10].

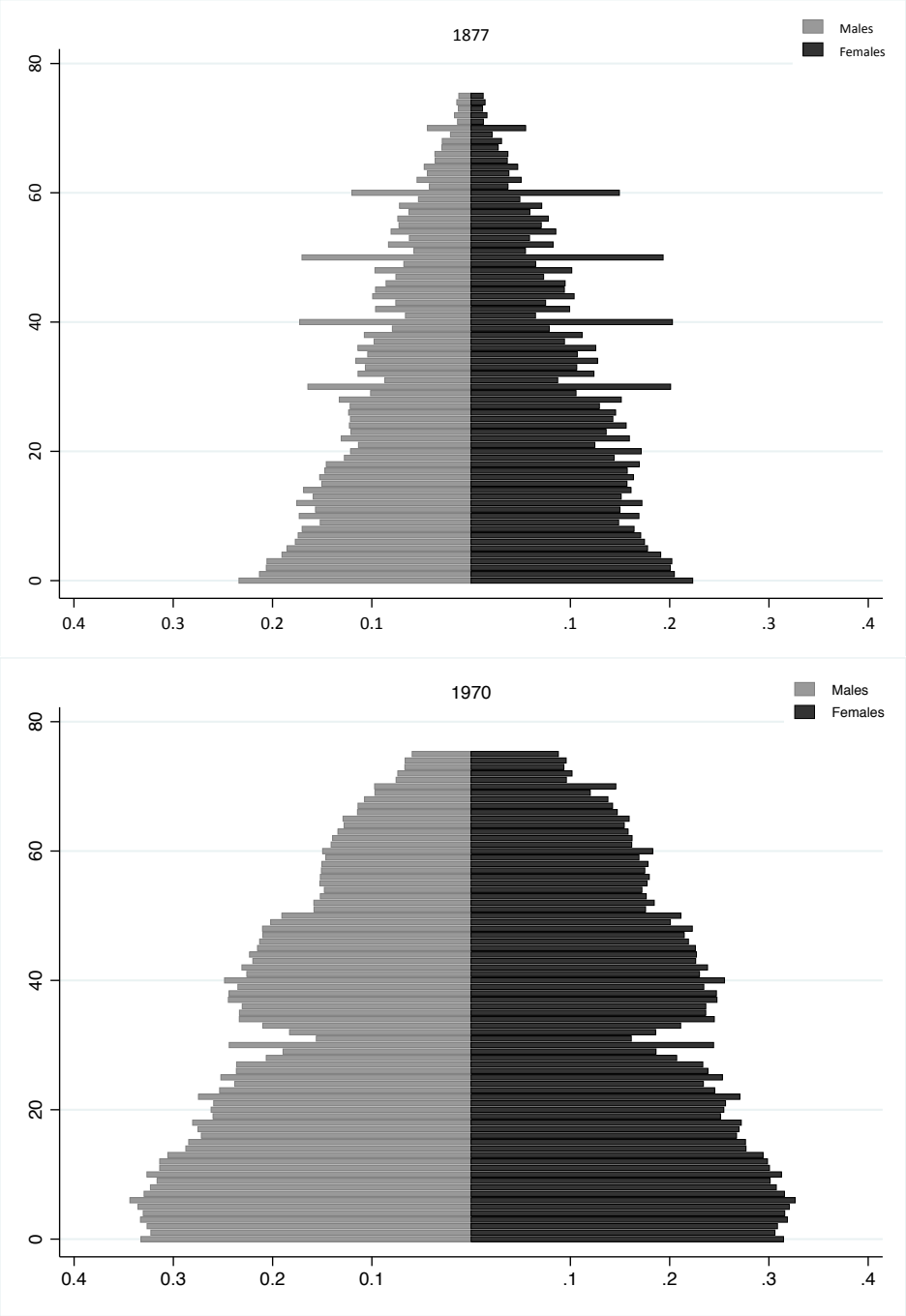
One of the simplest approaches to assess age-heaping is the Whipple index, which assumes that respondents are uniformly distributed over a specified age range. Suppose we know the number of individuals aged 23 to 62, where P_{23} stands for the total number of respondents who reported an exact age of 23. If respondents were uniformly distributed, no preference for ages ending in a specific digit should then be expected.

[9] 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.
[10] See Shyrock and Siegel (1973) for a brief summary.

$$\frac{(P_{23}+P_{33}+P_{43}+P_{53})}{1/10(P_{23}+\cdots P_{62})} = \cdots = \frac{(P_{32}+P_{42}+P_{52}+P_{62})}{1/10(P_{23}+\cdots P_{62})} = 1 \qquad (1)$$

But, if digit preference exists, then the above expression would not hold. This phenomenon is clearly observed in the Spanish population censuses as figure 2 illustrates. The population pyramid in 1877 clearly displays an age preference for 0s which is greatly reduced by 1970. Given that population is seldom uniformly distributed and that the elderly tends to overstate their age, age-heaping is typically computed for a restricted range, thereby excluding the top and bottom of the distribution. This also enables comparisons since other historical sources, such as military records or marriage registers, overwhelmingly concentrate on young adults.

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 and 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 given population. As previously said, this method 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,

$$W = \frac{(P_{25}+P_{30}+P_{35}+P_{40}+P_{45}+P_{50}+P_{55}+P_{60})}{1/5(P_{23}+P_{24}\ldots+P_{61}+P_{62})} \quad (2)$$

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. It is also common to find the Whipple index expressed in values going from 100 to 500. 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[11]:

$$W_0 = \frac{(P_{30}+P_{40}+P_{50}+P_{60})}{1/10(P_{23}+P_{24}\ldots+P_{61}+P_{62})} \quad (3)$$

$$W_5 = \frac{(P_{25}+P_{35}+P_{45}+P_{55})}{1/10(P_{23}+P_{24}\ldots+P_{61}+P_{62})} \quad (4)$$

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

$$\widetilde{W} = \sum_{i=0}^{9}(|W_i - 1|) \quad (5)$$

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

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

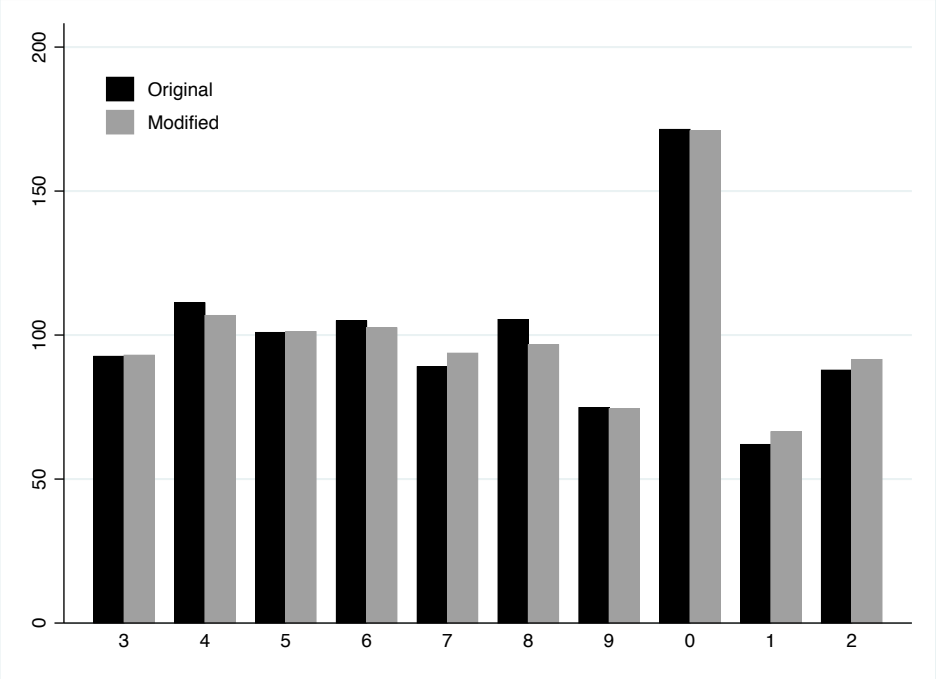
$$W_0 = \frac{(P_{30}+P_{40}+P_{50}+P_{60})}{1/5(P_{30}^5+P_{40}^5+P_{50}^5+P_{60}^5)} \quad (6)$$

…

$$W_9 = \frac{(P_{29}+P_{39}+P_{49}+P_{59})}{1/5(P_{29}^5+P_{39}^5+P_{49}^5+P_{59}^5)} \quad (7)$$

where P_{29} stands for the number of respondents who reported an age of 29 years old while P_{29}^5 is the population of those aged 27 to 31. 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. In any case, using the original and the modified approaches provide similar results, as figure 3 shows. It also confirms, relying on the data extracted from the population census of 1877, that in the case of Spain there was no preference for 5s[12].

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



Source: INE and authors' calculations.

[12] 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.

For the sake of clarity and to simplify 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:

$$\text{ABCC} = \left\{1 - \frac{(\text{W}-1)}{4}\right\} \times 100 \text{ for W} \geq 1 \quad (8)$$

$$\text{ABCC} = 100 \text{ elsewhere} \quad (9)$$

This ABCC index indicates the share of respondents that report age correctly and varies between 0 and 100. Equally, the modified version of the Whipple index ($\widetilde{\text{W}}$) can be easily transformed[13]. Table 1 summarises both the original (ABCC) and adjusted $(\overline{\text{ABCC}})$ indexes at the national level for all population censuses since 1877. These sources report the number of males and females by exact age. We compute two alternative indices, for the people aged 23-62 and 23-72 years. Unsurprisingly, using the traditional approach (ABCC) tends to slightly underestimate age-heaping. This is partly due to the fact that there seems to be no preference for the digit 5 in the Spanish censuses, as figure 2 shows. Additionally, although the indexes are greater when using the 23-62 age range than the 23-72, these differences are negligible. For the sake of comparability and robustness, we concentrate on the traditional ABCC index and the 23-62 age range in the remainder of this article[14].

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

[14] In order to avoid a survival bias, we follow Crayen and Baten (2010a) and omit population older than 62 or 72. Besides, Crayen and Baten (2010a: 93-96) proposed an age-adjustment procedure. However, given that Spanish censuses permit the study of age-heaping by age-cohort on a continuous basis we will not apply this correction and instead age-cohorts will be used. See also Budd and Guinanne (1991).

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

Census	23-62 yrs.			23-72 yrs.		
Year	ABCC	$\overline{ABCC}$	Obs.	ABCC	$\overline{ABCC}$	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: $\overline{ABCC}$ computed with the modified Whipple index or $\overline{W}$.
Source: INE and authors' calculations.

Apart from the information at the national level, we have gathered provincial data 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. In addition, the Spanish population censuses also include disaggregated information on males and females thus allowing the analysis by gender. Moreover, given that we have data on self-reported reading and writing skills from the same statistical sources we can explore the relationship between age-heaping and literacy over time.

4. Descriptive analysis

4.1. Age-heaping in Spain since the mid-nineteenth century

The census of 1877 was the first modern count reporting age-specific information for all the population. Since the ABCC index has been constructed for the 23-62 age group, it then uncovers information from individuals born between 1815 and 1854[15]. Table 2

[15] 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* educational system. By 1920, all the population considered went under the education system established by the *Ley Moyano*. Actually, in that date, the younger cohort 23-32 studied in a new educational system put forward after the creation of the *Ministerio de Instrucción Pública y Bellas Artes* in 1900.

displays age-heaping, measured with an ABCC index by birth decade (1850s), for a sample of European countries using historical information from the clio-infra project[16].

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

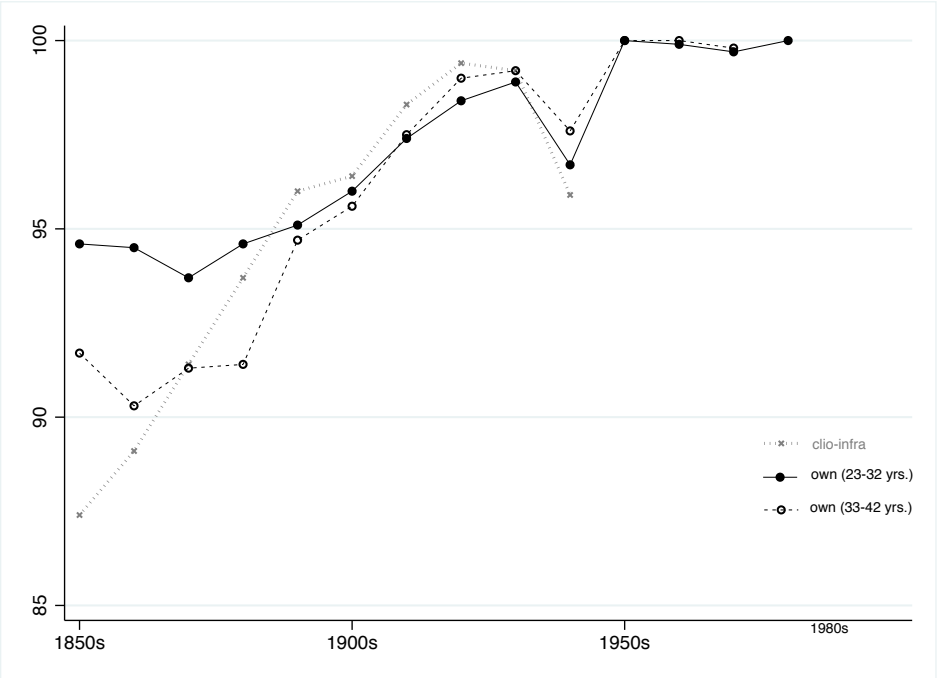
Country	ABCC index (%)
Belgium	100.0
Finland	100.0
France	100.0
Sweden	100.0
Switzerland	100.0
Germany	99.7
Italy	99.4
Denmark	99.2
Netherlands	99.0
Norway	98.4
Austria	98.0
UK	97.1
Spain	87.4
Russia	83.6

Note: ABCC index (%) illustrated above is a birth decadal average.
Source: Clio-infra

According to table 2, there appears to be a divide between Spain, Russia and the rest. Nevertheless, the above estimates must be taken with caution. In Spain, the ABCC index presented in table 2 actually refers to those aged 43-52 in the population census of 1900 (Manzel, 2007; Crayen and Baten, 2010a)[17]. Since age-heaping varies with age, it is important to check whether there is an age effect[18]. Crayen and Baten (2010b), for instance, find that the stronger age-heaping is the larger the age-effect[19]. In our case, data availability permits the comparison of different age-groups over time.

[16] Although the clio-infra historical estimates are based on several studies, the main references are A'Hearn *et al.* (2009), Crayen and Baten (2010a), and Prayon and Baten (2013). For a detailed list of sources and references see: https://www.clio-infra.eu/Indicators/NumeracyTotal.html. As regards to Spain, the main references are Manzel (2007), Crayen and Baten (2010a), Juif and Baten (2013) and Stolz *et al.* (2013).
[17] 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.
[18] See Crayen and Baten (2010a: 93-96).
[19] In particular, Crayen and Baten (2010a) 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.

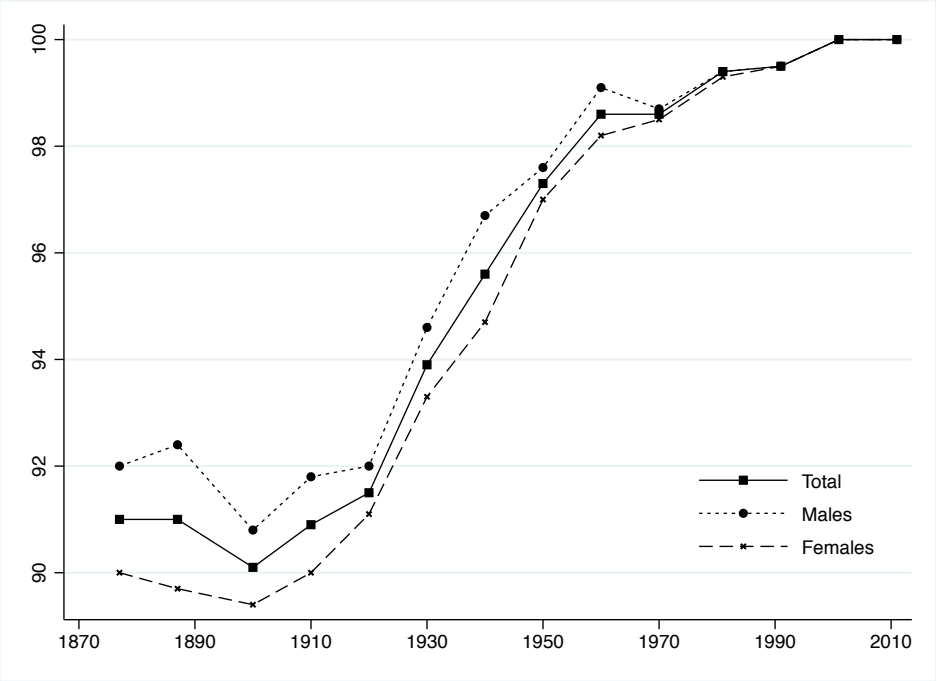
Figure 4. ABCC index (%) in Spain by birth decade, 1850-1980.



Note: Clio-infra makes use of distinct age-groups for constructing the ABCC indices.
Sources: Clio-infra; INE and authors' calculations.

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. First, our results show that the ABCC index in mid-nineteenth century Spain was not as low as described in the existing literature. Further, using the clio-infra ABCC indices, there appears to be a rapid improvement in age-heaping since the 1850s. If, however, we restrict to specific age-groups with the information from all population censuses, the ABCC indices remained relatively stable in the late nineteenth-century. This evidence thus casts doubts on the use of different age-groups from diverse data sources to evaluate the long-run dynamics of age-heaping. If ABCC indices for the whole 23-62 range and by gender are displayed (see figure 5), it also seems that age-heaping did not improve much until the population census of 1920. Besides, this trend is found both for males and females, although the latter show smaller values, particularly before the 1920s. Altogether this evidence suggests that there were limited, if any, advances in numeracy during the late-nineteenth and early-twentieth century.

Figure 5. 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.

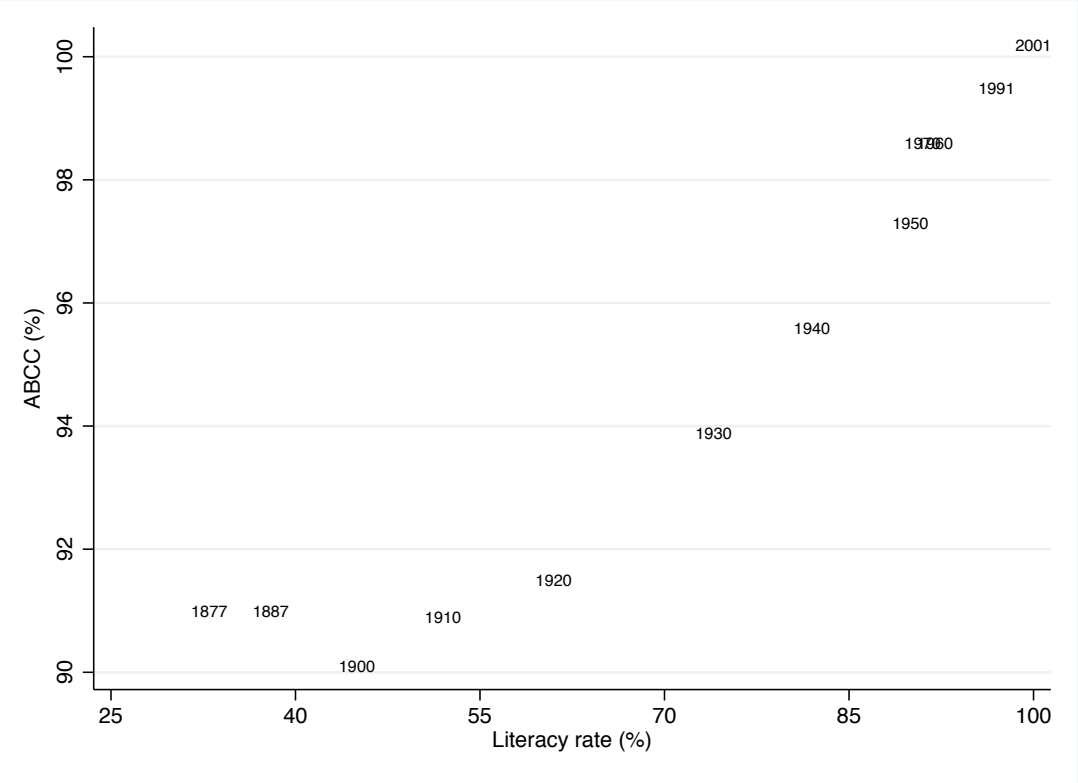
This result, in turn, goes in line with the existing evidence on living standards. Pérez Moreda *et al.* (2015) find that mortality rates in Spain increased between the 1850s and 1880s, falling gradually thereafter. Equally, Martínez Carrión (2016), using military records, finds that average adult heights only improved from the 1880s onwards[20]. Our findings also suggest that age-heaping for those born in the nineteenth-century remained unchanged. In addition, A'Hearn *et al.* (2016), using the Italian census of 1871, provide Whipple indices by age-groups and gender. When comparing these with those values obtained from the Spanish census of 1877, it seems that age-heaping was less prominent in Spain, thereby challenging the evidence in table 2 from the clio-infra project[21]. In sum, our findings provide a different story where not only was age-heaping less prevalent in the second half of the nineteenth-century, but also it did not improve until the twentieth-century.

Furthermore, the above findings raise questions on the interpretation of age-heaping. It would therefore be appropriate to assess whether age-heaping correlates with other indicators of human capital such as literacy rates. Broadly speaking, if age-heaping captures numeracy skills then the ABCC indices should be correlated with other cognitive skills like reading and writing. Although comparing age-heaping with literacy should be made with

[20] Baten et al. (2014) study the relationship between numeracy and living standards in industrializing Britain.
[21] See tables A.1 and A.2 in the appendix.

individual-data, we first gauge the potential existence of a relationship between these two indicators by looking at the evolution of both age-heaping and literacy rates between 1877 and 2001 at the national level[22]. As Figure 6 clearly illustrates, while literacy rates steadily increased throughout the whole period, the levels of age-heaping remained relatively stable up to 1920.

Figure 6. 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.

4.2 *A regional perspective on age-heaping and literacy in Spain, 1877-1930*

While the previous section provides information at the national level for Spain, we now rely on disaggregated information at the provincial level in order to shed more light into these issues[23]. To begin with, we focus on the topic that closed the previous section, that is, the existence of correlation between age-heaping and literacy rates in the case of Spain. As

[22] It is worth remembering that the ABCC index is computed using information for those aged 23-62 and the literacy rate is the share of the population aged 10 years or more capable of reading and writing.
[23] 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 6 censuses, our data set has been constructed on the basis of 11,760 observations. In addition, we also collected data by gender (which increases our total sample to 35,280 observations).

mentioned there, comparing age-heaping with literacy should be made with individual-data. In the United States, A'Hearn *et al.* (2009), using individual-data from population censuses, confirm the existence of a positive and statistically significant correlation between literacy and the probability of reporting a heaped age[24]. Still, not only individual characteristics seem relevant, the average literacy rate of individuals' birth region also plays a role. Likewise, A'Hearn *et al.* (2016) explore this relationship with Italian censuses. Although the correlation broadly emerges, they find some inconsistencies that challenge the conventional interpretation of age-heaping as a proxy for numeracy skills.

As mentioned above, Spain's population censuses report information at the province-level[25]. Following A'Hearn *et al.* (2009), we then study the relationship between age-heaping and literacy at this level for each census between 1877 and 1930[26]. Using 49 provinces as a unit of analysis, the correlation coefficients between the ABCC indices and literacy rates range from 0.586 and 0.707, whereas the spearman rank correlation goes from 0.649 to 0.770. Thus, and in line with the existing evidence, there is a reasonably strong and statistically significant relationship between these two variables. Additionally, the 1887, 1900 and 1910 censuses reported literacy rates by gender and age-group (21-30, 31-40, 41-50, 51-60). When ABCC indices are computed for these age-groups and correlated with the corresponding literacy rates, further evidence is found in support of the relationship between these two variables[27].

Figure 7 shows ABCC indices and literacy rates for the Spanish provinces (N=49) between 1877 and 1930. While the solid line is the regression line, the connected black-squares display the national average in each of the six censuses. At first glance, there appears to be a rather strong correlation between age-heaping and literacy. Still, figure 7 also illustrates that while literacy improved rapidly, age-heaping remained relatively unchanged until 1920 and 1930[28]. Furthermore, there is substantial noise or dispersion within Spain, thereby calling for a closer examination.

---

[24] 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).

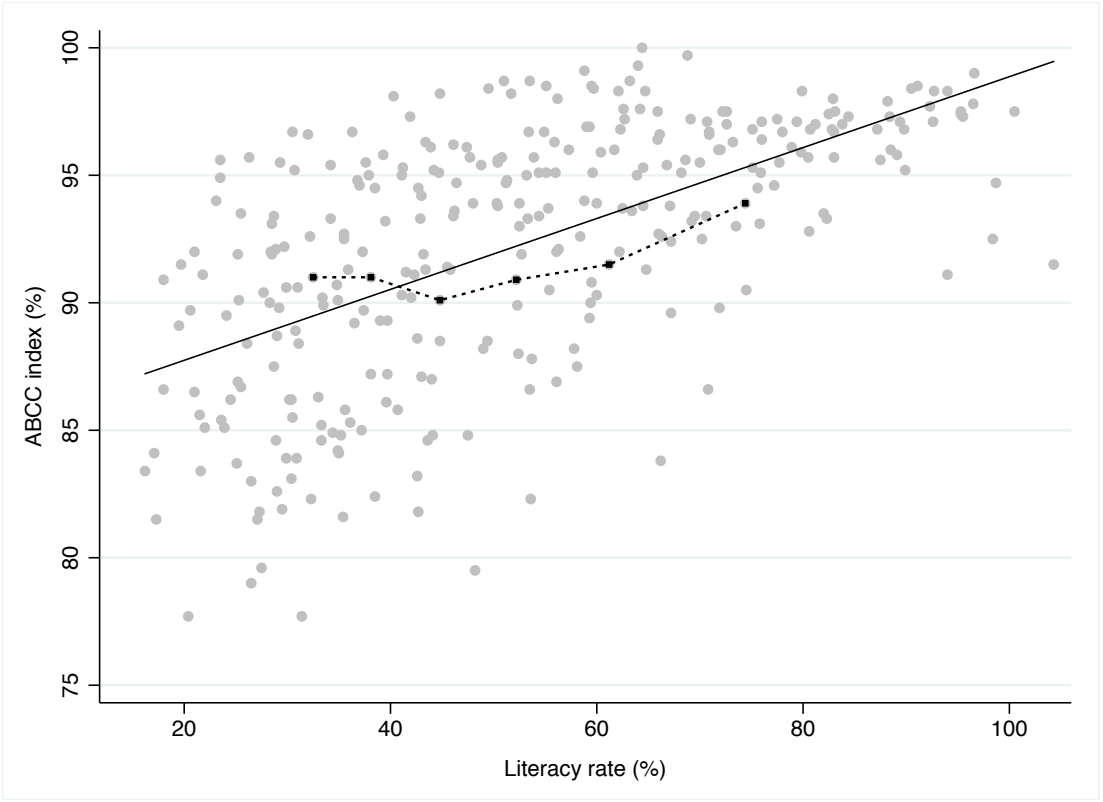
[25] In some cases, information is also provided for large municipalities. Alternatively, individual-data is reported in the *Padrones municipales*. See, for instance, Beltrán Tapia and Miguel Salanova (2017).

[26] As before, the differences in the information used to compute the ABCC index (those aged 23-62) and the literacy rate (the share of the population aged 10 years or more capable of reading and writing), should be considered.

[27] Figure A.1 in the Appendix shows provincial ABCC indices and literacy rates for males aged 21-30 in the censuses of 1887 and 1910.

[28] In the appendix, figure A.2 illustrates the relationship between age-heaping and literacy across provinces for each population census between 1877 and 1930.

Figure 7. ABCC index (%) and literacy in Spain by province, 1877-1930.



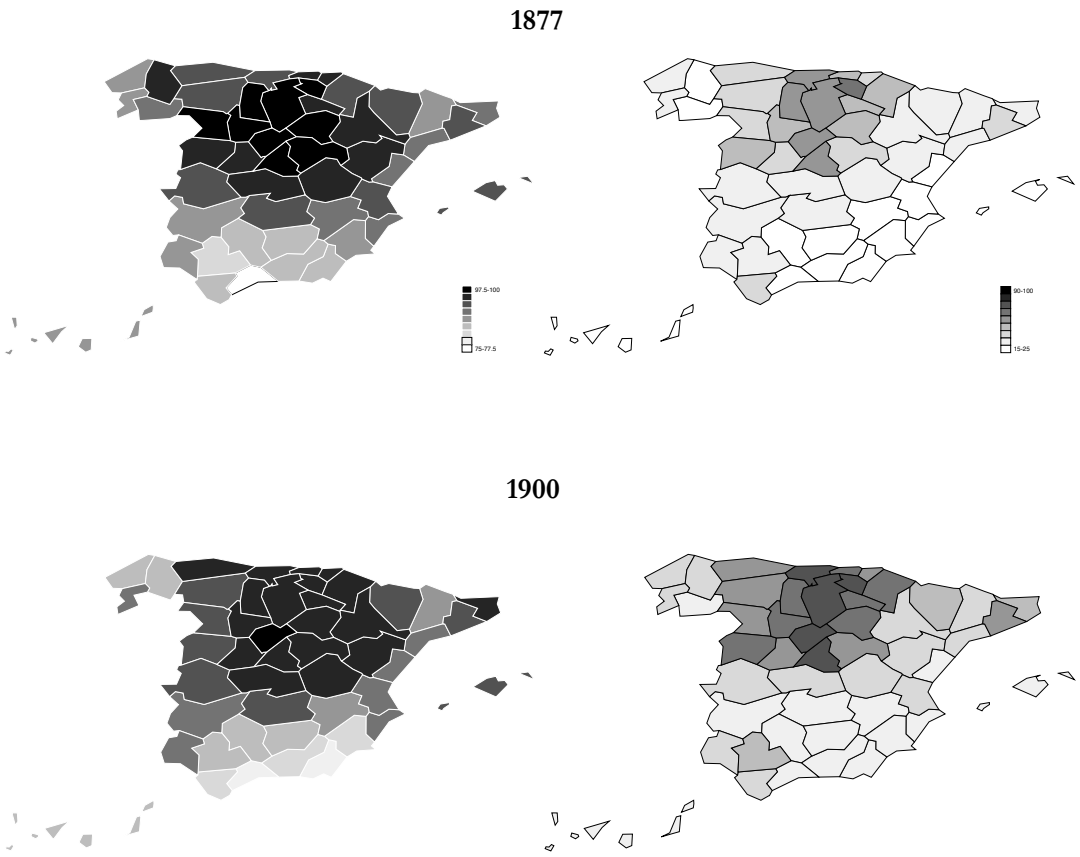
Notes: Data for the following population censuses: 1877, 1887, 1900, 1910, 1920 and 1930. The solid line represents the regression line, while the black-squares are the national average.
Source: INE and authors' calculations.

In addition, when we compare our findings with previous studies, age-heaping in Spain seems to be less prevalent than in United States or Italy. Although literacy rates in Italy and Spain were similarly poor, below 50 percent in the nineteenth century, age-heaping in the Spanish census of 1877 is consistently lower than the one reported in the Italian census of 1871. Even more, the ABCC index never falls below 90 percent, which is not far away from the age-heaping found in the United States census of 1870. In other words, Spain in that period had very low literacy rates and yet age-heaping was relatively low. This is, to some extent, puzzling and casts doubt on the complementarity between age-heaping and literacy. Plus, not only levels are puzzling, also the evolution of these two indicators is difficult to reconcile. If literacy rates were increasing since the mid-nineteenth century, one would expect that ABCC levels would also improve, but that was not the case: they remained stable up to the 1920s.

In order to shed more light into this issue, we investigate the spatial dimension. Figure 8 shows a series of maps that describe the spatial distribution of age-heaping and literacy in

1877, 1900 and 1930. In general, there appears to be a core, namely the centre and centre-north, where literacy was relatively high while age-heaping was practically inexistent. If, to begin with, we focus on the year 1877, we see that in Burgos, Guadalajara, Madrid, Segovia, Soria, and Valladolid, the ABCC index was above 98 percent, thereby implying that less than 2% of individuals reported their age incorrectly[29]. This level of accuracy would place these provinces with the most advanced societies in Europe[30]. 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. Although relatively high within the Spanish context, these figures would be rather modest in international perspective[31].

Figure 8. ABCC index (left) and literacy rate (right) in Spain, 1877-1930.



[29] 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".
[30] 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.
[31] In Spain, around 32.5% of those aged 10 or more years were capable of reading and writing in 1877. These values were far from the literacy rates reached in the main European countries by 1870: France (69%), UK (76%) and Germany (80%). Crafts (1997).

1930

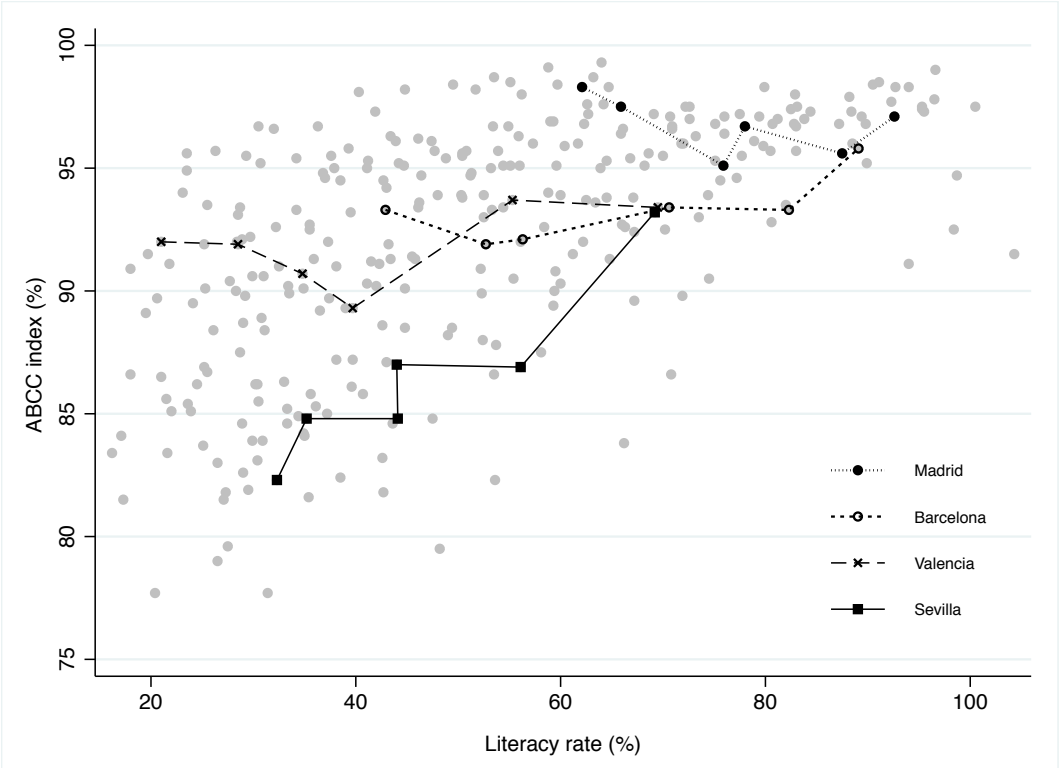
Notes: Provincial ABCC index (0-100) and literacy rates (%) classified into 9 equivalent categories for all censuses. In both cases the range unchanged over the period. Then, the ABCC indices range from 75 to 100 percent while literacy rates range from 15 to 100 percent.

Source: INE and authors' calculations.

Age-heaping and literacy were well below the national average in the other regions and especially so in Andalusia, Canary Islands, Galicia and the eastern coast. Consequently, there was a divide within Spain both in age-heaping and literacy, as figure 8 illustrates and previous studies suggested (Núñez, 1992). This is particularly evident on the eve of the Spanish Civil War (1936-1939), when most Northern provinces, apart from the Galician ones, had almost completed the transition to universal literacy. These differences were, nonetheless, not so marked in terms of numeracy, as lower ABCC indices in 1930 mainly concentrated in South-Eastern (and North-Western) Spain. In that year, the ABCC for Almería was 89.4 and Orense registered the lowest value with an ABCC index of 83.8.

To delve further into the spatial patterns observed, figure 9 displays the trajectories followed by four selected provinces between 1877 and 1930: Madrid, Barcelona, Valencia and Seville. These provinces contain the largest cities in Spain. All things considered, figure 9 presents interesting facts. On the one hand, in Madrid, the capital-province, age-heaping was practically inexistent even at the beginning of our period of analysis. Literacy was also relatively high, around 60% in the earliest count, and improved thereafter. However, as figure 9 displays, Madrid was not among the most advanced provinces. In Burgos, for example, age-heaping was practically inexistent, ranging from 96.8 to 98.5 percent, while literacy dramatically improved from 55.1 percent in 1877 to universal literacy (96.5 percent) in 1930. In these two cases, in spite of a modest literacy rate in the early years, age-heaping was virtually absent, which is somehow anomalous.

Figure 9. ABCC index (%) and literacy in Spain by province, 1877-1930.



Source: INE and authors' calculations.
Note: Marked dots correspond to the census years 1877, 1887, 1900, 1910, 1920, and 1930.

In Barcelona, however, literacy and numeracy rates were initially far from impressive but, while literacy increased throughout the whole period, the ABCC index remained stagnant up to the 1920 census. As regards to Valencia, the literacy rate in 1877 was very low (close to 20%) while the ABCC index was not far away from that of Barcelona where more than 40% of men and women could read and write. Then, the ABCC remained rather stable whereas literacy rates more than tripled by 1930. Finally, in Seville both measures were initially low but then increased rapidly hand in hand. All in all, these mixed and diverse stories present a snapshot of the diverging relationship between age-heaping and literacy in Spain.

An important feature absent in our analysis are internal migrations, which could affect the stock of human capital. In this regard, Beltrán Tapia and 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. Other destinations, however, attracted fewer literate migrants, especially those moving from rural areas to the provincial capital, often resulting in negative self-selection. This is indeed relevant, since the type (and importance) of migratory flows would influence the stock of human capital both at

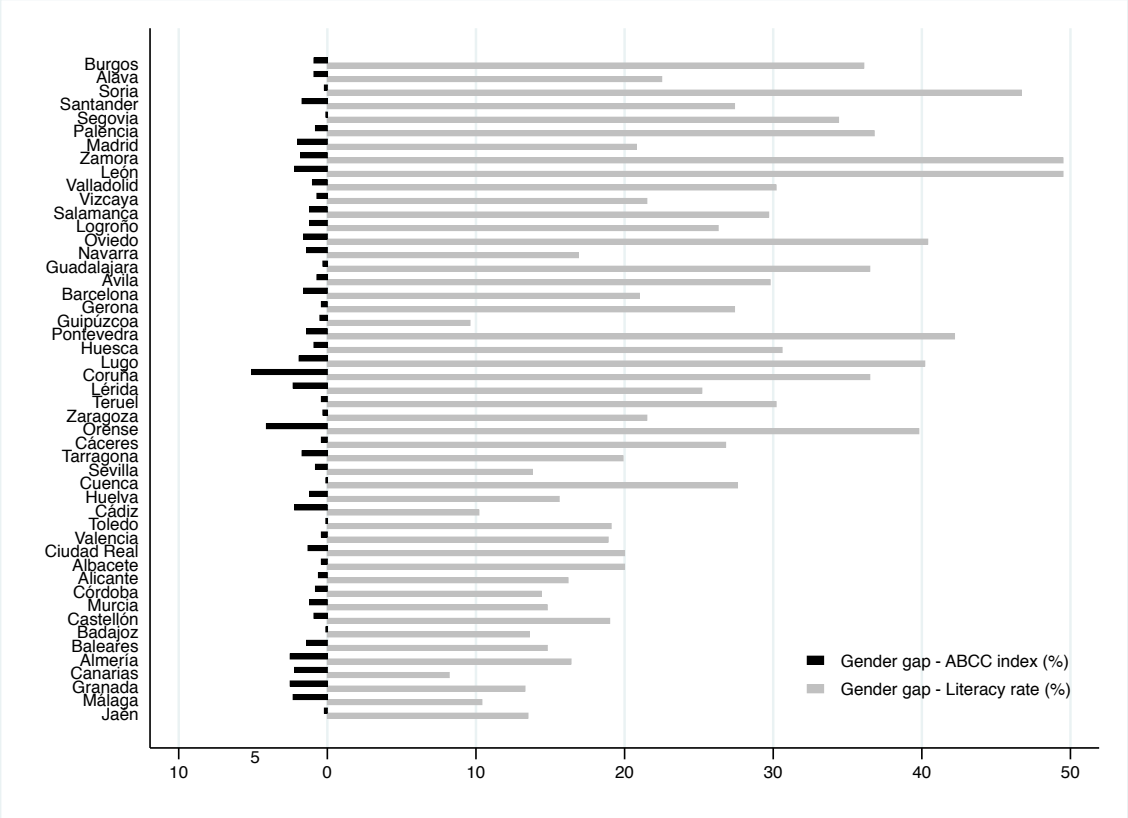
sending and receiving regions. We acknowledge the potential relevance of internal migrations, but, on the one hand, this approach just aims at describing age-heaping and literacy in Spain, and on the other, although internal migrations have been recurrent in Spanish history, permanent internal migrations remained low up to the 1920s (Silvestre, 2001, 2005).

To sum up, despite similar spatial patterns, the disparities mentioned above between the absolute levels of age-heaping and literacy are striking. This anomalous behaviour is even more acute when examining this information by gender. Whereas gender differences in age-heaping were insignificant, female and male literacy rates differed widely. Using information from the census of 1900, figure 10 shows the absolute gender gap in age-heaping and literacy within Spain[32]. In general, gender differences in literacy were vast while age-heaping for males and females appears to be rather similar. Although the small or even inexistent gender gap in age heaping has already been found in other contexts (De Moor and Van Zanden, 2010), and in spite of the nature (range) of each measure, this evidence, which is consistent over the period of study, calls for a careful discussion[33].

Some selected examples may illustrate this different result observed in the gender gap for age-heaping and literacy rates. In Burgos, for instance, the ABCC index for men and women in 1900 were 97.2% and 96.3% respectively, but only about half of all women could read and write. In Soria, 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%. Finally, in Madrid, the male and female literacy rates were 76.3% and 48.3%; ABCC indexes were 98.1 and 96.2, respectively. In addition, if we examine age-heaping for those aged 23-32, the gender gap fades. Likewise, if we explore the correlation between age-heaping and literacy by gender, the relationship weakens once data for females are used[34].

[32] Previous studies have constructed an indicator of «gender equality» in order to study this issue (See Manzel and Baten, 2009). In our case, we prefer to show the absolute differences to further clarify matters.
[33] Other studies suggest that this gender gap may be changing over the process of development depicting a U-inverted curve, as in the case of Latin America and Asia (Manzel and Baten, 2009; Friesen, Baten and Prayon, 2012). See Tollnek and Baten (2016) for a recent review.
[34] Figure A.3 in the Appendix shows the correlation between the gender gap in age-heaping and literacy from 1870 to 1930.

Figure 10. Gender gap in age-heaping and literacy in Spain, 1900.



Source: INE and authors' calculations.
Note: Provinces are ranked according the male literacy rate (%) in 1900, being Burgos and Jaén at the very top and bottom of the distribution.

4.3. Discussion

All these findings call for a careful use of age-heaping as proxy of numeracy. In particular, there appears to be some major issues, at least regarding the Spanish experience. First, and connected with the previous paragraph, the lack of gender gaps questions whether the information was self-reported by females. This is a common concern in studies based on census data. Although censuses have the virtue of including all the population and not a subset that might be affected by a selection bias, some authors have questioned their reliability as regards gender measurement. It has been argued that household heads, i.e. males, may be reporting the age of all family members. In this sense, Földvári et al. (2012) point that to properly measure female age-heaping only the information reported by unmarried woman should be used. In a recent paper, Blum et al. (2017), studying the case of Ireland in the nineteenth century, find that actually married woman showed higher ABCC indexes than unmarried woman concluding that male household heads reported the

age of their wives[35]. Thus, our estimates for female age-heaping may be suffering from this bias, given that in the particular case of Spain the elaboration of the censuses of population relied on family cards (*cédulas de inscripción familiares*)[36]. This may explain the absence of a gender gap in age-heaping in Spain in a period of large gender gaps in literacy rates.

Still, given the widespread illiteracy, age-heaping for the whole sample is abnormally low, which casts doubt on the capacity of the public administration to accurately carry out the data-collection process. Indeed, while it is worth stressing that age-heaping has been generally associated with ignorance or low numeracy skills of the respondents, it may also result from the enumerator (A'Hearn *et al.* 2009). Under these circumstances, the capacity of an administration to carry out such a vast endeavour needs to be considered. Concerning Spain, two elements are worth mentioning. First, the elaboration of the early Censuses of Population was in hands of the *Dirección General del Instituto Geográfico y Estadístico*.[37] The collection of information for the population statistics were coordinated from the headquarters in Madrid and conducted by the *Juntas Censales* (*provinciales* and *municipales*). At the end of the chain, the census-takers counted with the collaboration of both the local priests and primary teachers.

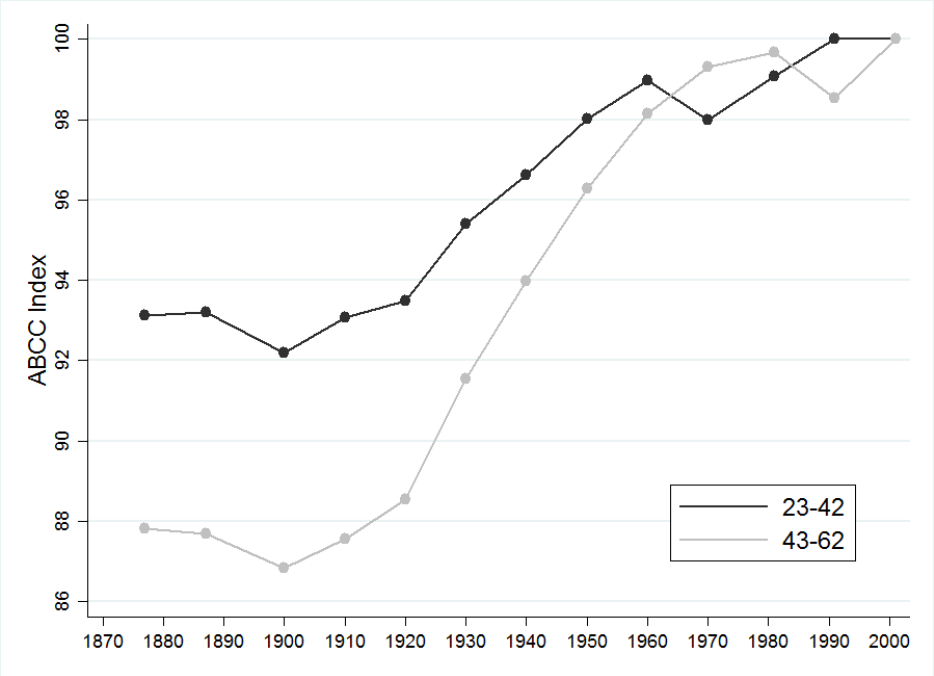
Second, the endeavour of conducting a population census and its outcome in the production of the information was probably affected by the low population density of the country and the different settlement patterns that characterise Spain. In the south, the counts had to face the existence of a relatively small number of large municipalities, such as the typical agro-towns. Contrarily, in the centre and north of the country, a very large number of small villages predominated.[38] It can be argued that the task of counting people was "easier" in small towns due to both the lower number of population to count and to the greater availability of time of the municipal employees leading to a higher accuracy in the responses.[39]

---

[35] They compare the results from the Census with those obtained from two alternative sources: prison and workhouse registers. They conclude that ABCC indexes overestimate age-heaping in the census by 26 points.
[36] For a sample of the registration cards employed, see Figure A.4 in the Appendix.
[37] The *Instituto Geográfico y Estadístico* was created in September 1870 although it adopted this name in June 1873. This institution was established to perform the task of collecting information to create the national statistics. In 1890, it was included in the *Ministerio de Fomento* and after 1900s it depended from different ministries.
[38] In Galicia and Asturias, in the north west of Spain, not only a large number of municipalities exist but also a relevant number of entities within municipalities.
[39] See Reher et al. (1993).

As a result, the information on age heaping based on Population Censuses may be pointing to the existence of limited numerical abilities, but also to a limited capacity of the state to collect the data. In this context, the accuracy of the census takers in performing their task is difficult to ascertain. It is recognised though that older people have more problems to remember their exact age, so it is expected that rounding numbers would be more pronounced as people gets older[40]. In order to explore this issue, we have divided our sample into two large groups by age: those between 23-42 and 43-62 years and calculated ABCC indices. Figure 11 shows the results. The larger age-heaping observed in the older generations, particularly in the early counts when age-heaping was more prevalent, suggests that public employees did well their job and makes us more confident about the capacity of the state to collect information and therefore about the accuracy of our age-heaping indicators.

Figure 11. ABCC indexes by two large age-groups, Spain 1877-2001.



Source: INE and authors' calculations.

[40] Yet, wealthier individuals tend to have both longer lives and higher educational levels, potentially offsetting the fact that elder people are more prone to round their age.

Moreover, A'Hearn *et al.* (2016) have 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. They define these factors as modernisation of the economy in a wide sense, including "*cultural change in the direction of secularism, individualism, and linear time-perception*" and, interestingly, also as an institutional change related with the state's administrative capacity to conduct the task of elaboration the Population Censuses.

Finally, the high ABCC indexes obtained for Spain in a period of widespread illiteracy shows that although both measures are related, there are also relevant differences between them. It is possible to hypothesise that while learning to read and write required a large investment in terms of resources, cost of opportunity, effort and time, numerical abilities could be acquired more easily. Indeed, numerical abilities may be obtained in a more intuitive way in daily life and work, especially when individuals are surrounded by other people that possess and use that ability. In this context, in-job training appears as a key factor for the acquisition of numerical activities, in particular, and human capital, in general. In the case of Spain, Rosés (1998) stressed the relevance of in-job training during the Catalan industrialization around mid-nineteenth century, a period characterized by modest literacy rates in Catalonia within the Spanish context. In addition, Nadal (1996) claimed that in the Catalan industrialization the role of technical schools (*liceus* and *escoles d'arts i oficis*) provided the required skills to the factory workers since the mid-nineteenth century. Hence, technical education rather than literacy rates would have played a key role in the spur of Catalan industrialization[41].

Anyhow, it is indeed difficult to disentangle the different sources behind the observed levels of age heaping in Spain. In particular, and to sum up, some concerns arise as to whether age heaping results from: a) low numeracy skills of the respondents, b) a poorly designed and implemented data collection process, c) a general modernization of the economy, or d) the less demanding effort required to obtain numerical abilities compared to literacy. While Blum and Krauss (2017) suggest a direct link to human capital, all these issues recommend caution in the census-data based analysis carried out for Spain and, importantly, open new avenues for further research.

[41] A similar argument can be found for Italy (Zamagni, 1993). As regards Britain, Zeev et al. (2017) find that in the eighteenth-century apprenticeship had a key contribution to an advantage in skilled mechanical labor.

5. 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 of it 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, the Spanish population censuses do not exhibit a preference for ages ending in 5. Thus, age-heaping emerges due to a preference for multiples 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). In any case, we follow the literature and use the traditional approach, for the sake of comparability, in the descriptive analysis.

Second, age-heaping, either for young adults (aged 23-32) or for the whole distribution (aged 23-62), was less relevant in the mid-nineteenth century than previous studies have suggested. Furthermore, age-heaping was relatively stable in the early censuses of the second half of the nineteenth-century. In fact, it did only improve significantly from 1920 onwards.

Third, relying on disaggregated information at the provincial level, literacy rates and the extent of age-heaping show relatively high levels of correlation. However, the level of age-heaping, both at national and provincial levels, was surprisingly low, especially compared to the rampant 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. Likewise, the rate of age-heaping in Spain remained stable up to 1920 despite the gradual improvement of literacy since the mid-nineteenth century. These results are difficult to reconcile with the widespread consideration of age-heaping as a proxy of numeracy. This conclusion is further (cautiously) confirmed by examining gender gaps: disparities in age-heaping between men and women were negligible even in provinces where the gender literacy gaps were noteworthy.

In sum, the study of age-heaping in Spain provides further insights into this subject, but also casts doubt on the extent to which digit preference proxies numeracy skills, and henceforth human capital. If the census-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, in the capital city of Madrid, and in the 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 needed because the age-heaping and literacy stories should in the end converge towards the same fate.

Acknowledgements

This paper was presented at the III Seminar in Economic History at the Bank of Spain in Madrid. We would like the participants for the comments received. This research is an outcome of the project 'Inequality, human capital and regional economic growth in Spain, 1787-1930', financed by the Bank of Spain (*Programa de ayudas a la investigación en macroeconomía, economía monetaria, finaciera y bancaria e historia económica, 2016-17*). We are also grateful for the financial support from the HAR2015-64076-P, ECO2015-65582, ECO2015-65049-C12-1-P and ECO2015-71534-REDT projects (*Ministerio de Economía y Competitividad*).

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. 2003. "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 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.

Blum, M. and Krauss, K.P. 2017. "Age heaping and numeracy: looking behind the curtain", *The Economic History Review* (forthcoming).

Blum, M., Colvin, C.L., McAtackney, L. and McLaughlin, E. 2017. "Women of an uncertain age: quantifying human capital accumulation in rural Ireland in the nineteenth century", *The Economic History Review*, 70 (1), 187-223.

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.

Calatayud, S., Millán, J. and Romeo M.C. 2016. "Introducción: la formación del Estado desde abajo. Poder, sociedad y conflictos en la España del siglo XIX". In S. Calatayud, J Millán and M.C. Romeo (eds), *El Estado desde la sociedad. Espacios de poder en la España del siglo XIX*, Alacant: Publicacions de la Universitat de Alacant, 9-40.

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

Crafts, N.F.R. 1997. "The Human Development Index and changes in standard of living: some historical comparisons", *European Review of Economic History*, 1, 299-322.

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.

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.

Földvári, P., Van Leeuwen, B. and Van Leeuwen-Li, J. 2012. "How did women count? A note on gender-specific age heaping differences in the sixteenth to nineteenth centuries", *The Economic History Review*, 65 (1), 304-313.

Friesen, J., Baten, J. and Prayon V. 2012. Women count. Gender (in-)equalities in the human capitl development in Asia, 1900-60, *University of Tübingen Working Papers in Economics and Finance*, No 29.

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.

Goldin, C. 2015. "Human capital". In C. Diebolt and M. Haupert (eds.), *Handbook of Cliometrics*, ch. 3, Berlin and Heidelberg: Springer, 55-86.

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).

Hobsbawm, E. 1962. *The age of revolution, 1789-1848*, New York: Vintage Books.

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.

Mankiw, N.G., Romer, D. and Weil D.N. 1992. "A contribution to the empirics of economic growth", *The Quarterly Journal of Economics* 107 (2), 407-437.

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

Manzel, K. and Baten, J. 2009. "Gender equality and inequality in numeracy: the case of Latin American and the Caribbean, 1880-1949", *Revista de Historia Económica / Journal of Iberian and Latin American Economic History*, 27 (1), 37-74.

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.

McCloskey, D.N. 2010. *Burgeois dignity: why economics can't explain the modern world*, Chicago: University of Chicago Press.

Mitch, D. 1993. "The role of human capital in the First Industrial Revolution". In J. Mokyr (ed.), *The British Industrial Revolution: an economic perspective*, Boulder: Westview, 267-307.

Mitch, D. 1999. "The role of education and skill in the British Industrial Revolution". In J. Mokyr (ed.), *The British Industrial Revolution: an economic perspective*, 2nd ed., Boulder: Westview Press, 241-279.

Mitch, D. 2004. "Education and the skill of the British labor force". In R.C. Floud and P. Johnson (eds.), *The Cambridge Economic History of Britain, 1700–2000*, vol. 1, Cambridge: Cambridge University Press, 332-356.

Mokyr, J. 2009. *The Enlightened economy: an economic history of Britain 1700-1850*, New Haven: Yale University Press.

Mokyr, J. and Voth, J. 2010. "Understanding growth in Europe, 1700-1870: theory and evidence". In S. Broadberry and K. H. O'Rourke (eds.), *The Cambridge Economic History of Europe*, vol. 1, Cambridge: Cambridge University Press, 7-42.

Nadal, J. 1996. "El factor humà en el retard econòmic espanyol. El debat entre els historiadors", *Revista Económica de Catalunya*, 29, 9-14

Nicolau, R. 2005. "Población, salud y actividad". In A. Carreras and X. Tafunell (eds.), *Estadísticas históricas de España: siglos XIX-XX*, Bilbao: Fundación BBVA, 77-154.

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. 1991. "El gasto público en educación entre 1860-1935", *Hacienda Pública Española*, 1, 121-146.

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.

Núñez, C.E. 2005. "A modern human capital stock. Spain in the nineteenth and twentieth centuries". In M. Jerneck, M. Morner, G. Tortella and S. Akerman (eds.), *Different paths of modernity. A Nordic and Spanish perspective*, Lund: Nordic Academic Press, 122-142.

Pamuk, S. and Van Zanden, J.L. 2010. "Standards of living". In S. Broadberry and K.H. O'Rourke (eds.), *The Cambridge Economic History of Modern Europe, volume 1: 1700-1870*, ch. 9, Cambridge: Cambridge University Press, 217-234.

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*.

Reher, D.S., Nogueras, B. and Pombo, M.N. 1993. *España a la luz del Censo de 1887*, INE: Madrid.

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. 1998. "Measuring the contribution of human capital to the development of the Catalan factory system (1830-1861)", *European Review of Economic History*, 2 (1), 25-48.

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.

Zamagni, V. 1993. "Instrucción y desarrollo económico en Italia, 1861-1913". In C.E. Núñez and G. Tortella (eds.), *La maldición divina: ignorancia y atraso económico en perspectiva histórica*, Madrid: Alianza, 181-222.

Zeev, N.B., Mokyr, J. and Van Der Beek, K. 2017. "Flexible supply of apprenticeship in the British Industrial Revolution", *The Journal of Economic History*, 77 (1), 208-250.

APPENDIX: TABLES and FIGURES.

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	8,655	132	4,189	143	4,466
23-72	136		132		140	

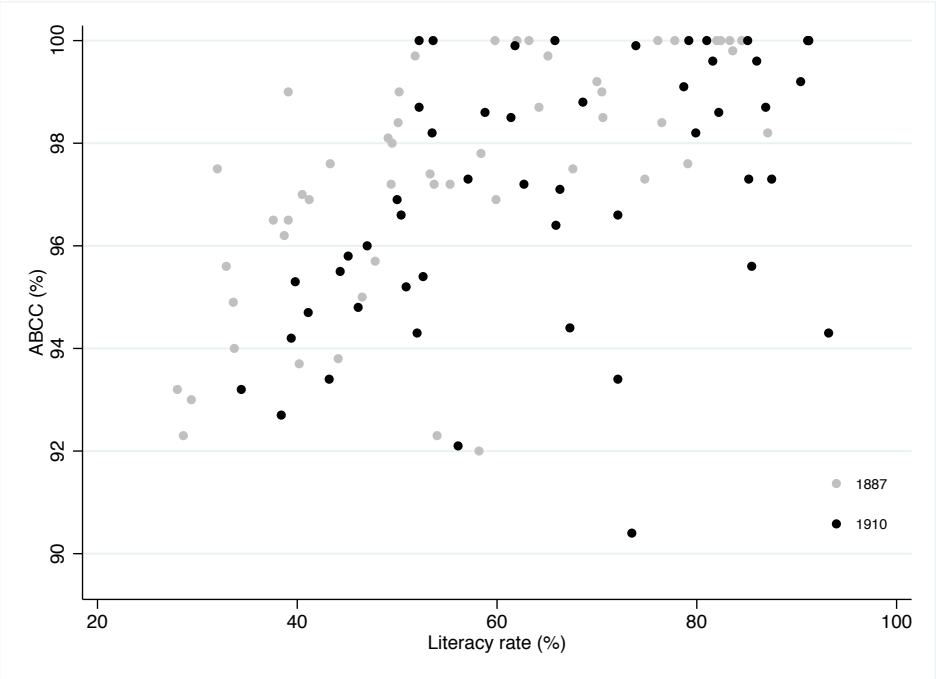
Notes: Number of observations in thousands
Source: INE and authors' calculations.

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	13,743	144	6,892	159	6,851
23-72	147		140		154	

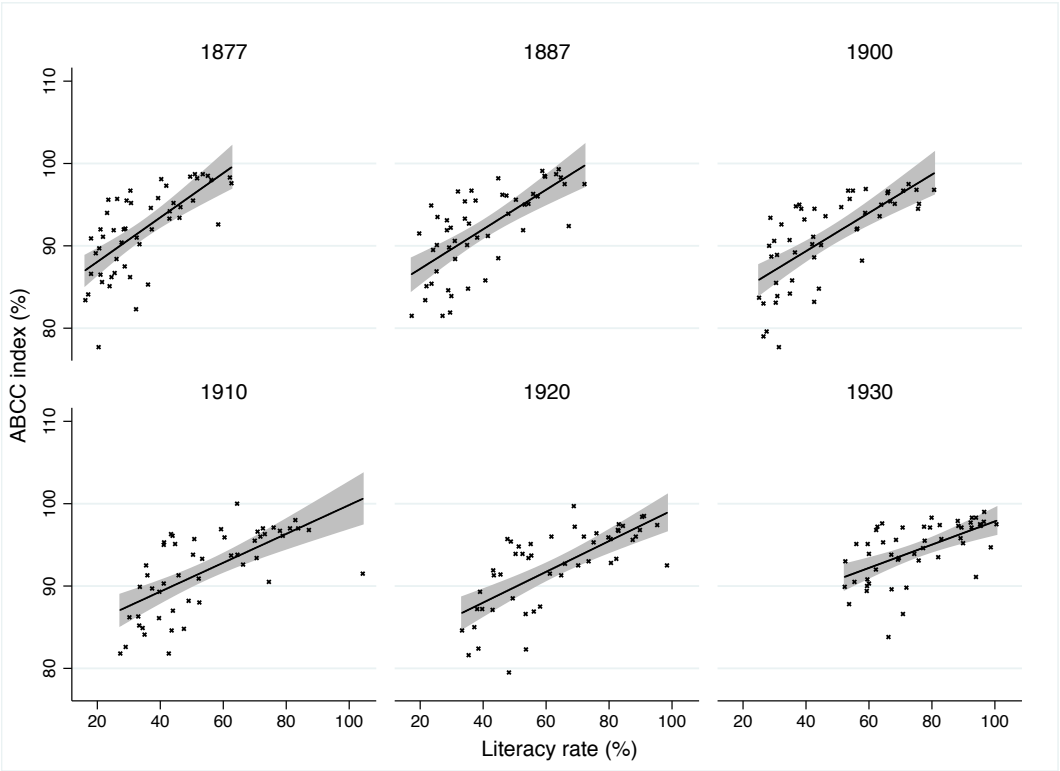
Notes: Number of observations in thousands
Source: A'Hearn *et al.* (2016: Table 2.1)

Figure A.1 Male age-heaping and literacy by age-group (21-30) in Spain, 1887 and 1910.



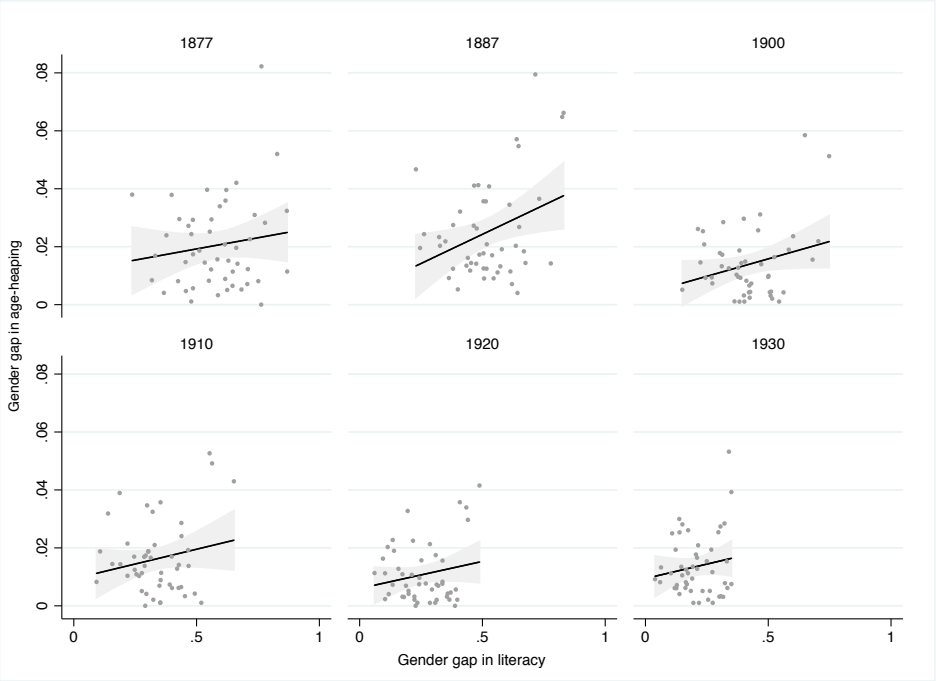
Notes: Each dot represents a province in 1887 and 1910.
Source: INE and authors' calculations.

Figure A.2 Provincial age-heaping and literacy in Spain, 1877-1930.



Notes: Each line represents a linear fit with a 5% confidence interval.
Source: INE and authors' calculations.

Figure A.3 Gender gap in age-heaping and literacy within Spain, 1877-1930.



Notes: Data for the following population censuses: 1877, 1887, 1900, 1910, 1920, and 1930. The gender gap in age-heaping is measured as $(\text{ABCC}^{male} - \text{ABCC}^{female}/\text{ABCC}^{male}) * 100$ while as regards literacy the gap is measured as $(\text{Literacy}^{male} - \text{Literacy}^{female}/\text{Literacy}^{male}) * 100$

Source: INE and authors' calculations.

DIRECCIÓN GENERAL DEL INSTITUTO GEOGRÁFICO Y ESTADÍSTICO.

CÉDULA NÚM._____

CENSO DE POBLACIÓN.

Provincia de_____ **Partido judicial de**_____ **Distrito municipal de**_____

Pueblo de_____ **Sección núm.**_____ **nombre**_____ **Calle de**_____ **Casa núm.**_____ **piso**_____

Cédula de inscripción que, para la formación del Censo general, presenta D._____ *de esta vecindad, como cabeza de familia, de todas las personas correspondientes á la misma, presentes y temporalmente ausentes, y de las demás que pernoctaron en su casa la noche del 31 de Diciembre de 1877 al 1.º de Enero de 1878.*

Número de orden.	NOMBRES Y APELLIDOS.	SEXO.	EDAD. Años.	EDAD. Meses.	EDAD. Días.	ESTADO CIVIL.	Parentesco ó razón de conveniencia con el cabeza de familia.	INSTRUCCIÓN ELEMENTAL. ¿Sabe leer?	INSTRUCCIÓN ELEMENTAL. ¿Sabe escribir?	RELIGIÓN.	Defectos físicos notorios.	NATURALEZA. Para los españoles. Pueblo.	NATURALEZA. Para los españoles. Provincia.	NATURALEZA. Para los extranjeros. Nación.	Condición de su residencia en este pueblo.	TIEMPO DE RESIDENCIA EN ESTE PUEBLO. Años.	TIEMPO DE RESIDENCIA EN ESTE PUEBLO. Meses.	TIEMPO DE RESIDENCIA EN ESTE PUEBLO. Días.	Profesión, oficio, ocupación ó posición social.	PUNTOS EN QUE LOS AUSENTES SE ENCUENTRAN. De España. Pueblo.	PUNTOS EN QUE LOS AUSENTES SE ENCUENTRAN. De España. Provincia.	PUNTOS EN QUE LOS AUSENTES SE ENCUENTRAN. Del extranjero. Nación.	VECINDAD Ó DOMICILIO LEGAL DE LOS TRANSEUNTES. En España. Pueblo.	VECINDAD Ó DOMICILIO LEGAL DE LOS TRANSEUNTES. En España. Provincia.	VECINDAD Ó DOMICILIO LEGAL DE LOS TRANSEUNTES. En el extranjero. Nación.	OBSERVACIONES.
1																										
2																										
3																										
4																										
5																										
6																										
7																										
8																										

La inscripción se hará por el orden siguiente: 1.º el cabeza de familia, su mujer, hijos, parientes y deudos; 2.º los ayos, secretarios, dependientes, criados y demás personas que vivan en su compañía, y 3.º los que accidentalmente se encontraren en la casa. Cada una de estas secciones se cortará á su final por medio de una raya. Se consignarán los dos apellidos de cada individuo; si sólo se supiese uno, se expresará éste, y si se ignorasen ambos, se marcará una cruz á continuación del nombre. Si la persona es de padres desconocidos, se pondrá *expósito* en lugar de los apellidos. A los ausentes se les señalará á continuación de los apellidos con una *A*, á los extranjeros con una *E* y á los transeuntes con una *T*.

Se indicará el sexo con las abreviaturas *V.* para el varon, y *Hra.* para el femenino.

La edad se expresará por años cumplidos. Para los niños que el dia de la inscripción no hayan cumplido un año se hará por meses, y para los que no tengan un mes por días.

En esta casilla se hará constar si el inscripto es soltero, casado ó viudo.

Se expresará, en el que no sea pariente, si es ayo, escribiente, administrador ó dependiente, criado, etc., y si es huesped ó vive en familia.

Por medio de las particulas Sí y No se manifestará en la casilla respectiva la instrucción que se posea ó la carencia de ella.

Se expresará la religión á que pertenece cada uno de los individuos inscriptos en esta cédula.

Sólo en los que sean ciegos, sordomudos, lixiados, dementes ó locos, idiotas ó bobos se hará constar, añadiendo si el defecto ó defectos son de nacimiento ó adquiridos.

Se consignará en estas casillas el pueblo y provincia, ó nación, si fuese extranjero, en que nació cada uno de los individuos que figuren en la cédula.

Esta condición se fijará con arreglo á lo prescrito en los artículos 10 y 11 de la Ley municipal de 20 de *Agosto* de 1870 insertos al pie *. Los extranjeros dirán además si se encuentran ó no naturalizados.

Este dato se expresará por años cumplidos: los que no lleven un año por meses, y los que no lleven un mes, por días.

El que ejerza varias profesiones, las hará constar todas ellas, comenzando por la que le produzca mas por utilidad. En las artes y oficios se expresará si es maestro, oficial ó aprendiz.

Cuando se ignore el paradero de las personas ausentes de su casa el día de la inscripción, se pondrá el que se crea más probable.

Estas casillas se llenarán con arreglo á lo dispuesto en los párrafos 1.º y 2.º del artículo 11 de la Ley municipal antes citada.

En esta casilla se consignará todo lo que sirva para aclarar cualquier concepto dudoso de la cédula ó ilustrarla sobre algunos extremos, como por ejemplo: la causa de la ausencia, el número de mujeres contraidas por los casados y viudos, cuando sea más de una, la circunstancia de estar separado ó divorciado el cabeza de familia, &., &.

* LEY MUNICIPAL. { Art. 10. Los habitantes de un término municipal se dividen en residentes y transeuntes. Los residentes se subdividen en vecinos y domiciliados.
Art. 11. Es *vecino* todo español emancipado que reside habitualmente en un término municipal y se halla inscripto con tal carácter en el padron del pueblo. Es *domiciliado* todo español, que sin estar emancipado, reside habitualmente en el término formando parte de la casa ó familia de un vecino. Es *transeunte* todo el que no estando comprendido en los párrafos anteriores, se encuentra en el término accidentalmente.

Figure A.4 Registration card (*cédula de inscripción*), Population Census of 1877.

European Historical Economics Society

EHES Working Paper Series

Recent EHES Working Papers

2018

EHES 138 Two Worlds of Female Labour: Gender Wage Inequality in Western Europe, 1300-1800
Alexandra M. de Pleijt and Jan Luiten van Zanden

EHES 137 From Convergence to Divergence: Portuguese Economic Growth
Nuno Palma and Jaime Reis

EHES 136 The Big Bang: Stock Market Capitalization in the Long Run
Dmitry Kuvshinov and Kaspar Zimmermann

EHES 135 The Great Moderation of Grain Price Volatility: Market Integration vs. Climate Change, Germany, 1650–1790
Hakon Albers, Ulrich Pfister, Martin Uebele

EHES 134 The age of mass migration in Latin America
Blanca Sánchez-Alonso

EHES 133 Gravity and Migration before Railways: Evidence from Parisian Prostitutes and Revolutionaries
Morgan Kelly and Cormac Ó Gráda

EHES 132 On the economics of forced labour. Did the employment of Prisoners-of-War depress German coal mining productivity in World War I?
Tobias A. Jopp

EHES 131 Well-being Inequality in the Long Run
Leandro Prados de la Escosura

All papers may be downloaded free of charge from: www.ehes.org
The European Historical Economics Society is concerned with advancing education in European economic history through study of European economies and economic history. The society is registered with the Charity Commissioners of England and Wales number: 1052680