TWO STORIES, ONE FATE: AGE-HEAPING AND LITERACY IN SPAIN, 1877-1930

FRANCISCO J. BELTRÁN TAPIA
Norwegian University of Science and Technology[a]

ALFONSO DÍEZ-MINGUELA
Universitat de València[b]

JULIO MARTINEZ-GALARRAGA
Universitat de València[c]

DANIEL A. TIRADO-FABREGAT
Universitat de València[d]

ABSTRACT

This study looks at human capital in Spain during the early stages of modern economic growth. We have assembled a new dataset for age-heaping and literacy in Spain with information about men and women from six population censuses and forty-nine provinces between 1877 and 1930. Our results show that, although age-heaping was less prevalent during the second half of the 19th century than previously thought, it did not decline until the early 20th century. Given that literacy increased throughout the whole period, our study thus unveils stark differences between age-heaping and literacy, which raises further questions regarding sources, methods and interpretation.

Keywords: Spain, age-heaping, literacy, 19th-century

JEL codes: I25, N01, N9, O15

[a] Department of Modern History and Society, Trondheim, Norway. francisco.beltran.tapia@ntnu.no
[b] Departament d'Anàlisi Econòmica, València, Spain. alfonso.diez@uv.es
[c] Departament d'Anàlisi Econòmica, València, Spain. julio.martinez-galarraga@uv.es
[d] Departament d'Anàlisi Econòmica, València, Spain. Daniel.Tirado@uv.es

Vol. 40, No. 3: 405–438. doi:10.1017/S0212610921000033

RESUMEN

Este trabajo analiza el capital humano en España durante las primeras etapas del crecimiento económico moderno. Para ello, hemos construido una nueva base de datos que recoge el nivel de age-heaping y de alfabetización a nivel provincial para hombres y mujeres a partir de la información contenida en seis censos de población entre 1877 y 1930. Los resultados muestran que durante la segunda mitad del siglo XIX el nivel de age-heaping era menor de lo sugerido en estudios previos, y que éste no comenzó a reducirse de manera significativa hasta el inicio del siglo XX. Dado que la alfabetización aumentó a lo largo de todo el período, este trabajo revela la existencia de marcadas diferencias entre los niveles de age-heaping y de alfabetización, lo cual genera cuestiones relacionadas con las fuentes, la metodología y la interpretación de los indicadores.

Palabras clave: España, age-heaping, alfabetización, siglo XIX

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 the role played by human capital. While some authors claim that human capital (or at least the upper tail of the knowledge distribution) made a significant contribution to economic growth (Mokyr 2009; Mokyr and Voth 2010; Galor 2011), other works are less optimistic (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, such information is scarce or non-existent. Population censuses, for instance, provide information on literacy rates from the mid-19th century but, in order to proxy literacy for previous periods, researchers have often relied on the ability to sign official documents such as marriage registers (Cipolla 1969; Allen 2003; Reis 2005)[1].

[1] There are alternative approaches to capture human capital in the distant past. While Baten and Van Zanden (2008) used *per capita* book production as an indicator of advanced literacy skills,

Still, even literacy is a broad and complex concept (Graff 1987). Though defined as having the ability to read and write, there might be qualitative differences across literates which are unmeasurable, and hence neglected in quantitative analyses. This is particularly relevant when literacy, as a proxy for human capital, is related to economic development (García Abad *et al*. 2007; González Portilla and Urrutikoetxea Lizarraga 2016). In order to shed further light on these issues, much recent literature has proposed the broadening of the number of indicators by constructing historical numeracy indices as a complementary measure of human capital.

It is argued that the ability to understand and work with numbers, or numeracy, might be even more crucial for economic development than literacy (Crayen and Baten 2010a). In order to capture this skill—and thus human capital in a broader sense—a growing number of studies have compiled historical data from several distinct sources to compute the prevalence of age-heaping[2]. Following this line of enquiry, this study first calculates age-heaping in Spain with the information reported in the population censuses. Spanish population censuses provide summary tables with the number of individuals, men and women, in each province by their exact age. Furthermore, information about the ability to read and write of all these individuals is also provided, which permits the computation of literacy rates.

Comparing both proxies allows a better understanding of the evolution of human capital in Spain in a period marked by profound socioeconomic transformations, namely the late 19th and the early 20th century. In this regard, socioeconomic change could have encouraged investment in particular skills which were not so much in demand in a preindustrial, or traditional, society. Similarly, greater numeracy and literacy could also have stimulated technological and socioeconomic change. The analysis of the complex relationship between human capital and economic development requires a conceptual framework. For the Basque Country, for instance, González Portilla and Urrutikoetxea Lizarraga (2016) proposed an evaluation of this relationship in stages. First, a starting point, around 1877-87, which reflected the legacies of the past. Then, the turn of the 20th century, in which some signs of change emerged, and finally the period from 1900 to 1930 in which human capital and economic development went hand in hand.

In this study, we follow a similar approach and contextualise the findings bearing in mind that socioeconomic change occurred more rapidly during the period 1900-30. Our results show that age-heaping remained relatively unchanged until the early 20th century whereas literacy gradually improved, albeit modestly. Furthermore, the level of age-heaping during

Squicciarini and Voigtländer (2015) gathered information about subscriptions to the *Encyclopédie* per 1,000 inhabitants in French cities.

[2] A recent survey of this literature can be found in Tollnek and Baten (2016).

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

the second half of the 19th century seemed, in some cases, abnormally low. Prior to the arrival of modern economic growth, age-heaping was practically inexistent in some territories. Lastly, we find no differences in age-heaping between men and women, which raises concerns about the data collection process given the sizable gender gaps in literacy. Our work also points to another technical issue: in Spain, age-heaping derived from population censuses emerges due to a preference for rounding ages in multiples of 10 (and does not exhibit a preference for ages ending in 5).

The rest of the paper is structured as follows. In section 2, a brief historical background is presented. The concept of age-heaping is defined and discussed in section 3 and the methodology and data are introduced and explained. The descriptive analysis for Spain is presented in section 4, while section 5 focuses on the results obtained for age-heaping at a regional level (provinces). Section 6 provides a discussion of our main findings and section 7 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). From an economic perspective and compared with its European counterparts, Spain underwent a rather slow industrialisation process and it was only after WWI when GDP growth rates showed a substantial increase (Prados de la Escosura 2017). In 1910, Spain was still an agrarian economy, where roughly 66 per cent of the labour force was involved in agriculture (Nicolau 2005).

Still, this overall picture hides substantial regional differences. For example, Catalonia, the Basque country and the capital region of Madrid, made considerable advances. In the first two territories, industrial development and structural change characterised the late 19th and early 20th centuries (Carreras 1990). In Barcelona, the labour force employed in industrial activities increased rapidly between 1877 and 1930 (being close to 60 per cent of the total active population in 1930). Similarly, in the Basque provinces of Guipúzcoa and Vizcaya, industrial employment tripled, reaching almost 40 per cent of total employment. Likewise, this period witnessed the emergence of the capital city of Madrid, which essentially resulted from the expansion of the manufacturing, construction and service sectors. In short, by 1930, while their population represented 11.8, 3.8 and 5.9 per cent of Spain, Catalonia, the Basque Country and Madrid contributed 34.6, 9.2 and 9.3 per cent to the industrial output, respectively (Tirado and Martinez-Galarraga 2008). As a result, these diverging paths led to an upswing in regional income inequality during the early stages of modern economic growth (Rosés *et al*. 2010).

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

In a context of economic backwardness and sociopolitical instability, education was a source of concern for the successive liberal governments throughout the 19th century. The first national law of education (*Ley de Instrucción Pública*, known as the *Moyano Act*) was passed in 1857. The *Moyano Act* established compulsory schooling for children aged 6-9 years, but funding relied on local councils. Primary schooling thus depended on local demands[3]. The lack of or slow progress in terms of literacy witnessed in the second half of the 19th century led to the creation of the *Ministerio de Instrucción Pública y Bellas Artes* in 1900 (Núñez 1991).

Accordingly, literacy rates in 1870 were indeed rather modest. While in Germany or the United Kingdom, around 80 per cent of the adult population could read and write, literacy rates were around 30 per cent in Spain and Italy (Crafts 1997; Pamuk and Van Zanden 2010). This figure is slightly lower if we consider the population census of 1860. In that year, only 26 per cent of the population above 10 years reported that they were able to read and write, an average that hides a significant gap between men and women (Núñez 2005)[4]. Things were not much better regarding years of schooling. In 1870, the average number of years of schooling in Spain was 1.5, well behind other Western European countries such as Switzerland (6.1 years), Germany (5.4), France (4.1) or the United Kingdom (3.6)[5].

This backwardness hides, nonetheless, large regional differences. A dual structure existed with the northern provinces reaching higher literacy rates than those in the south (Núñez 1992). In 1860, literacy rates in provinces such as Almería (south) were only 14 per cent, but in Álava (north) 53 per cent of the population above the age of 10 was able to read and write. It has been argued that the existence of marked regional differences mimicked the unequal structure of access to land and could be one of the main reasons for the uneven development within Spain (Núñez 1992; Núñez and Tortella 1993; Beltrán Tapia 2013; Beltrán Tapia and Martinez-Galarraga 2018)[6].

3. METHODOLOGY, DATA AND SOURCES

Self-reported age regularly appears in parish and military records, tax rolls, civil and legal documents, passenger lists, and population censuses.

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

[4] While male literacy rates were 40 per cent, female rates were only 12 per cent.

[5] Spain ranked number 20, just below Bulgaria and Uruguay. Other southern European countries, nonetheless, such as Greece (1.4), Italy (0.8) or Portugal (0.5), had a more disappointing performance. Data are from the Clio-Infra project.

[6] On the long-term evolution of regional economies, see Martinez-Galarraga *et al.* (2015) and Tirado *et al.* (2016). 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 the mid-19th century.

Age, however, has often been misreported because of custom and tradition, poor numeracy skills, or a badly designed and executed data collection process[7]. Digit preference for numbers ending in 0 and 5, as well as an aversion to 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; Szoltysek *et al*. 2017). In doing so, researchers have used a battery of tools to gauge age-heaping[8].

3.1 Methodology

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 23-62, 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.

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

However, if digit preference exists, then the above expression would not hold. This phenomenon is clearly observed in the Spanish population censuses as Figure 1 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 tend 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.

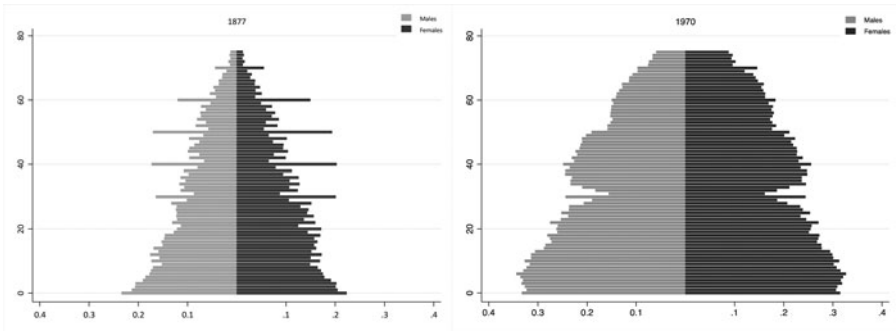
Bearing this in mind, the Whipple index is possibly the most popular indicator to measure preference for ages ending in 0 and 5 (Spoorenberg and Dutreuilh 2007). It can be computed as the ratio of the number of reported ages ending in 0 or 5 to a fifth of the given population. As previously noted, this method assumes that respondents are uniformly distributed as regards age, consequently the oldest and youngest individuals are excluded. The Whipple index is thus 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 final digits. Then, if P_{25} denotes the total number of

[7] 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-year age groups.

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

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

FIGURE 1
POPULATION PYRAMIDS BY AGE (0-75 YEARS OLD) AND SEX IN 1877 AND 1970.



Source: INE.

respondents who reported an age of 25, the Whipple index for multiples of 5 for the range 23-62 years can be computed as follows,

$$W = \left(\frac{(P_{25} + P_{30} + P_{35} + P_{40} + P_{45} + P_{50} + P_{55} + P_{60})}{1/5(P_{23} + P_{24} \ldots + P_{61} + P_{62})} \right) \times 100 \qquad (2)$$

By definition, the Whipple index varies between 100, which indicates no preference for ages ending in 0 or 5, and 500, which implies perfect heaping[9]. In this way, however, we neglect the fact that age preference might occur in other terminal digits. Besides, this approach assumes uniformity over an age range of 10 years, which might be 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:

$$\tilde{W} = \sum_{i=0}^{9} (|W_i - 1|) \qquad (3)$$

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

$$W_0 = \frac{(P_{30} + P_{40} + P_{50} + P_{60})}{1/5(P_{30}^5 + P_{40}^5 + P_{50}^5 + P_{60}^5)} \qquad (4)$$

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

[9] The Whipple index is usually presented or expressed in the range 100-500. For this reason, equation (1) is multiplied by 100.

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

where P_{29} stands for the number of respondents who reported an age of 29 while P_{29}^{5} is the population of those aged 27-31. 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. Using information on preference for each individual digit, 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.

3.2 Data and Sources

In Spain, population censuses have been carried out on a regular basis since the mid-19th century[10]. Since 1877, the information on the number of men and women by exact age has been summarised at the provincial level. In this study, we have thus gathered provincial data from the following censuses: 1877, 1887, 1900, 1910, 1920, 1930, 1940, 1970, 1991 and 2001. Similarly, censuses also report the number of men and women who could read and write, thereby allowing for the computation of literacy rates. Sadly, literacy is not reported by exact age, only in few censuses by age groups (4-6, 7-10, 11-20, etc.).

Regarding the data collection process, and as established in the enumerators' instructions, information was completed on family cards (*cédulas de inscripción familiares*) by the head of the household, or the local agent if household members could not write, on December 31[11]. This implies that the raw data could be recorded by one person for the whole household. Therefore, information was not self-reported. Additionally, it is stipulated that family cards must contain information on the sex, age and level of education, among other characteristics, for each member of the household. More specifically, while the level of education was recorded marking one of the following alternatives («can read», «can read and write» or «cannot read or write»), the exact age was written down on the family card.

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

$$\text{ABCC} = \left\{1 - \frac{(W - 100)}{400}\right\} \times 100 \quad \text{for } W \geq 100 \tag{6}$$

[10] Although population counts had been carried out in the 18th century, modern population censuses are usually associated with the creation of the General Statistics Commission in 1856. Muro *et al.* (1996), for instance, offer an overview of this process. Similarly, Reher and Valero Lobo (1995) provide an overview of historical demographic sources.

[11] For a sample of the registration cards employed, see Figure A.1 in the Appendix. The instructions to carry out the census of 1877 were established by Royal Order (RO, 27 October 1877). In Art. 24, it is specified who must fill in the family cards and what information must be reported.

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

$$\text{ABCC} = 100 \text{ elsewhere} \quad (7)$$

Theoretically, the ABCC index thus indicates the share of respondents that report age correctly and varies between 0 and 100[12]. Equally, the modified version of the Whipple index ($\check{W}$) can be easily transformed[13]. Table 1 summarises both the original (ABCC) and adjusted ($\overline{\text{ABCC}}$) indices at the national level for all population censuses since 1877. We compute two alternative indicators, for 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 1 shows. Additionally, although the indices are greater when using the 23-62 age range than 23-72, these differences are negligible. For the sake of comparability, we concentrate on the traditional ABCC index and the 23-62 age range in the remainder of this article[14].

4. AGE-HEAPING IN SPAIN SINCE THE MID-19TH CENTURY

4.1 Exploiting the Censuses: a Larger Dataset, Alternative Results

The census of 1877 was the first modern count reporting age-specific information for the whole population. Since the ABCC index has been constructed for the 23-62 age group, it then provides information about individuals born between 1815 and 1854[15]. Table 2 displays age-heaping, measured with an ABCC index (per cent), for men and women born in the 1850s across Europe as computed by the Clio-Infra Project[16].

[12] The above interpretation should be taken with caution since age misreporting might still occur even if age-heaping is not detected. For the presence of misreporting of age among the elderly in early 20th-century Spain, see García-Quismondo (1999).

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

[14] In order to avoid a survival bias, we follow Crayen and Baten (2010a) and omit the population older than 62 or 72. Besides, Crayen and Baten (2010a, pp. 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 age cohorts will be used instead. See also Budd and Guinanne (1991).

[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, at 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.

[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 Spain, the main references are Manzel (2007), Crayen and Baten (2010a), Juif and Baten (2013) and Stolz *et al*. (2013).

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

TABLE 1
AGE-HEAPING IN SPAIN BY POPULATION CENSUS

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

According to Table 2, there appears to be a divide between Russia, Spain and the rest. Nevertheless, the estimates must be taken with caution. For Spain, the ABCC index 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 age-heaping increases with age[19]. Our data indeed make it possible to compare different age groups over time with previous estimates.

In Figure 2, we compare the age-heaping of those aged 23-32 and 33-42 in each population census with the existing estimates. All values are presented by birth decade. First, our results show that the ABCC index in

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

[18] See Crayen and Baten (2010a, pp. 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.

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
The Netherlands	99.0
Norway	98.4
Austria	98.0
UK	97.1
Spain	87.4
Russia	83.6

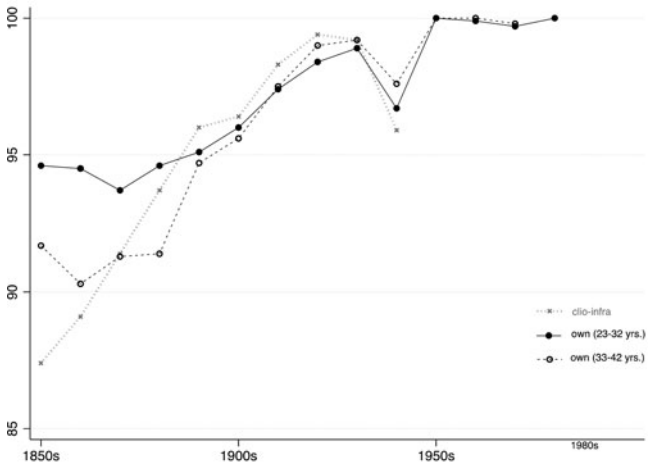
Note: ABCC index (per cent) illustrated above is a birth decadal average.
Source: Clio-Infra

mid-19th century Spain was not as low as described in the existing literature. Furthermore, if we focus on specific age groups, the ABCC indices remained relatively stable in the late 19th century. Following age groups (23-32; 33-42) over time then provides a somewhat different view of the dynamics of age-heaping. If ABCC indices are displayed, on the other hand, for the whole range (23-62), sex and census (Figure 3), it can be argued that age-heaping did not improve much until the population census of 1920. The period of birth of those included in each population census is shown in the *x*-axis. This evidence reinforces the above finding: age-heaping remained relatively stable until the 20th century.

If age-heaping is correlated with socioeconomic development, then this finding is in line with the existing evidence on living standards. Although Spain witnessed population and economic growth during the second half of the 19th century (Prados de la Escosura 2017), some studies have indicated that improvements in living standards generally occurred from the 1880s onwards. 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

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

FIGURE 2
ABCC INDEX (PER CENT) IN SPAIN BY BIRTH DECADE, 1850-1980.



Note: Clio-Infra uses distinct age groups to construct the ABCC indices. More specifically, 1850s (43-52 years in 1900 census); 1860s (63-72 years in 1930 census); 1870s (53-62 years in 1930 census); 1880s (53-62 years in 1940 census); 1890s (43-52 years in 1940 census); 1900s (33-42 years in 1940 census).

Sources: Computed with data from Clio-Infra and INE.

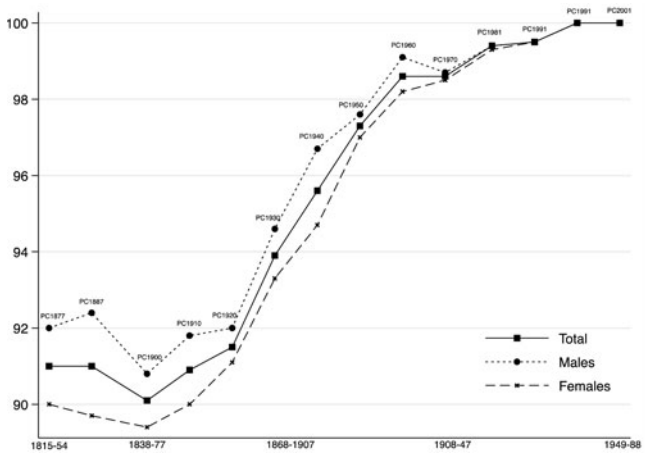
average adult heights only improved from the 1880s onwards[20]. Life expectancy, which remained stagnant at around 30 years between the 1860s and the 1880s, began to show an increasing trend in the 1890s (Felice and Pujol-Andreu 2016). Finally, nutrition also improved in the late 19th and early 20th century (Cussó *et al*. 2018; Medina-Albaladejo and Calatayud 2020).

Additionally, the detailed information offered in the censuses can be used to compare Spain with neighbouring countries. A'Hearn *et al*. (2016), using the Italian census of 1871, also compute Whipple indices by age groups and sex. When comparing these indices with those derived from the Spanish population census of 1877, it seems that age-heaping was slightly less prevalent than previously claimed[21]. In fact, it appears that age-heaping was less prevalent than in Italy, which challenges the conventional view (in Table 2) that Spain was lagging behind Italy by the mid-19th century. Although age-heaping was still more prevalent than in other

[20] As Otero Carvajal and De Miguel Salanova (2018) have recently stressed, urban development in the late 19th century is at the core of the issue.
[21] See Tables A.1 and A.2 in the Appendix.

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

FIGURE 3
ABCC INDEX (PER CENT) IN SPAIN BY SEX AND POPULATION CENSUS, 1877-2001.



Note: ABCC index (per cent) constructed for those aged 23-62 years. Population censuses are illustrated for each point while the *x*-axis displays the birth period.

Source: Calculated with data from INE.

countries (Belgium, Finland, France, etc.), it was less widespread than previously documented.

4.2 Female Age-heaping in Spain's Censuses of Population

When examining age-heaping by sex (Figure 3; Table A.1), there appears to be another important fact. In line with previous studies, we find that differences in age-heaping between males and females were small. This has also been found in other contexts. In the 17th and 18th centuries, women appeared to be as numerate, if not more so, than men in the Low Countries (De Moor and Van Zanden 2010)[22]. Using municipal registers, Gómez-i-Aznar (2019) shows that male and female levels of age-heaping were rather similar in 18th-century Catalonia. There were, however, sizable differences in literacy thereby opening the way for an interesting discussion.

The apparent lack of a relationship between these two proxies of human capital has brought about a compelling debate. On the one hand, and given

[22] 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 *et al.*, 2012). See also Tollnek and Baten (2016).

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

its distinct nature, these skills might not be interchangeable. If, for instance, counting required less effort or training than literacy and both men and women actively participated in day-to-day activities then it would be reasonable to find similar levels of age-heaping. On the other hand, it has been pointed out that age might not have been self-reported. Instead, the head of the household, or a local agent when family members were illiterate, was responsible for this. As Blum *et al*. (2017) have recently shown, rural women in 19th-century Ireland heaped more than the censuses suggested, thereby recommending caution when using administrative data[23].

In our case, although there were sizable differences in literacy between men and women, this was not the case as regards age-heaping as Figure 3, Table A.1 and Figure A.2 show[24]. To shed further light on this issue, we use information for forty-nine provinces and six censuses to test the relationship between literacy and age-heaping, both for males and females. The panel structure of the data allows for time and province fixed effects[25]. After controlling for all these dimensions, the relationship appears to be statistically significant only for males, as Table 3 shows. In fact, while female literacy is not associated with female age-heaping, male literacy is, thus strongly suggesting that it was usually men who reported the ages of their wives.

In this regard, Földvari *et al*. (2012) found that married women heaped significantly less than those who were not married. This could be the result of selection, learning within marriage, or women adapting their ages to that of their husbands. Blum *et al*. (2017), however, find that married women exhibited higher ABCC indices than those who were unmarried thereby suggesting that, instead of the previous interpretations, age was not self-reported in censuses or administrative data, but filled in by the head of the household, mostly men[26]. Using information from the population census of 1930, Table 4 presents age-heaping for women by marital status. Sadly, this information is not available for the previous censuses.

[23] Baten *et al.* (2017, p. 38), for instance, «took care to not include any sources in which some cross-checking by priests or officials might have taken place. In those cases, there was almost no heaping present».

[24] Figure A.2 in the Appendix illustrates the differences observed in age-heaping and literacy by sex across provinces in 1900. Both indicators have been standardised [0,1]. In Soria, for instance, whereas the male and female ABCC indices were 96.5 per cent and 96.7 per cent, the male and female literacy rates were 90.5 per cent and 43.8 per cent. Flecha García (2014) offers a brief overview of the causes and consequences of the gender imbalances in education in Spain.

[25] Although 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, it is worth remembering that the data are from the same source.

[26] Blum *et al.* (2017) compare data from census with evidence from prison and workhouse registers and find that ABCC indices overestimate age-heaping in the census by 26 points.

TABLE 3
LITERACY AND NUMERACY IN SPAIN, 1877-1930

	Dependent variable: age-heaping			
	All	**Male**	**Female**	**Female**
	(1)	**(2)**	**(3)**	**(4)**
Literacy	0.02	0.09***	–0.03	0.09**
	(0.04)	(0.03)	(0.03)	(0.03)
Constant	91.17***	88.19***	91.42***	86.66***
	(1.27)	(1.50)	(0.58)	(1.88)
Time FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
Observations	294	294	294	294
R^2	0.254	0.314	0.251	0.270
Provinces	49	49	49	49

Robust standard errors in parentheses; ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$. Literacy in columns (1) to (3) refers to the corresponding group (all, male and female). In column (4), however, literacy refers to male literacy. The data include six time periods: 1877, 1887, 1900, 1910, 1920 and 1930.

As shown here, the ABCC index for married women is essentially similar to that of men. The ABCC index was however significantly lower for singles and widows, whose figures are mostly identical[27].

Summing up, we show that differences in age-heaping between men and women were small for all the censuses. Moreover, the relationship between age-heaping and literacy only arises for males. In this regard, according to the instructions given by the census enumerators, age does not seem to be self-reported. Instead, the head of the household (or member) and the local agent filled in the cards. Taken together, the evidence suggests that female age-heaping computed from Spanish population censuses should be taken with caution since the ABCC index might not be reflecting their numeracy but that of their husbands. In the next section, we will describe the main regional patterns. Given the previous discussion, we will only focus on male age-heaping and literacy.

[27] The fact that single women in the youngest group (23-32) have the highest ABCC index is likely to be related to their age composition relative to those married and widowed in that age group (while singles will be closer to 23, widows will tend to be in their late 20s and early 30s). The ABCC index in this age group is therefore over- and under-estimated for singles and widows, respectively (both figures should thus be more similar).

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

TABLE 4
FEMALE AGE-HEAPING BY MARITAL STATUS AND AGE COHORT IN SPAIN, 1930

	23-32	33-42	43-52	53-62	23-62
Men					
ABCC	96.6	95.3	92.4	92.3	94.6
Obs.	1,839,694	1,413,511	1,135,672	858,206	5,247,083
Women					
ABCC	95.4	94.1	91.4	90.1	93.3
Obs.	1,959,281	1,509,203	1,227,157	979,490	5,675,131
Married					
ABCC	94.8	95.2	93.4	93.2	94.4
Obs.	1,192,306	1,173,223	902,107	599,877	3,867,513
Single					
ABCC	96.5	90.2	85.2	85.6	93.1
Obs.	737,176	242,930	141,389	96,090	1,217,585
Widowed					
ABCC	90.9	89.6	86.2	85.0	86.4
Obs.	26,841	90,949	181,797	281,898	581,485
Not known					
ABCC	91.5	88.4	87.0	85.3	88.6
Obs.	2,958	2,101	1,864	1,625	8,548

Source: Calculated with data from INE.

5. A REGIONAL PERSPECTIVE ON AGE-HEAPING AND LITERACY IN SPAIN, 1877-1930

While the previous section provides information at the national level, in this one we rely on disaggregated information on male age-heaping and literacy at the provincial level[28]. In general, there appears to be a strong relationship between age-heaping and literacy (Hippe 2012). In the United States, for instance, and using individual data from censuses, there is a statistically significant correlation between literacy and the probability of reporting a heaped age[29]. Likewise, A'Hearn *et al*. (2016, 2019) have looked

[28] Spanish population censuses report information at province level and, in some instances, for large municipalities. Alternatively, individual data can be found in the *Padrones municipales*. See, for instance, Beltrán Tapia and Miguel Salanova (2017).

[29] 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).

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

at this relationship with Italian censuses. Though the correlation is present in broad terms, they find some inconsistencies that might challenge the conventional interpretation of age-heaping as a proxy for numeracy skills.

In Spain, we study age-heaping and literacy at the provincial level for each census between 1877 and 1930[30]. Using forty-nine provinces as the main unit of analysis, the correlation between the male ABCC indices and literacy rates ranges from 0.75 in 1877 to 0.59 in 1930, while the Spearman rank correlation goes from 0.80 in 1877 to 0.67 in 1930. Thus, and in line with the existing evidence, there is a strong and statistically significant relationship. In this regard, the 1887, 1900 and 1910 censuses reported literacy rates by sex 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 this relationship[31].

Figure 4 shows male 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 dots 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. Figure 4, however, also shows that while male literacy improved rapidly, age-heaping remained relatively unchanged until the censuses of 1920 and 1930. Furthermore, there is substantial noise or dispersion within Spain, thereby calling for a closer examination.

Map 1 shows the spatial distribution of male age-heaping and literacy in 1877, 1900 and 1930. In brief, there appears to be a core, namely the centre and centre-north, where literacy was relatively high and age-heaping was practically non-existent. By 1877, in the provinces of Burgos, Álava, Segovia, Palencia, Soria, Valladolid, Madrid and Guadalajara, the ABCC index was above 98 per cent, thereby implying that less than 2 per cent of males aged 23-62 years old reported their age incorrectly[32]. This level of accuracy would place these territories with the most advanced European societies[33]. Male literacy rates in these provinces, however, ranged from 62 per cent in Guadalajara to 83 per cent in Burgos[34].

[30] 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.

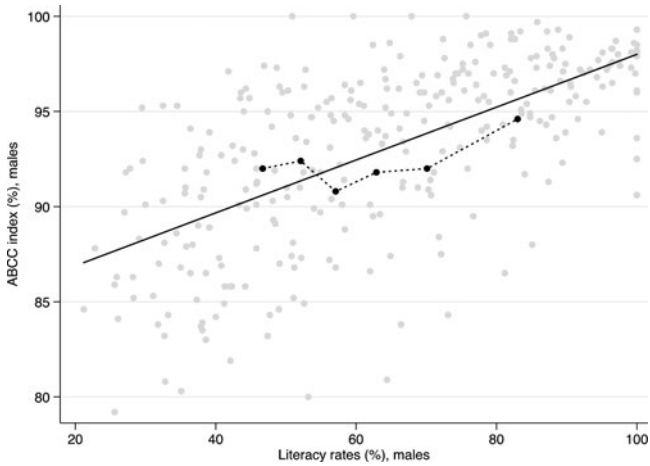
[31] The correlation coefficient ranges from 0.6615 in 1887 to 0.7063 in 1900 for males aged 21-30. Similarly, for those aged 31-40, the correlation coefficient ranges from 0.6350 to 0.7189.

[32] For the United Nations Statistics Division (UNSD), a Whipple index below 105 (or an ABCC greater than 98.8 per cent) is considered «highly accurate data».

[33] Using Clio-Infra historical estimates, age-heaping in the north and centre-north provinces of Spain will be at similar levels to those computed for Norway or the Netherlands, see Table 2.

[34] In Spain, around 32.5 per cent 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 per cent), UK (76 per cent) and Germany (80 per cent) Crafts (1997).

FIGURE 4
ABCC INDEX (PER CENT) AND LITERACY RATES (PER CENT) IN SPAIN BY PROVINCE, MALES 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.

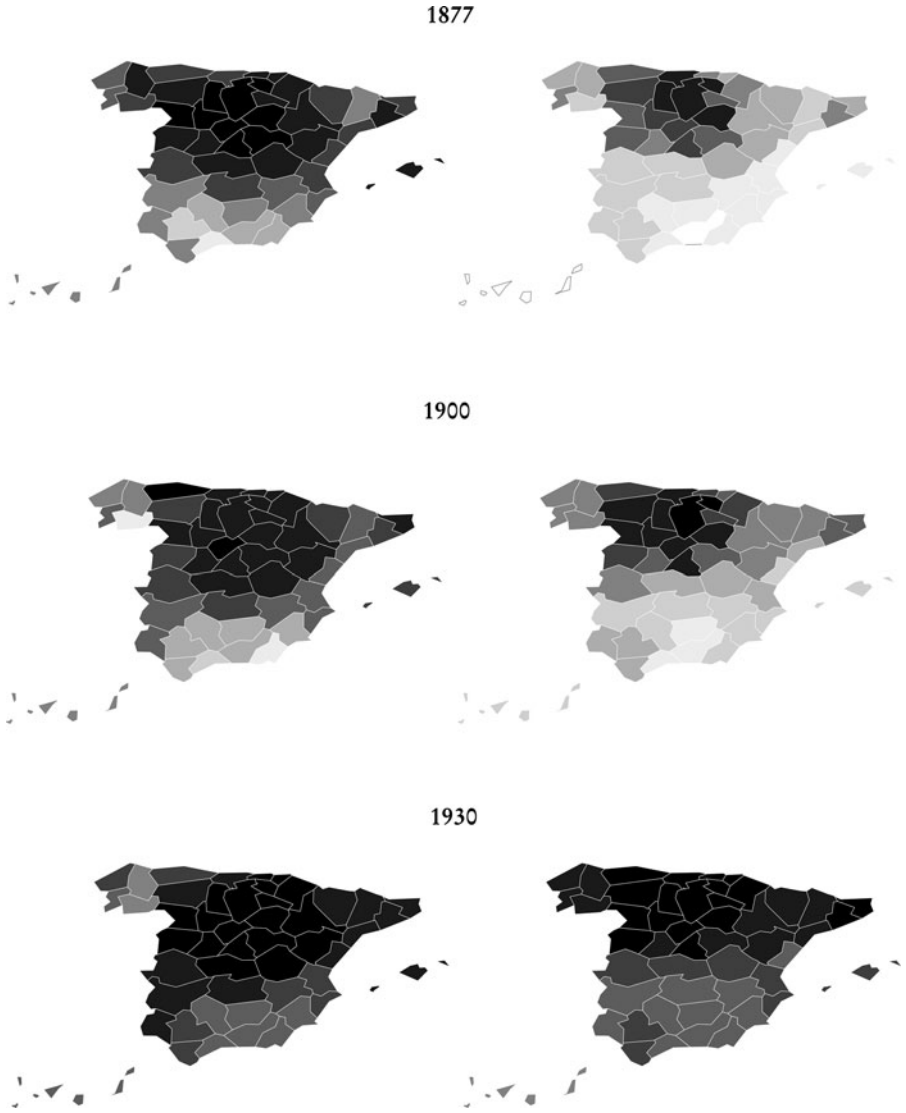
Source: Calculated with data from INE.

In other provinces, especially in Andalusia, the Canary Islands, Galicia and the eastern coast, age-heaping and literacy were well below the national average. This great divide, as Map 1 shows and previous studies suggested (Núñez 1992), eventually diminished. On the eve of the Spanish Civil War (1936-1939), most of the northern provinces had almost completed the transition to universal male literacy. Similarly, age-heaping was negligible except for the north-west, Galicia and Asturias. In southern and south-eastern provinces, literacy rates and the ABCC index gradually improved, although they remained low.

In Figure 5, we standardise both measures [0,1] and depict the absolute change in the vertical axis (1877-1900; 1900-1930) and the initial level in the horizontal axis. As shown previously, catching up in human capital essentially occurred in the period 1900-30. That is to say, age-heaping and literacy improved more in provinces with a lower initial level. Yet, in the late 19th century, things were different. When considering all provinces, male literacy improved evenly between 1877 and 1900, irrespective of the initial levels. The situation regarding age-heaping, however, is less clear. Though some catching up can be observed, the absolute change in most provinces was negligible, in some cases even negative, implying little, if any, improvement during this period.

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

MAP 1
ABCC INDEX (LEFT) AND LITERACY RATE (RIGHT) IN SPAIN, MALES 1877-1930.

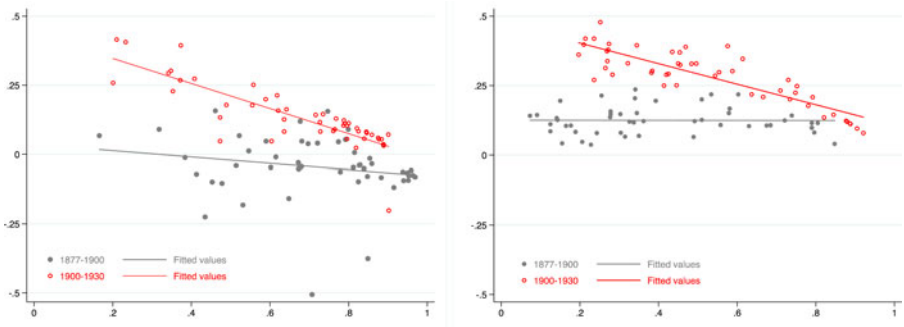


Notes: Provincial ABCC indices (per cent) and literacy rates (per cent) classified into nine equivalent categories for all censuses. In both cases, the range unchanged over the period. The ABCC indices range from 75 to 100 per cent while literacy rates range from 15 to 100 per cent.

Source: Calculated with data from INE.

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

FIGURE 5
ABCC INDEX (LEFT) AND LITERACY RATE (RIGHT) IN SPAIN, MALES 1877-1900 AND 1900-1930.



Notes: Provincial ABCC index (per cent) and literacy rates (per cent) have been standardised using a fixed range for both periods, 1877-1900 and 1900-1930. The ABCC indices range was 75-100 while for literacy it was from 15 to 100 per cent. Thus, the *y*-axis reflects absolute change whereas the *x*-axis shows the initial level.

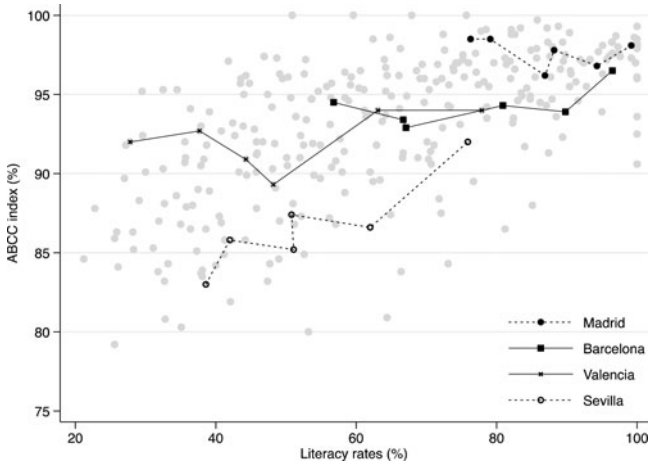
Source: Calculated with data from INE.

To delve further into the spatial patterns, Figure 6 displays the trajectories followed by four selected provinces: Madrid, Barcelona, Valencia and Seville. These provinces contain the largest cities in Spain. All things considered, the figure presents some interesting facts. On the one hand, in Madrid, the capital province, male age-heaping was practically inexistent even at the beginning of our period of analysis. Literacy was also relatively high, nearly 80 per cent in the earliest count, and improved thereafter. In Barcelona, however, literacy and age-heaping were initially far less impressive but, while literacy increased throughout the whole period, the ABCC index remained unchanged until the 1930 census. In Valencia, literacy was low (28 per cent) while the ABCC index was not far away from that of Barcelona where more than 40 per cent of men and women could read and write. Finally, in Seville, both indicators were initially low but then increased rapidly. All in all, these stories present a snapshot of the peculiar relationship between age-heaping and literacy in Spain.

An important feature absent from our analysis is internal migration, which could affect the stock of human capital. Beltrán Tapia and Miguel Salanova (2017) found that the migrants moving to Madrid from 1880 to 1930 were on average more literate than their counterparts who 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

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

FIGURE 6
ABCC INDEX AND LITERACY IN SPAIN BY PROVINCE, MALES 1877-1930.



Source: Calculated with data from INE.

Note: Marked dots correspond to the census years 1877, 1887, 1900, 1910, 1920 and 1930.

relevant, since the type (and importance) of migratory flows would influence the stock of human capital both in sending and receiving regions. We acknowledge the potential relevance of internal migrations, but it must be born in mind that although internal migrations have been recurrent in Spanish history, permanent internal migrations remained low until the 1920s (Silvestre 2001, 2005).

6. DISCUSSION

Taken together, the findings reported here suggest that assessing the relationship between age-heaping and literacy during the early stages of modern economic growth is subject to some major issues, at least regarding the Spanish experience. First, and connected with the previous sections, female age in Spanish censuses, at least from 1877 to 1930, seems not to be self-reported. This, as pointed out previously, is a common concern in studies based on administrative data[35]. While censuses have the virtue of including the whole population, thereby removing concerns about

[35] The issues related to the fact that in household interviews most of the information is supplied by a single respondent, often the head of the household, have been widely discussed (National Research Council, 1981).

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

potential selection bias or representativity, some authors have questioned their reliability. In our case, sizable differences between males and females are found in literacy but not in age-heaping which calls for caution when making explicit interpretations.

Besides, and although these relative indicators cannot be directly compared, age-heaping in late 19th-century Spain seems abnormally low given the widespread illiteracy. As reported in the census of 1877, male literacy in Almería, Granada, Málaga or the Canary Islands was around 20-25 per cent while ABCC indices ranged from 79 to 87 per cent. Given this level of accuracy and following the standards used by the United Nations Statistics Division (UNSD), these data would be labelled as «rough». Yet, these figures would be «fairly accurate», almost «highly accurate», for several northern provinces even though male literacy rates were in some cases below 75 per cent, thereby implying that one out of four adult males were not able to read and write[36].

This leads to a recurrent theme: whether the accuracy of the data captures numeracy or whether it is affected by the data-collection process. In the first case, our findings would provide further evidence in support of the view that numeracy, reading and writing have a distinct nature. It is possible to hypothesise that, while learning to read and write required a larger investment of resources, effort and time, numeracy could be acquired more easily. Indeed, the latter may be obtained in a more intuitive way in daily life and work, especially when individuals are surrounded by other people who possess and use this skill. In this context, on-the-job training might be a crucial factor. In Spain, Rosés (1998) stressed the relevance of on-the-job training during the industrialisation of Catalonia around the mid-19th century, a period marked by relatively modest literacy in comparison with other territories. Similarly, Nadal (1996) argued that technical schools (*liceus* and *escoles d'arts i oficis*) provided the required skills to the factory workers in Catalonia[37].

On the other hand, if information is corrected and cross-checked with other sources, such as parish registers or *padrones*, by the local agent (or census taker) or clerical workers, age-heaping would then not be fully capturing numeracy. A'Hearn *et al*. (2016) have suggested that age-heaping is not directly measuring numeracy, at least in 19th-century Italy, but a «broader mix of contextual factors» related to the process of economic

[36] In the Demographic Yearbooks, the quality of the data is grouped according to the level of age-heaping (Whipple index) as follows: (I) highly accurate data (less than 105); (II) fairly accurate data 105-109.9; (III) approximate data 110-124.9; rough data 125-174.9; very rough data 175 and more.

[37] A similar argument can be found for Italy (Zamagni, 1993). As regards Britain, Zeev *et al.* (2017) find that in the 18th century, apprenticeships made a key contribution to an advantage in skilled mechanical labour.

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

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 carrying out the population censuses.

In Spain, population censuses were undertaken on December 31. This colossal enterprise, coordinated by the *Dirección General del Instituto Geográfico y Estadístico*, was conducted at the local level by the *Juntas Censales* (*municipales* and *provinciales*)[38]. As specified in the instructions, households were given registration cards some days before. The head of the household, or local agent if household members could not write, then filled in the information. On 1 January, the registration cards were collected and the *Juntas municipales* began the examination of the reported information, correcting and filling in the gaps or omissions found. Finally, the *Juntas municipales* tabulated the data in a specific format for the municipality and prepared the *padrones*[39].

Therefore, in parish and civil registers and *padrones*, information might have been filled in, cross-checked and corrected. Interestingly, the civil registry was introduced in the 1870s, after the passing of the *Ley Provisional de Registro Civil* of 1870 and subsequent acts. In both parish and civil registers, the date of birth was often recorded thereby allowing for a simple calculation of the age in each specific year. Besides, municipalities maintained a register, or *padrón*, of the residents or *empadronados*, which provided «a line for each *empadronado*» with basic information (name, address, date of birth, marital status…). The *Ley Municipal* of 1870 stipulated that all citizens must be registered in a municipality. Additionally, every 5 years municipalities had to carry out a *padrón*.

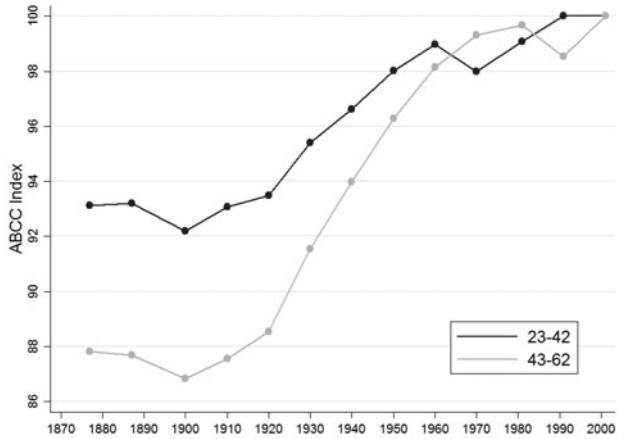
As a result, the information on age-heaping based on the Spanish population censuses may be pointing to the existence of limited numerical abilities, but also to a limited capacity of the state to collect the data. Differences in age-heaping could also be related to the presence of regional differences in administrative capacity. As previously mentioned, the state designed the data-collection process and sent instructions and materials (*cédulas*, etc.) to the municipalities. Local authorities and agents proceeded to collect, correct and tabulate the information before sending it to the provincial board, where it was summarised. We do not have evidence to claim that the allocation of funding varied, and it seems that

[38] 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 during the 1900s it depended on different ministries (*Ministerio de Instrucción Pública y Bellas Artes*, *Ministerio de Trabajo, Comercio e Industria*).

[39] *Art. 25. Real Orden de 27 de octubre de 1877*.

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

FIGURE 7
ABCC INDICES BY TWO LARGE AGE GROUPS, SPAIN 1877-2001.



Source: Calculated with data from INE.

much was devoted to printing the instructions and materials. However, given the spatial heterogeneity of the Spanish municipalities, it might be the case that the costs associated with collecting the data were greater in some places and the municipal budgets might have allocated different amounts of resources. In addition, there might be other relevant differences regarding the existence or not of parish registers or *padrones* which permitted cross-checking. The dissimilar settlement patterns in northern and southern Spain could also imply different costs when collecting the information. All these issues open up a new avenue for research that requires more detailed micro-level studies.

Although the accuracy of the census takers in performing their task is difficult to ascertain, it is recognised that older people find it more difficult to remember their exact age, so it is expected that rounding numbers would be more pronounced as people get 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 7 shows the results. The increased age-heaping observed in the older generations, particularly in the early counts when age-heaping was more prevalent, is in line with the existing literature and makes us more

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

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

confident about the capacity of the state to collect information and therefore about the accuracy of our male age-heaping indicators.

Finally, since age-heaping is calculated for those aged 23-62 or 23-72, the developments described above might have an impact on the age-heaping reported in the population censuses of the early 20th century. As shown in Figure 1 and Map 1, age-heaping remained relatively unchanged in the early counts, improving thereafter. Besides, even if accuracy improved, this occurred unevenly as Map 1 shows. By 1930, age-heaping remained relatively high in the north-west (Galicia, Asturias). Given that migrations were particularly acute in the early 20th century, the latter might be explained by the selection biases these flows introduced regarding the characteristics of the population who migrated or stayed behind (Beltrán Tapia and Miguel Salanova 2017). The general patterns might also be related to the different settlement patterns characterising Spain (Oto-Peralías 2020). In the north-west, the density of population entities is much higher than in other territories and most of these are relatively small, thus facilitating enumerating their populations[41]. It is therefore plausible to argue that the costs associated with the undertaking of a census might be affected by territorial specificities and their dynamism[42].

Literacy, on the other hand, improved throughout the whole period but especially in the early 20th century. Our study, then, shows that both stories are plausible. Literacy and age-heaping are correlated which reinforces the view that the latter captures a dimension of human capital, numeracy. Likewise, this story is consistent with a steady improvement of the data collection methods, especially after the introduction of the civil registry. Although we lack the evidence to disentangle this issue, this study calls for a careful consideration of the sources in historical contexts. Lastly, these results also point to the importance of territorial specificities that require further research.

7. CONCLUSIONS

This paper explores human capital in late 19th and early 20th-century Spain. In doing so, we use information published in six population censuses for age and the level of education. Subsequently, and following the existing literature, we calculate age-heaping and literacy rates. Our approach thus combines two distinct proxies of human capital.

[41] In the south, there is a low density of large entities, while the centre and north are essentially characterised by a high density of small and medium-sized municipalities.

[42] See Reher *et al.* (1993). We do not, however, have precise information on the numbers and areas covered by census takers.

In brief, there are three major findings. First, age-heaping was relatively stable in the late 19th century. In fact, it only significantly improved in the early 20th century. Literacy, however, gradually improved throughout the period of study. This dissociation might be the result of the distinct nature of both skills and/or issues related to the data-collection process. Second, by the mid-19th century, age-heaping, either for young adults (23-32 years) or for the whole distribution (23-62 years), was less significant than what previous studies have suggested. What is more, when provincial information is used, we find a correlation between age-heaping and literacy but also abnormally low levels of age-heaping in certain provinces where a sizable part of the population could not read and write. Likewise, although Italy and Spain exhibited similar levels of literacy during the 1870s, age-heaping was less relevant too. Third, differences in age-heaping between men and women were minor even though there were marked gaps as regards literacy.

In sum, this study provides further insights, but also raises questions about data sources and interpretation. The absence of a gender gap in age-heaping reinforces the view that female information in administrative data, such as population censuses, is not self-reported. In this case, the head of household, or local agent if household members were illiterate, filled in the information. Still, even if the analysis is restricted to males, the dissociation between age-heaping and literacy in the late 19th century and the abnormally low levels of age-heaping in certain provinces raise concerns about the sources, and hence the interpretation of these findings. If omitted information is filled in, cross-checked or corrected with other sources (parish and civil registries, *padrones*) by the *Juntas municipales*, then accuracy might not just be reflecting numeracy but the capacity of the local administration. Although our study does not permit digging further into these issues, human capital and data quality improved, especially in the early 20th century, whereas further research is needed to disentangle the late 19th-century story.

ACKNOWLEDGEMENTS

Received 10 October 2019. Accepted 23 November 2020.

The authors would like to thank the editor and three anonymous referees for their comments. We are also grateful for the financial support of the Bank of Spain through the «*Programa de Ayudas a la Investigación en Macroeconomía, Economía Monetaria, Financiera y Bancaria, e Historia Económica*». This paper was presented at the III Seminar in Economic History at the Bank of Spain in Madrid. We would like to thank the participants for the comments received and Fran Marco-Gracia who kindly shared information regarding the population censuses with us. Financial

support from the Spanish Ministry of Science and Innovation, Projects ECO2015-65049-C12-1-P, ECO2017-90848-REDT, PGC2018-095821-B-I00, HAR2015-64076-P, PGC2018-095529-B-I00 (MCIU/AEI/FEDER, UE), the *Generalitat Valenciana* (AICO/2018/1930 and PROMETEO/2020/083) and the Norwegian Research Council, Project 249723, is also gratefully acknowledged.

REFERENCES

A'HEARN, B., BATEN, J., and CRAYEN, D. (2009): «Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital», *The Journal of Economic History* 69 (3), pp. 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.

A'HEARN, B., DELFINO, A, and NUVOLARI, A. (2019): «Cognition, Culture, and State Capacity: Age-Heaping in XIX Century Italy», CEPR Discussion Paper in Economic History, DP14261.

ALLEN, R. C. (2003): «Progress and Poverty in Early Modern Europe», *The Economic History Review* 56 (3), pp. 403-443.

BARRO, R. J. (2013): «Education and Economic Growth», *Annals of Economics and Finance* 14 (2), pp. 301-328.

BATEN, J., and VAN ZANDEN, J. L. (2008): «Book Production and the Onset of Modern Economic Growth», *The Journal of Economic Growth* 13 (3), pp. 217-235.

BATEN, J., SZOLTYSEK, M., and CAMPESTRINI, M. (2017): «'Girl Power' in Eastern Europe? The Human Capital Development of Central-Eastern and Eastern Europe in the Seventeenth to Nineteenth Centuries and its Determinants», *European Review of Economic History* 21 (1), pp. 29-63.

BELTRÁN TAPIA, F. J. (2013): «Enclosing Literacy? Common Lands and Human Capital in Spain, 1860-1930», *Journal of Institutional Economics* 9 (4), pp. 491-515.

BELTRÁN TAPIA, F. J., and DE MIGUEL SALANOVA, S. (2017): «Migrants' Self-Selection in the Early Stage of Modern Economic Growth, Spain (1880-1930)», *The Economic History Review* 70 (1), pp. 101-121.

BELTRÁN TAPIA, F. J., and MARTINEZ-GALARRAGA, J. (2018): «Inequality and Education in Pre-Industrial Spain: Evidence from Spain», *Explorations in Economic History* 69 (3), pp. 81-101.

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), pp. 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), pp. 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*. Alicante: Publicaciones de la Universidad de Alicante, pp. 9-40.

CARRERAS, A. (1990): *Industrialización española. Estudios de historia cuantitativa*. Barcelona: Espasa-Calpe.

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, pp. 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), pp. 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», *The Economic History Review* 63 (2), pp. 452-478.
Cussó, X., Gamboa, G., and Pujol-Andreu, J. (2018): «El estado nutritivo de la población española, 1860-2010: una aproximación a las diferencias de género y generacionales», *Nutrición Hospitalaria* 35 (5), pp. 11-18.
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), pp. 179-208.
Felice, E., and Pujol-Andreu, J. (2016): «GDP and Life Expectancy in Italy and Spain Over the Long-Run: A Time-Series Approach», *Demographic Research* 35 (1), pp. 813-866.
Flecha García, C. (2014): «Desequilibrios de género en educación en la España contemporánea: causas, indicadores y consecuencias», *Áreas-Revista Internacional de Ciencias Sociales* 33, pp. 49-60.
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), pp. 304-313.
Friesen, J., Baten, J, and Prayon V. (2012): «Women Count. Gender (In-)Equalities in the Human Capital 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, pp. 441-493.
García Abad, R., Pareja Alonso, A., and Zarraga Sangroniz, K. (2007): ««¿Sabe leer? ¿Sabe escribir?» El proceso de alfabetización en el País Vasco (1860-1930)», *Revista de Demografía Histórica* 25 (1), pp. 23-58.
García-Quismondo, F. (1999): «Declaración errónea de la edad de adultos y ancianos en España: evaluación de la consistencia de los datos para el análisis de la mortalidad, 1900-1930», *Boletín de la Asociación de Demografía Histórica* 7 (2), pp. 29-60.
Gennaioli, N., La Porta, R., Lopez-De-Silanes, F., and Shleifer, A. (2013): «Human Capital and Regional Development», *Quarterly Journal of Economics* 128 (1), pp. 105-164.
Goldin, C. (2015): «Human Capital», in C. Diebolt and M. Haupert (eds), *Handbook of Cliometrics*, ch. 3. Berlin and Heidelberg: Springer, pp. 55-86.
Gómez-I-Aznar, E. (2019): «Human Capital at the Beginnings of the 18th Century Catalonia: Heaping and Numeracy in a Changing Economy», Documento de Trabajo – AEHE, No. 1904.
González Portilla, M., and Urrutikoetxea Lizarraga, J. (2016): «El capital humano en la primera modernización industrial vasca (1876-1930). Viejas herencias e innovaciones recientes», *Revista de Demografía Histórica* 34 (2), pp. 53-83.
Graff, H. (1987): *The Legacies of Literacy: Continuities and Contradictions in Western Culture and Society*. Bloomington: Indian University Press.

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).

Hippe, R. (2012): «How to Measure Human Capital? The Relationship Between Numeracy and Literacy», *Economies et Sociétés* 45, pp. 227-241.

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), pp. 227-241.

Lucas, R. (1988): «On the Mechanics of Economic Development», *Journal of Monetary Economics* 22 (1), pp. 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), pp. 407-437.

Manzel, K. (2007): «Human Capital Development and the Impact of the Hungry Forties in Spain», Unpublished manuscript.

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, pp. 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), pp. 502-517.

McCloskey, D. N. (2010): *Burgeois Dignity: Why Economics Can't Explain the Modern World*. Chicago: University of Chicago Press.

Medina-Albadalejo, F. J., and Calatayud, S. (2020): «Unequal Access to Food During the Nutritional Transition: Evidence from Mediterranean Spain», *The Economic History Review* 73 (4), pp. 1023-1049.

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, pp. 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 edn. Boulder: Westview Press, pp. 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, pp. 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, pp. 7-42.

Muro, J. I., Nadal, F., and Urteaga, L. (1996): *Geografía, estadística y catastro en España 1856-1870*. Barcelona: Ediciones del Serbal.

Nadal, J. (1996): «El factor humà en el retard econòmic espanyol. El debat entre els historiadors», *Revista Económica de Catalunya* 29, pp. 9-14.

National Research Council (1981): *Age Misreporting and Age-Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analyses*. Washington, DC: The National Academies Press.

Nicolau, R. (2005): «Población, salud y actividad», in A. Carreras and X. Tafunell (eds), *Estadísticas históricas de España: siglos XIX-XX*. Bilbao: Fundación BBVA, pp. 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), pp. 1038-1041.

Núñez, C. E. (1991): «El gasto público en educación entre 1860-1935», *Hacienda Pública Española* 1, pp. 121-146.

Núñez, C. E. (1992): *La fuente de la riqueza: educación y desarrollo económico en la España contemporánea*. 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, pp. 122-142.

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

Otero Carvajal, L. E., and De Miguel Salanova, S. (2018): *La escuela y la despensa. Indicadores de modernidad, España, 1900-1936*. Madrid: Catarata.

Oto-Peralías, D. (2020): «Frontiers, Warfare and Economic Geography. The Case of Spain», *Journal of Development Economics* 146, in print https://doi.org/10.1016/j.jdeveco.2020.102511

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*, vol. 1, 1700-1870, ch. 9. Cambridge: Cambridge University Press, pp. 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», Unpublished manuscript.

Reher, D. S., and Valero Lobo, A. (1995): *Fuentes de información demográfica en España*. Madrid: Centro de Investigaciones Sociológicas.

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, pp. 195-225.

Romer, P. M. (1986): «Increasing Returns and Long-Run Growth», *Journal of Political Economy* 94 (5), pp. 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), pp. 25-48.

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), pp. 244-257.

Shyrock, H. S., and Siegel, J. (1973): *The Methods and Materials of Demography*. Washington, DC: 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, pp. 247-286.

Silvestre, J. (2005): «Internal Migrations in Spain, 1877-1930», *European Review of Economic History* 9 (2), pp. 233-265.

Spoorenberg, T., and Dutreuilh, C. (2007): «Quality of Age Reporting and Application of the Modified Whipple's Index», *Population* 62 (4), pp. 729-741.

Squicciarini, M., and Voigtländer, N. (2015): «Human Capital and Industrialization: Evidence From the Age of Enlightenment», *Quarterly Journal of Economics* 130 (4), pp. 1825-1883.

Stolz, Y., Baten, J., and Reis, J. (2013): «Portuguese Living Standards 1720-1980 in European Comparison: Heights, Income and Human Capital», *The Economic History Review* 66 (2), pp. 545-578.

Szoltysek, M., Poniat, R., and Gruber, S. (2017): «Age Heaping Patterns in Mosaic Data», *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 51 (1), pp. 13-38.

Tirado, D. A., and Martinez-Galarraga, J. (2008): «Una nova estimació retrospectiva del VAB regional industrial. Espanya (1860-1930)», UB Economics Working Papers, E08/192.

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), pp. 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, pp. 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, pp. 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), pp. 208-250.

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

APPENDIX

TABLE A1
AGE-HEAPING IN SPAIN BY AGE GROUP AND SEX 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: Calculated with data from INE.

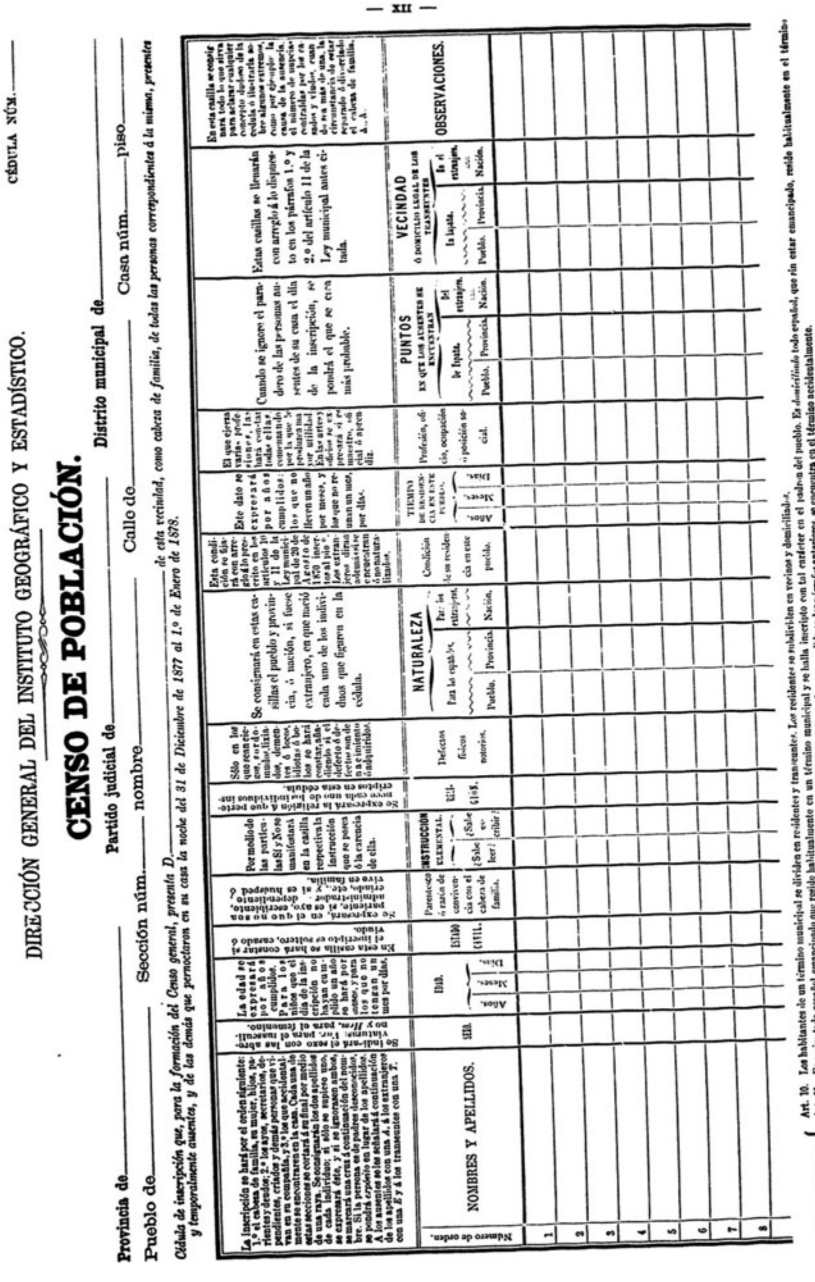
TABLE A2
AGE-HEAPING IN ITALY BY AGE GROUP AND SEX IN THE POPULATION CENSUS OF 1871

	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)

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

FIGURE A1.
REGISTRATION CARD (*CéDULA DE INSCRIPCIÓN*), POPULATION CENSUS OF 1877.



— XII —

DIRECCIÓN GENERAL DEL INSTITUTO GEOGRÁFICO Y ESTADÍSTICO.

CÉDULA NÚM.

CENSO DE POBLACIÓN.

Provincia de — Partido judicial de — Distrito municipal de

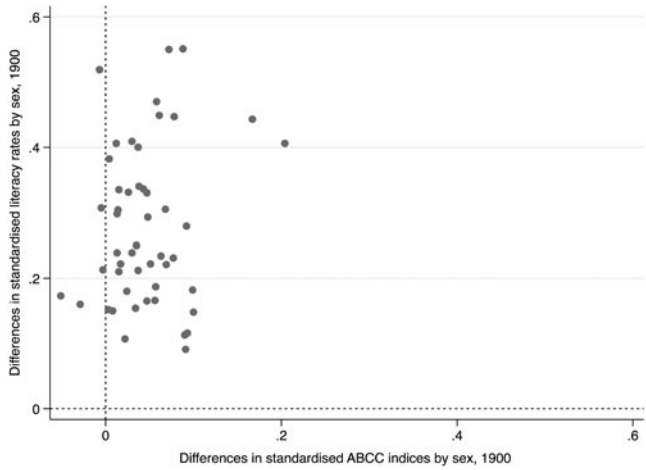
Pueblo de — Sección núm. — nombre — Calle de — Casa núm. — piso

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

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press

FIGURE A2.
DIFFERENCES IN AGE-HEAPING AND LITERACY BY SEX AND PROVINCE IN SPAIN, 1900.



Notes: ABCC indices (per cent) and literacy rates (per cent) by sex and province have been standardised using a fixed range. The ABCC indices range was 75-100 while for literacy it was from 10 to 100 per cent. The *y*-axis reflects the differences in literacy while the *x*-axis shows the differences in age-heaping.

Source: Calculated with data from INE.

https://doi.org/10.1017/S0212610921000033 Published online by Cambridge University Press