

**Artículo de revista:**

Pujadas-Mora, Joana María & Pérez-Artés, María Carmen (2023). Numeracy and consistency in age declarations: a case study on nineteenth and twentieth century Catalonia. *Cliometrica* (2023). (ISSN 1863-2513)

<https://doi.org/10.1007/s11698-023-00277-w>



# Numeracy and consistency in age declarations: a case study on nineteenth and twentieth century Catalonia

Joana María Pujadas-Mora<sup>1</sup> · María Carmen Pérez-Artés<sup>2</sup>

Received: 7 March 2023 / Accepted: 28 October 2023  
© The Author(s) 2023

## Abstract

A rich tradition exists on measuring human capital in historical societies through basic mathematical skills using the method of age heaping. In this article, we novelty propose to calculate it also through the analysis of the consistency in individual consecutive age declaration statements in five towns of Baix Llobregat's county (Barcelona, Spain) during the nineteenth and early twentieth centuries. The region underwent a proto-industrialization in the eighteenth century and was one of the main wool and cotton textile producers in the nineteenth century in Catalonia. Our main results show that (I) numeracy levels in this county were above the Spanish average in the whole period and, during the first years of the nineteenth century, ahead of countries such as Italy, Portugal and the USA. However, its numeracy dropped from 1880's due the early industrialization of Catalonia in the Spanish and Southern Europe context, the existence of child labor in the factories and that also coincided with the phylloxera crisis. (II) To this effect, women presented lower levels of numeracy than men, but their levels tended to converge. (III) As regards emigrants, those who came to the rest of Spain had the lowest level of numeracy compared to the natives. (IV) Moreover, we found that literacy, the 1901 education reform and a higher socioeconomic status had a positive effect on reporting ages consistently over time. (V) Finally, we prove a positive association between systematically reporting ages ending in 0 or 5 and having a higher individual error in age statements along the life course.

**Keywords** Human capital · Numeracy · Catalonia · Nineteenth century · Twentieth century · Life course

**JEL Classification** N33 · N34 · J10

---

Extended author information available on the last page of the article

## 1 Introduction

The importance of human capital for economic growth has been a subject of study since the 1960s. Following the works by Schultz (1961) and Becker (1962, 1964), numerous studies have analyzed the influence of human capital on the development of countries and regions (Mincer 1974; Romer 1989; Barro and Lee 1996; Núñez 1992; Beltrán Tapia et al. 2019). Traditionally, the proxy used to estimate human capital levels has been the ability to sign one's name, understood as showing the ability to read and write (Viñao Frago 1999; Rodríguez and Ben-nassar 1978; Vincent 1987; de la Pascua Sánchez 1989). Numeracy, i.e., basic math skills, has also proved to be one of the best indicators for estimating human capital. Measuring age heaping (the rounding up of ages ending in 0 or 5) helps ascertain the percentage of population who declare their age correctly. As the required data are people's declared age, this is a high the universal information from highly diverse and widely available sources. Since the pre-industrial period, documents have been kept that record people's age, such as censuses, but also marriage and death registers, lists of immigrants and more.

In the case of Spain, ages were always recorded in the censuses and more specifically in the population registers, known as *padrones de población*, that were established through municipal initiatives around the country in the early nineteenth century. National censuses were also definitively introduced from the second half of the nineteenth century onwards. Both the population registers and censuses were the product of the developing Liberal State (Salas-Vives and Pujadas-Mora 2021) seeking universal coverage (Thorvaldsen 2018). This is not the case for other sources, such as signed documents, book production and school registers, which are not always available homogeneously over time and for the whole population (Barro 1991; Baten and Van Zanden 2008). Furthermore, a lack of specific sources on literacy has led to the increasing acceptance of numeracy since the pioneering work of A'Hearn et al. (2009) showed that age heaping estimation was a valid indicator for assessing numeracy levels and thus a society's human capital. Since then, the body of research conducted using this methodology has not stopped growing. Furthermore, results from numeracy studies have been compared to literacy and schooling, finding a high degree of correlation between both (Tollnek and Baten 2016). In a recent article, Baten et al. (2022) once again test the association between numeracy and literacy in Italy and between the tendency for age heaping in parents' ages and the results of math tests performed by their children in African countries.

As has been noted, the level of numeracy in a given population is usually measured using aggregated individual age declarations in demographic sources from both a cross-sectional and longitudinal perspective. Mostly numeracy has been analyzed for groups of people with common characteristics, such as year/cohort of birth, origin, occupation or religion, among others (Crayen and Baten 2010; Tollnek and Baten 2017; Juif and Quiroga 2019; Juif et al. 2020). To date, only two studies have examined numeracy and its progress over the life course individually. In the case of the article by Blum and Krauss (2018), age

declarations by the same individual are compared across two different registers from a sample of 162 immigrants traveling from the Holy Roman Empire to the Kingdom of Hungary in the 1780s. Using the lists of emigrants and locating these individuals in ecclesiastical registers and family inheritance books, they conclude that there is a correlation between the age heaping and the accuracy of individual declarations. The most recent paper by Baten and Nalle (2022) uses the testimony of the defendants during the Spanish Inquisition who had to defend themselves by reporting their lives at different ages and declaring their current age. In this way, it is possible to know whether they were able to keep this account accurate. Comparing a sample of 218 defendants born between 1426 and 1687, the authors argue that those who were off by more than 5 years tended to age heaping more.

Hence, we present the individual longitudinal perspective as an analytical opportunity for understanding the development of numeracy among populations in the past. This is because an individual can be followed in the censuses and population registers over time and place and hence, their individual development of numeracy. Indeed, as well as measuring attraction to the numbers 0 and 5 in age declarations, the study also seeks to measure the consistency of age declarations throughout individuals' lives. Although it is true that consistency involves an accurate declaration of age, a numerate individual is unlikely to lack coherence between different declarations; thus, we analyze the association between numeracy (or not age heaping to numbers ending in 0 or 5) and declaring age consistently over time. To do this, we use the individual population register *par excellence*, the so-called *padrones*, while providing new series on numeracy, not calculated until now, for a sample from five municipalities in the Catalan region now known as Baix Llobregat, bordering the city of Barcelona, from 1857 to 1950. This region was notable for its early industrialization compared to the rest of Spain, except the Basque Country and southern Europe (Martínez-Galarraga and Prat 2016; Brea-Martínez and Pujadas-Mora 2017).

In line with Blum and Krauss (2018), this analysis also provides additional evidence for current debates on the viability of using numeracy as a proxy for basic math skills. Furthermore, yet, there have been no studies in Spain analyzing numeracy at the municipal level beyond the eighteenth century. This is also true for other countries taking into account the whole population, although international estimates based on random samples are available (Crayen and Baten 2010). Therefore, this research is pioneering in that it studies numeracy levels in Spain in municipalities and for all inhabitants aged 23–72 in the nineteenth century and the first half of the twentieth century from the cross-sectional and longitudinal perspectives. In total, we analyze 25,394 individual age declarations in five different municipalities in the same region next to Barcelona from 1857 to 1950, using Baix Llobregat Database (BALL Database).<sup>1</sup>

The article is structured into the following sections. The first section provides a historiographical review of numeracy in Spain and presents the geographical area of

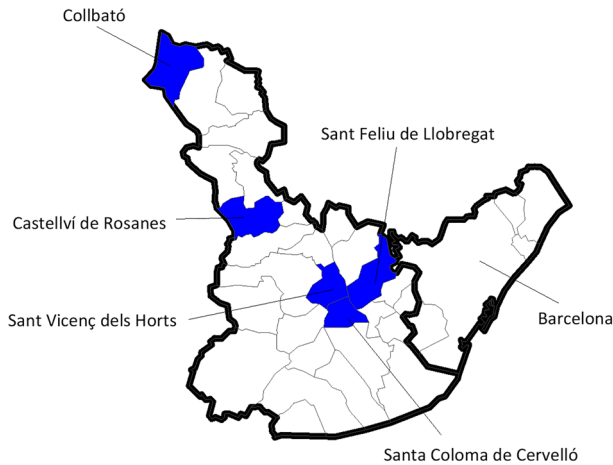
<sup>1</sup> It should be borne in mind that the first preserved population register for the city of Barcelona dates from 1930, making the type of study we propose impossible.

the study. The next section presents the source of our data, the BALL database. This collects individual census data from both population registers and national censuses and portrays the descriptive and multivariable statistical methods applied. The subsequent sections explain the results of the study, focusing on estimates of numeracy and error levels, the consistency of the declared age over the life course and the relation between both. Finally, the article closes with conclusions and discussion and suggests new lines of research opened up by the results.

## 2 Numeracy in Spain and the geographical scope of the study

In the case of Spain, the study of numeracy is still an incipient topic. It is worth highlighting the study by Juif et al. (2020), which shows that, at the end of the fifteenth and eighteenth centuries, the elite groups of the catholic majority had a numeracy level similar to that of the Jewish population on average. Furthermore, Pérez-Artés and Baten (2021) show that the regions characterized by medium-sized farms developed better than the regions with large landowners and many unskilled agricultural day laborers who did not own their own land. In a more recent paper, Pérez-Artés (2023) finds that Spaniards who left the country to settle in Hispanic America were positively self-selected in terms of numeracy. In Catalonia, the study area of this article, Gómez-i-Aznar's study (2019), estimates numeracy levels from information on declared ages in individual population registers dating from 1716 to 1724 for 13 municipalities throughout Catalonia. The results suggest that the level of numeracy in Catalonia was relatively high before industrialization. Indeed, 73% of the population already correctly declared their age around 1700, higher than Catholic Germany (68%) and Switzerland (66%).

For Spain, there are also studies analyzing numeracy and literacy levels jointly, such as Álvarez and Ramos Palencia (2018) and Beltrán Tapia et al. (2022). The first authors analyze different municipalities in the provinces of Madrid, Guadalajara and Palencia around 1750, concluding that the level of human capital may have contributed to income inequality. The research by Beltrán Tapia et al. (2022) compares the six national population censuses from 1877 to 1930 at the provincial level. Their analyses conclude that while literacy started rising from the start of the period, numeracy remained stable and did not increase until the early twentieth century. If Catalonia was part of the great leap forward in literacy levels from 1860 to 1900, reaching a literacy rate of 40% of the population by the latter date, this was thanks to the work of the Ministry of Public Instruction in 1900, which helped reduce the literacy gap among municipalities (Beltrán Tapia et al. 2019; Núñez 1992). Indeed, the first general education act in Spain had been passed decades earlier, in 1857. This law established obligatory schooling from ages 6 to 9, later extended to 12 in 1901 and then to 14 in 1923. It was not until 1990 that it was raised to 16 (de Gabriel 1997; Tiana Ferrer 1987; Viñao Frago 2004). In addition, from 1902, the Ministry of Public Instruction and Fine Arts assumed responsibility of paying for staff and materials in state primary schools, which would help homogenize municipalities in terms of education (Gaceta de Madrid 1901). Thus, the generalization of literacy was a gradual process involving different education acts, although this does



**Fig. 1** Map of the municipalities analyzed in Baix Llobregat region. Province of Barcelona

not mean people with no compulsory schooling were unable to acquire the skill individually and later in life.

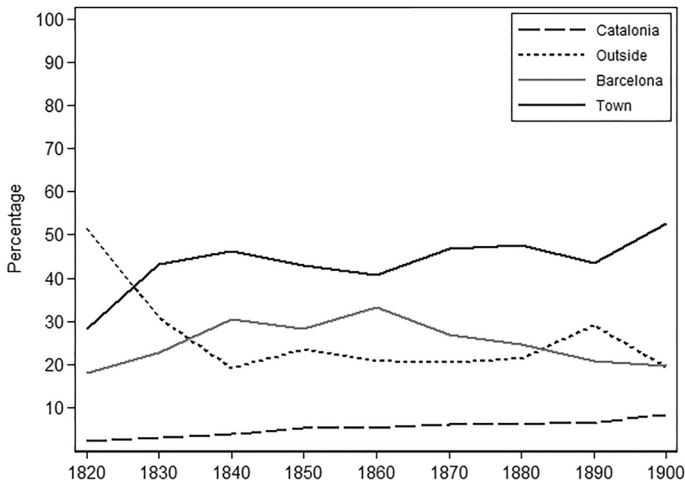
Our research broadens these numeracy results in Spain with an analysis of the inhabitants of the region of Baix Llobregat in Barcelona, a province that experienced early industrialization, particularly in the cotton industry (Fig. 1). Since 1736, a highly distinctive cotton manufacturing sector developed in Barcelona, and in the 1790s, the use of spinning mills became widespread. Catalan producers adopted early technology from the British industrial revolution, focusing technological changes in a series of dynamic sectors. Together with other factors like market size, this transformed Barcelona into a significant industrial center during the nineteenth century (Martínez-Galarraga and Prat 2016). However, the growth was hindered by the Carlist War and the Liberal Revolution, leading to a slowdown until the following decade. Subsequently, there were multiple cycles of expansion. During the 1860s, Catalan fabrics became more competitive than the French and British ones in the Spanish market. This led to progress not only in the cotton industry but also in other sectors such as wool (Carreras 1990). From this point onwards, cotton establishments began to settle along the riverbanks due to the utilization of hydraulic energy, particularly in the basins of the Ter and Llobregat rivers, being an alternative to coal, given its scarcity. In 1861, the Llobregat basin accounted for 21.83% of the total number of spindles and 12.68% of the cotton looms in Catalonia. Seventy years later, these percentages had risen to 41.21% and 37.81%, respectively (Nadal 1992:115). The industrial importance of this area is also evidenced by the presence in Sabadell of immigrant textile and industrial workers who had previously lived in the municipalities of Baix Llobregat in the nineteenth century (Camps 1992).

On the other hand, this region experienced a boom in vineyards as a commercial crop from the second half of the eighteenth century onwards. In fact, the municipalities on the left border are currently part of the “Penedès Denomination of origin”, including Castellví de Rosanes, which is included in our sample.

By the mid-nineteenth century, 76% of the 9800 hectares (i.e., 7500 hectares) in the river delta were arable land. In this sense, it is important to highlight how in 1855 the works of the Canal de Riego de la Derecha del Llobregat-irrigation canal of the Llobregat river began, following the Infanta canal that had been built between 1817 and 1820 (Retuerta Jiménez 2012). Among the municipalities encompassed by this latter canal, it is noteworthy that only Sant Feliu de Llobregat falls within the scope of our study. Conversely, the more contemporary canal originates from Sant Vicenç dels Horts, traversing the localities of Santa Coloma de Cervelló, Sant Boi, and Prat de Llobregat (Rocamonde Lourido and Sabaté, 2019; Alba Molina and Aso Pérez 2008). In fact, these canals marked a significant change, catalyzing the evolution from rain-fed, expansive farming practices to a more intensive, irrigation-based approach. This transformation focused on the cultivation of orchards, fruit trees and rice. As a result, these cities assumed the role of pantry for Barcelona, gradually expanding their influence to encompass international markets. This trend was accentuated especially in the second decade of the twentieth century, as will be consolidated in its flourishing exports to Europe. Thus, the irrigation increased the productivity of agriculture in the region beyond the Llobregat Delta. From then on, investment in the region began to be productive, leading to a process of concentration of property ownership among the Barcelona bourgeoisie and nobility, who had the capitalist mentality and the economic means to do so (Codina i Vilá 1995). Between 1885 and 1900, however, the Baix Llobregat region fell victim of the phylloxera plague, which had a negative impact on the growth of the wine industry (Recaño Valverde 1995). We therefore argue that our study considers not only industrialized municipalities, but also agricultural ones, as the total population of the region in 1900 was split between 53% urban population and 47% industrial population (Camps 1995).

More specifically, we analyze the municipalities of Collbató, Castellví de Rosanes, Santa Coloma de Cervelló, Sant Feliu de Llobregat and Sant Vicenç dels Horts (Fig. 1). In 1860, a total of 46,261 individuals lived in Baix Llobregat, of whom 316 were inhabitants of Castellví de Rosanes, 859 of Collbató, 2478 of Sant Feliu de Llobregat, 1732 of San Vicenç dels Horts and 192 of Santa Coloma de Cervelló (Codina i Vilá 1995). It was mainly the municipalities in the south of the region (in our sample, Sant Feliu de Llobregat, Santa Coloma de Cervelló and Sant Vicenç dels Horts) that had a strongly commercial agricultural economy and industry from the end of the nineteenth century, linked to the economic expansion of the Barcelona. This period is also characterized by a decline in mortality and sharp decline in birth rates, resulting in little natural growth but compensated by an influx of immigration. This constant flow of immigrants rejuvenated Baix Llobregat population in the twentieth century. The region's proximity to the capital boosted its growth and economic transformation during the nineteenth century (Recaño Valverde 1995). Besides this proximity, a further factor to bear in mind is the construction of the Martorell railway line (17 km north-east of Sant Feliu de Llobregat) and its impact on integrating the regional economy with that of the capital (Calvo 1995).

In addition, it is worth noting that Sant Feliu de Llobregat had already been the capital of the judicial district since. In 1862, 57.5% of the working population was employed in agriculture, 34.1% in manufacturing and crafts, and 7.4% in services



**Fig. 2** Percentage of population by origin and birth decade in Baix Llobregat (1820–1900). Source: BALL Database

(Carbonell i Porro 1995). On the other hand, in 1881, according to BALL database data, the percentage of men who reported being employed in the primary sector was only 29%, dropping to between 19 and 17% between 1906 and 1936, while at the latter date, the secondary sector accounted for 38%. These figures show a switch from a purely agrarian to an industrial society occurred in this municipality, a process that had been consolidated by 1920's. The business structure also played a very important role. The two biggest textile companies, Bertrand and Solà Sert, employed over 56% of the workers in the sector in 1923 (Carbonell i Porro 1995). Thus, while following in the footsteps of the capital toward industrialization, this was a heterogeneous region in which the small, more distant villages remained agricultural. These steps were followed by the municipality Santa Coloma de Cervelló since in 1936, its primary sector represented 26% and the secondary sector 44% since the Güell's colony was located in the municipality. This colony was established in 1890 due to the relocation of a factory from the Sans neighborhood in Barcelona. However, Collbató stands as a municipality characterized by an occupational landscape primarily centered around agriculture. The male workforce engaged in the primary sector consistently constituted more than 70% until 1936, when this proportion declined to 64%, accompanied by a corresponding surge in the tertiary sector to 16%—a notable shift from its previous level of under 5%. A similar pattern emerges in the case of Sant Vicenç dels Horts, where the primary sector still represented a substantial 55% in 1935. This trend is also evident in the municipality of Castellví de Rosanes, despite a more pronounced emphasis on the primary sector. Remarkably, in 1936, a staggering 93% of men reported their occupation as agriculture related.

Furthermore, it is crucial to take into account the migratory composition of the population to be analyzed given its establish correlation with numeracy levels, as we will explore in the following section. In this sense, to get an idea of what the native and migrant population represents in our sample, we have calculated the percentage



of people between 23 and 72 years of age, our reference population, in each cohort according to their origin (Fig. 2). Given the industrialization of part of the municipalities, as we have explained above, there is a continuous arrival of people of working age, so it is not surprising that Baix Llobregat has a higher percentage of internal migrants than autochthonous people. Among individuals aged 23–72 years, less than 50% stated that they were originally from the municipality where they lived, followed by those coming from other municipalities in the province of Barcelona or from outside Catalonia. Lastly, the population comes from the other three Catalan provinces: Girona, Lleida and Tarragona. These figures would indicate a high migratory component of the populations studied, basically marked by internal migrations that have already been described for other industrial municipalities in Catalonia such as Sabadell (Camps 1992).

### 3 The BALL database and methodology

As mentioned in the introduction, from the so-called BALL Database, we chose the population registers of five municipalities in Baix Llobregat that was economically and demographically representative of the province of Barcelona, with the aim of calculating numeracy levels and the consistency of individual age declarations as a proxy for human capital in Catalonia in the nineteenth and twentieth centuries. We used a total of 30 population registers and censuses dating from 1857 to 1950, representing 25,394 individual observations from 19,014 individuals aged 23–72.<sup>2</sup> Table 1 shows the breakdown of the number of individual declarations by municipality and year of the census. Not included are population registers from the first half of the nineteenth century, as individuals' ages and dates of birth were not systematically recorded. The municipalities for which we have the most population registers with age declarations are Collbató and Sant Feliu de Llobregat, with nine population registers from different years for each town.<sup>3</sup> Notably in the latter town, there was a striking jump from 310 declarations in 1857 to 1056 in 1878. This is explained by its 159% growth in population between 1862 and 1877, coinciding with the most important period of industrialization in the municipality (Carbonell i Porro 1995).

In Spain, population registers were introduced by Decree on 3 February 1823. They were carried out by municipal initiative, so were not initially conducted simultaneously or at a set frequency throughout the country. The Municipal Act of 20 August 1870 established a period of every 5 years to conduct a new individual population register. However, this was not generally observed until the early twentieth century. Throughout the nineteenth century, the population registers evolved from simple lists of inhabitants, hardly distinguishing between households, to become full registers from the second half of the century, including the full name, age or date of birth, marital status and occupation of each individual, as well as their family or work relationships to the head of the household and the full address. For some

<sup>2</sup> From now on, we use the words census or individual population register as synonyms.

<sup>3</sup> We use the terms "town" and "municipality" as synonyms.

Table 1 Number of observations by town and year of census

Town/ Year	1857	1866	1878	1880	1881	1896	1900	1901	1905	1906	1910	1915	1916	1920	1921	1924	1930	1935	1936	1940	1945	1950	Total
COLL				242		274	254	267			278		285	251		252			217				2,320
CR	160															162	162		147	157	146	174	1,108
SCC								252											790				1,042
SFLL	310		1,056		1,383					1,704	1,857	2,166		2,245		2,614			3,497				16,832
SVH															992		1,486	1,614					4,092
Total	310	160	1,056	242	1,383	274	254	252	267	1,704	2,135	2,166	285	2,496	992	3,028	1,648	1,614	4,651	157	146	174	25,394

Source: BALL Database. (COLL) Collbató, (CR) Castellví de Rosanes, (SCC) Santa Coloma de Cervelló, (SFLL) Sant Feliu de Llobregat and (SVH) Sant Vicenç dels Horts. The years 1857, 1900, 1910, 1920, 1930, 1940 and 1950 correspond to national censuses. The other years refer to municipal population registers

**Table 2** Sociodemographic characteristics reported in the population registers and national censuses of the municipalities in Baix Llobregat (Barcelona)

Variables	Year of population registers/national census and Municipalities															
	1857		1866		1878		1880		1881		1896		1900		1901	
	SFLL	CR	SFLL	CR	SFLL	CR	COLL	SFLL	COLL	SFLL	COLL	COLL	COLL	SFLL	COLL	COLL
Street name		×	×		×		×		×		×		×		×	
Number of the household		×	×		×		×		×		×		×		×	
Name	×		×		×		×		×		×		×		×	
First surname	×		×		×		×		×		×		×		×	
Second surname	×		×		×		×		×		×		×		×	
Relation to the head of household	×		×		×		×		×		×		×		×	
Occupation		×														
Sex																
Marital status	×		×		×		×		×		×		×		×	
Date of birth					×		×		×		×		×		×	
Age	×		×		×		×		×		×		×		×	
Citizenship																
Literacy																
Income or wage																
Municipality and province of usual residence																
Municipality and province of birth																
Territorial/Industrial contribution																
Boss for whom you work																
Years living in the municipality																
Classification of inhabitant																

**Table 2** (continued)

Variables	Year of population registers/national census and Municipalities											
	1921	1924	1930	1935	1936	1940	1945	1950				
	SVH	SFLL	COLL	CR	SVH	SFLL	COLL	SCC	CR	CR	CR	CR
Street name	×	×	×	×	×	×	×	×	×	×	×	×
Number of the house-hold	×	×	×	×	×	×	×	×	×	×	×	×
Name	×	×	×	×	×	×	×	×	×	×	×	×
First sur-name	×	×	×	×	×	×	×	×	×	×	×	×
Second surname	×	×	×	×	×	×	×	×	×	×	×	×
Relation to the Head of house-hold	×	×	×	×	×	×	×	×	×	×	×	×
Occupation	×	×	×	×	×	×	×	×	×	×	×	×
Sex		×	×	×	×	×	×	×	×	×	×	×
Marital status	×	×	×	×	×	×	×	×	×	×	×	×
Date of birth	×		×		×	×	×	×	×	×	×	×
Age	×	×	×	×	×	×	×	×	×	×	×	×
Citizenship		×	×	×	×	×	×	×	×	×	×	×
Literacy	×	×	×	×	×	×	×	×	×	×	×	×
Income or wage			×	×								

**Table 2** (continued)

Variables	Year of population registers/national census and Municipalities									
	1921	1924	1924	1930	1935	1936	1936	1940	1945	1950
Municipality and province of usual residence	×	×	×	×	×	×	×	×	×	×
Municipality and province of birth	×	×	×	×	×	×	×	×	×	×
Territorial/Industrial contribution										
Boss for whom you work					×	×	×	×		
Years living in the municipality	×	×	×	×	×	×	×	×	×	×
Classification of inhabitant	×	×	×	×	×	×	×	×	×	×

Source: BALL Database. SFL, Sant Feliu de Llobregat; CR, Castellví de Rosanes; COLL, Collbató; SCC, Santa Coloma de Cervelló; SVH, Sant Vicenç dels Horts

periods and municipalities, information is also available on individuals' ability to read and write, their migratory situation and income (see Table 2). There was no legislation to standardize the content of the population registers until the Municipal Statute of 8 March 1924 (García Pérez 2007; García Ruipérez 2012).

The population registers are almost the only census data to have been preserved in Spain containing individual data, which are essential for our research aim objectives. Individual sheets from national censuses of municipalities were generally destroyed; once the population had been counted, the main variables aggregated. The first national census in Spain was conducted in 1857. By 1950, a further 10 censuses had been conducted, dating from 1860, 1877, 1887, 1897, 1900, 1910, 1920, 1930, 1940 and 1950 (Reher and Valero Lobo 1995). Exceptionally, we found census sheets with individual information for the municipalities of Sant Feliu de Llobregat for the years 1857, 1910 and 1920; Collbató for 1897, 1900, 1910 and 1920; Sant Vicenç dels Horts for 1930 only, and Castellví de Rosanes for 1930, 1940 and 1950, as also indicated in Table 1. The censuses provide more information on socioeconomic indicators and literacy than the population registers, which is also extremely useful for the purposes of our analysis (see Table 2). Thus, individual observations from the population registers together with the censuses permit the compilation of various individuals and allow the collection of individual numeracy measurements over the life course. It should be noted that the disparity in reference years for the age declarations in the different municipalities does not affect the estimates, given the type of statistical models applied, as explained before.

It is worth stressing that age data were always collected individually in both population registers and censuses for men and women of any age in every municipality included in this study from the first half of the nineteenth century. For this reason, the earlier population registers have not been included, since they do not include age statements (see Table 2). However, date of birth was collected more rarely and always together with age. It is used to obtain two age measurements: one declared and the other exact. It is worth noting that the date of birth variable is more frequently found in the municipal population registers than the national censuses.

These census data, from both population registers and censuses, were collected from the BALL Database, as mentioned previously, using the AI application and Computer Vision in crowdsourcing and gamesourcing environments designed ad hoc (Pujadas-Mora et al. 2022). Using such methodologies help to significantly reduce the time taken to collect the original information, already a time-consuming difficult task helping increase the volume of information. This explains the large number of sociodemographic and economic variables (see Table 2) gathered exhaustively for each citizen who lived in the abovementioned municipalities for their whole lives or at some point in their lives. It also helps track individual changes in age declarations. It is worth noting that the database has not been used before for this type of study, making it an original aspect of our paper.

For the article, these data underwent both descriptive and multivariable analyses in order to present the evolution of numeracy from the cross-sectional and aggregated perspectives and to model the accuracy of the age declaration over time from a longitudinal perspective. Thus, we first conducted a cross-sectional analysis using the so-called ABCC Index in order to assess the level of numeracy by municipality for the year in which the population registers were conducted, followed by a longitudinal analysis by cohort. Secondly, we tracked individuals over time to assess the consistency or not of their age declarations, based on differences in declared age between one register or census and the next. We also estimated the association between this inconsistency in successive declarations and age declarations ending in 0 and 5. All these analyses were controlled for a number of sociodemographic variables, as explained below.

We calculated the ABCC Index based on the age heaping to estimate the level of numeracy (A'Hearn et al. 2009).<sup>4</sup> This index is a modification of the Whipple Index (1) which ranges from 100 to 500, where the latter value indicates that all individuals declared in multiples of 5, while the ABCC Index (2) varies from 0 to 100, the latter value indicating that no age heaping was noted and therefore, 100% of the individuals would have declared their age correctly. As is well established, this index only measures the phenomenon of age heaping for individuals aged between 23 and 72 (Crayen and Baten 2010; Tollnek and Baten 2016). Thus, we also corrected the age effect, as younger generations seem to systematically age heap to a lesser extent. Consequently, 0.2 Whipple units were added to the 23–32-year-old age group for each Whipple unit over 100 in the 33–42 age group, as proposed by Crayen and Baten (2010):

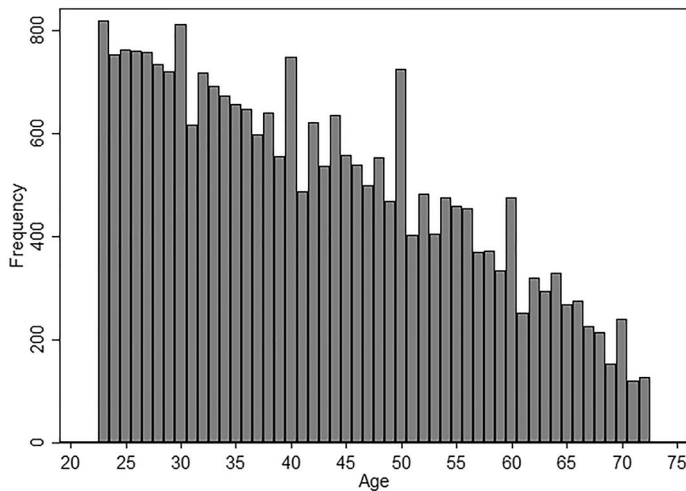
$$Wh = \left( \frac{(Age25 + Age30 + Age35 + \dots + Age60)}{\frac{1}{5} \times (Age23 + Age24 + Age25 + \dots + Age62)} \right) \quad (1)$$

$$ABCC = \left( 1 - \frac{(Wh - 100)}{400} \right) \times 100 \text{ if } Wh \geq 100; \text{ else } ABCC = 100 \quad (2)$$

Figure 3 shows the number of single age declarations for all individuals aged 23–72 in our database. As can be seen, there is a clear tendency to age heap, especially to digits ending in 0. The number 50 is particularly striking: around 750 people declared their age as 50, while a little over 400 declared it as 49 and barely this number declared it as 51. The same pattern is repeated at 30, 40, 60 and 70.

With regard to tracking individuals from the municipalities of Collbató, Castellví de Rosanes, Santa Coloma de Cervelló, Sant Feliu de Llobregat and Sant Vicenç dels Horts, a total of 19,223 declarations were gathered from 9427 individuals who declared their age and whose information is complete (or it can be assigned in the case of women) for the relevant variables, such as occupation, sex, origin and

<sup>4</sup> This index comes from the initials of the authors' surnames plus that of Gregory Clark, who provided his advice on the matter.



**Fig. 3** Number of individual age declarations from 1857 to 1950 for the complete sample. Source: BALL Database

**Table 3** Number of age inconsistency observations per number of individuals

Town	1	2	3	4	5	6	7	8	Total
COLL	276	204	210	232	260	318	301	88	1889
CR	82	92	102	112	250	0	0	0	638
SCC	694	32	0	0	0	0	0	0	726
SFLL	3148	1778	2253	2132	2510	1764	0	0	13,585
SVH	1006	1376	3	0	0	0	0	0	2385
Total	5206	3482	2568	2476	3020	2082	301	88	19,223

Source: BALL Database. SFLL: Sant Feliu de Llobregat; CR: Castellví de Rosanes; COLL: Collbató; SCC: Santa Coloma de Cervelló; SVH: Sant Vicenç dels Horts

Note: 1 observation in the model corresponds to two consecutive observations –age declarations– in time, while 2 correspond to 3 and so on

whether they could read or write (see Table 3). A 93% of those total individuals showed to have from two to eight age declarations. By municipality, Collbató and Sant Feliu de Llobregat showed more individuals with complete information.

The longitudinal approach to numeracy used inconsistency of age based on the difference between declared age in a population register or census and the previous one (3). It also took into account the number of years elapsed between both censuses or population registers and turned them into an absolute value (negative values become positive by considering the consistency error as the total distance from the correct value). Thus, the error in an individual's age declaration and in a census year  $j$  is formulated in the following way:



$$\text{Inconsistency in declared age}_{ij} = \text{absolute error} [(age_{ij} - age_{i(j-1)}) - (year_{ij} - year_{i(j-1)})] \quad (3)$$

where  $age_{ij}$  is an individual's age declaration ( $i$ ) at the census year  $j$ , while  $age_{i(j-1)}$  corresponds to the age declaration of the same individual in the previous census year ( $j-1$ ),  $year_{ij}$  is the calendar year in which the census was carried out, and  $year_{i(j-1)}$  refers to the year for the previous census.

One feature of this method is that the first age declaration of all individuals is lost. To counteract the problem, if an individual also declared their year of birth, the difference between the declared age and the subtraction between the census year and the declared year of birth was calculated and transformed to an absolute value (4), i.e.:

$$\text{Inconsistency in declared age}_{ij} = \text{absolute error} [age_{ij} - (year_{ij} - birth_{ij})] \quad (4)$$

where  $age_{ij}$  is an individual's age declaration ( $i$ ) at the census year  $j$ , while  $age_{i(j-1)}$ ,  $year_{ij}$  is the calendar year in which the census was carried out and  $birth_{ij}$  corresponds to the individual's birth year.

Declarations differing by a year are not calculated as a lack of consistency, as age declaration is a declaration of completed age between an exact age ( $x$ ) and the next exact age ( $x+1$ ), i.e., between one year and the next. Thus, the value 0 is assigned to those declarations that differ by 0 or one unit from the correct age, and 1 to those that differ by 2 units, 2 to those that differ by 3 units, and so on. This provides the metric to calculate the inconsistency of the age declaration. The metric does not carry over errors made in the previous declaration, as it only measures the error between two consecutive age declarations.

Furthermore, this inconsistency in declared age is a dependent variable that is later modeled with random effect Poisson regression models, given that this is a count variable (absolute error between consecutive ages/dates of birth). This type of regression means variability of a single individual can be controlled, as it uses different age declarations at different moments, as well as individuals with many declarations and others with few. This model uses the following formula (5):

$$\log(E[\text{Error}_{ij} / X]) = B0 + XB + u_{ij} \quad (5)$$

Note: The logarithmic transformation is intended to linearize the distribution of our original dependent variable, as it is positively skewed and contains a large proportion of zeros, following the approach of Wu and Little (2011).

Where:  $E[\text{Error}_{ij} / X]$  is the expected mean error for the individual between two census year where  $j$  corresponds to the second one, taking into account the independent variables  $X$ ;  $B0$  is the regression intercept,  $B$  the regression coefficients; and finally,  $u_{ij}$  is the random error for the individual  $i$  at the time  $j$ . Note that  $\text{Error}_{ij} \sim \text{Pois}(\lambda)$  and  $u_{ij} \sim N(0, \sigma^2)$ , with  $\lambda$  being the mean inconsistency for the individual and with the observations  $X$  and  $\sigma^2$  being the variance of the error for all observations.

The specific independent variables ( $X$ ) considered for this study are:

- Year: the year in which the census was conducted (numerical).
- Educational reform: reform in education in 1901. This considers a 5-year delay in the complete application of the reform (category “No” before 1906; “Yes” for 1906 and later). With this variable, we control for the possible effect of the 3-year increase (from 9 to 12) in obligatory schooling and its influence on raising literacy. Indeed, there was a major rise in literacy in Catalonia from 1900 (48% literacy among the population) to 1910 (58%) (Núñez 1992: 132). This date is also the one Beltrán Tapia et al. (2019) set as the start of the acceleration in the Spanish process and convergence in educational levels in different regions.
- Age: individual’s declared age at the time of the register gathered by 5-year age groups (23–32 years old, 33–42; 43–52; 53–62 and 63–72), which gives us information on whether there is a worsening or improvement in terms of numeracy during the life course of the individuals.
- Literate: individuals stating they can write (category “Yes” if they know how to write, otherwise “No”). Where data on an individual are missing, if they have previously stated they do know how to write, subsequent missing data are also completed as positive, given that this ability is not usually lost over time. Traditionally, literacy is the most widely used proxy for measuring human capital in societies, so its inclusion as a variable makes the analysis more robust.
- Class: Individual occupations have been transformed into social groups based on the HISCLASS classification (van Leeuwen and Maas 2011). The 12 HISCLASS labels have been reorganized into eight distinct classes. This new grouping brings together labels 3, 4 and 6 (Lower manager, Lower professional and Clerical sales, Foremen) under the heading of a single label (3). We have also merged Low-skilled and Unskilled workers in label 7, and Low-skilled farm workers and Unskilled farm workers in label 8. These changes respond to the need to better adjust to the socio-occupational structure of the period. Moreover, this distribution has already been tested in other publications (Pujadas-Mora et al. 2018).
- Sex: Sex of the individual to control for men/women and observe possible differences.
- Migration status: The individual place of origin (category, “Town” for those born in the municipality itself; otherwise “Outside town”). This variable provides information on the importance of human capital with regard to emigration and controls for the possible effect of being an emigrant on the numeracy level in each municipality.
- Municipality of residence: This variable is used to test whether there are differences between municipalities in terms of inconsistency. If it is found that there are statistically significant differences, it is convenient to model the municipalities separately.
- Declaration: Order of the declaration within the individual’s total declarations (numerical count, greater than zero).

Specifically, the proposed model is formulated as follows (6):

$$\begin{aligned}
\log(\text{Error}_{ij}) = & B0 + B1\text{year}_{ij} + B2\text{educational\_reform}_i \\
& + B3\text{declaration\_number}_{ij} + B4\text{age groups}_{ij} \\
& + B5\text{literate}_{ij} + B6\text{social group}_{ij} + B7\text{sex}_i \\
& + B8\text{Migratory status}_{ij} + B9\text{Municipality of residence}_i \\
& + (1 | id) + u_{ij}
\end{aligned} \tag{6}$$

Note: Variables with suffix  $j$  are time varying.

It should be stressed that the model was adjusted for each of the five municipalities to assess whether there were any differences in terms of consistency in the age declaration in relation to their different socioeconomic profiles. Likewise, the average predicted error was also calculated to estimate how changes in the dependent variable (inconsistency in age declaration) relate to changes in a given dependent variable, such as sex, educational reform, with the aim of estimating the partial effects of certain variables. All other co-variables were assumed to be constant.

In addition, another model was derived from the previous one to analyze the relation between declaring an age ending in 0 or 5 and inconsistency in age declarations in population registers. We adjust Poisson regressions for each municipality, adding the independent variable being non-numerate. This takes the value of 1 when the individual declared an age ending in 0 or 5. The model helps test whether there is a correlation between correctly declaring age over the different population registers and the tendency to declare ages in multiples of 5.

The formula for this model is (7):

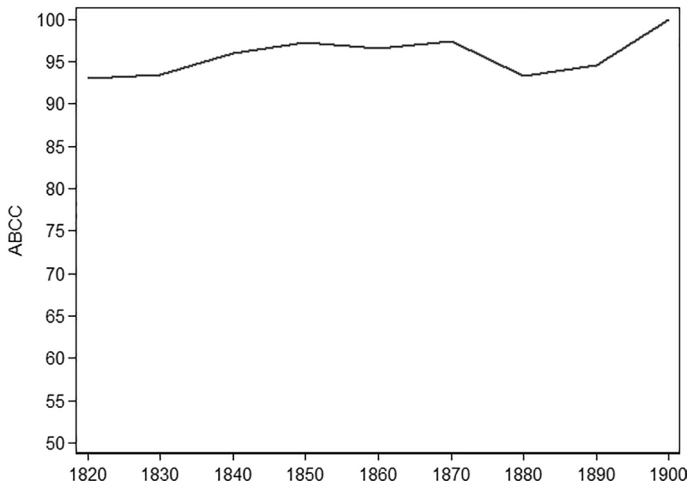
$$\begin{aligned}
\log(\text{Error } i) = & B0 + B1 \text{ report age ending 0/5}_{ij} + B2 \text{ year}_{ij} \\
& + B3 \text{ educational\_reform}_i + B4 \text{ age groups}_{ij} \\
& + B5 \text{ literate}_{ij} + B6 \text{ social group}_{ij} + b7 \text{ sex}_i \\
& + B8\text{Migratory status}_{ij} + B9\text{Municipality of residence}_i \\
& + (1 | id) + u_i
\end{aligned} \tag{7}$$

Note: Variables with suffix  $j$  are time varying.

#### 4 Levels of numeracy and error over a century (1857–1950)

This section provides the descriptive analysis of the two proxies used in the study to estimate the level of human capital through basic mathematical abilities. First of all, the ABCC Indices are calculated for each cohort, sex, origin and compared with other countries. Secondly, calculated individual errors between two declarations at two different moments in the nineteenth and twentieth centuries are shown. Thus, Fig. 4 shows the general trend in the level of numeracy for five the municipalities in our sample by birth decades between 1820 and 1900.<sup>5</sup> It should

<sup>5</sup> Numeracy levels by municipality and decade of birth are shown in Fig. 10 in the Appendix.



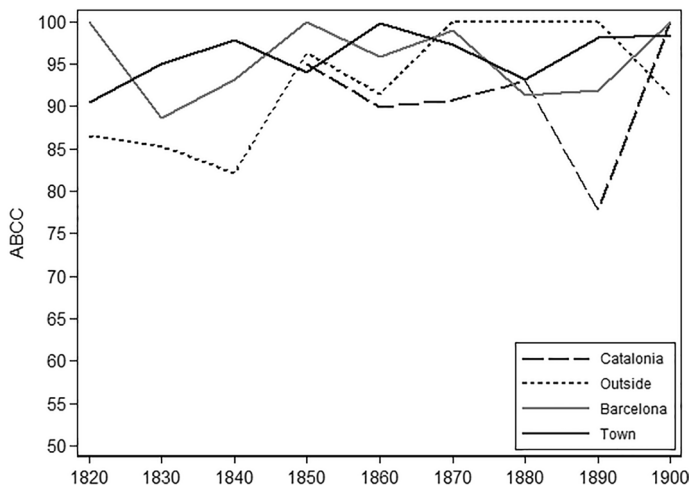
**Fig. 4** ABCC Index by birth decade in Baix Llobregat (1820–1900). Source: BALL Database

be noted that we have at least 600 observations in all the birth decades taken into account. The general trend is of lower numeracy in the nineteenth century than in the twentieth century. Indeed, the drop in numeracy levels, especially in the last two decades of the nineteenth century and early 20th, in line with the results calculated by Beltrán Tapia et al. (2022) for the whole province of Barcelona. This could be due to the following reasons: (I) It was during these decades that the phylloxera plague struck this region during the so-called turn-of-the-century crisis. First detected in France in 1860, the disease spread rapidly and by the end of the 1870s had affected the main producing regions of Spain, leading to both an economic and social crisis. These types of shocks have been shown to affect numeracy as Baten et al. (2014) present in the case of industrialized England between 1780 and 1850, coinciding with the Napoleonic wars. (2) Furthermore, during the years of the phylloxera crisis, the population of the Penedès region fell by  $-0.63\%$  per year, much more than in the rest of Catalonia ( $-0.07\%$ ), as a result of emigration to Barcelona or nearby industrialized areas—where our study municipalities are located—and a decline in the birth rate as part of the evolution of Demographic Transition (Recaño Valverde 1995; Colomé Ferrer and Valls-Junyent 2012). (3) Another reason may have been the importance of child labor in factories at a time when some municipalities in the Baix Llobregat were already industrialized and families did not see the need to send their children to school (Borrás Llop 2002; Iturralde Valls 2017; Reher 2022).

Secondly, we calculate the level of abcc by sex and birth decade as is shown in Fig. 5. In the 1820 cohort, the gap between men and women was 11%. By the 1840s, the gap had narrowed to 1%, with women's numeracy levels 2% higher by the following decade. It is noticeable that the causes mentioned above (positive self-selected emigration, nutrient shortages due to agricultural crises and factory work) seem to have affected men earlier than women, as the female cohorts in 1850 and 1880 had a higher numeracy level than the male cohort, only to decrease significantly in 1890.



**Fig. 5** ABCC Index by birth decade and sex in Baix Llobregat (1820–1900). Source: BALL Database



**Fig. 6** ABCC Index by birth decade and origin in Baix Llobregat (1820–1900). Source: BALL Database

At the end of the period analyzed, both sexes reached 100% of numeracy. Research using numeracy as a proxy for human capital shows that women might have had higher levels than men in specific times and places. For instance, Gómez-i-Aznar (2019) found that in some Catalan municipalities (Olot 1716 and Badalona 1717), women tended to have a level of numeracy 3–4 points higher than men.

We also analyzed the level of numeracy by origin (see Fig. 6). This origin was divided into four categories: non-migrant (i.e., originating from the municipality), migrants from the province of Barcelona, migrants from outside the province of Barcelona but within Catalonia and, finally, migrants from outside Catalonia. The

analysis is based on at least 110 individual observations. In general, the highest levels of numeracy are found among individuals from the same municipality or from the province of Barcelona. Immigrants from outside Catalonia had a lower level of numeracy than the other groups at the beginning of the period, but this trend changed in the last third of the nineteenth century. This could be explained by the presence of emigrants in the province of Barcelona, especially from the Balearic Islands and the Valencian Community, and in some cases from areas that also had specialized textile workers, such as the municipality of Alcoy in Alicante (Camps 1995; Recaño Valverde 1996). This would explain why immigrants from outside Catalonia sometimes had higher numeracy levels than those from Catalonia itself. As the literature shows, individuals working in artisan industries, commerce and administration show higher levels of numeracy (Pérez Artés and Baten 2021). Along these same lines, Beltrán Tapia and Salanova (2017) found that in internal migrations from 1880 to 1930, those who decided to go to Madrid were positively selected in terms of literacy, while those who went to growing neighboring towns were negatively selected.

Since the final objective of this article is to measure the inconsistency of consecutive age statements, we have estimated the mean error in years between two successive declarations. In this sense, younger generations show greater consistency in age declarations, except for cases of Santa Coloma de Cervelló and Castellví de Rosanes (see Fig. 7). In the case of Collbató, we find that the maximum mean individual error of all the municipalities analyzed for those born from 1810 to 1819 of up to a little more than 1 year, which in real terms would be 2 years' difference between the declared age and the time lapse between two population registers/censuses. However, such high values are not observed in the rest of the municipalities examined. The highest average values in the latter municipalities are only close to 1 year, although they do not reach it, in the case of Sant Feliu de Llobregat for those born between 1810 and 1829 and in Santa Coloma de Cervelló for the generation born between 1860 and 1869. Likewise, a certain rise in error is observed in some cohorts in the second half of the nineteenth century as in Sant Vicenç dels Horts or Collbató which coincides with the fall in literacy levels that Núñez (1992) found in 1900 for the province of Barcelona. This is not incompatible with the fact that Barcelona was, together with those of the Basque Country, the province with the greatest economic dynamism. In recent papers, Beltrán-Tapia et al. (2021) have proved out that "Still, literacy seems not to be closely related to socioeconomic progress" (p. 21) and Reher (2022) adds "From an economic standpoint, unquestionably the two more dynamic areas of Spain at the time were in the Basque Country and in Catalonia, one located in high literacy area and the other characterized by literacy levels that were moderate at best" (p.24). The statistical association between the two measures, numeracy and age inconsistency, will be the subject of Sect. 6.

Finally, we have compared the level of numeracy in Baix Llobregat (called Barcelona province) with the data published for Spain as a whole and other countries (Fig. 8). Our results show that, with the exception of those born during the crisis of the 1880s and 1890s, numeracy levels in Baix Llobregat were above the Spanish average throughout the period. Internationally, during the first years of the nineteenth century, the arithmetic skills of our region of study were ahead of countries such as Italy, Portugal and the USA. For example, the 1820 cohorts in the Catalan

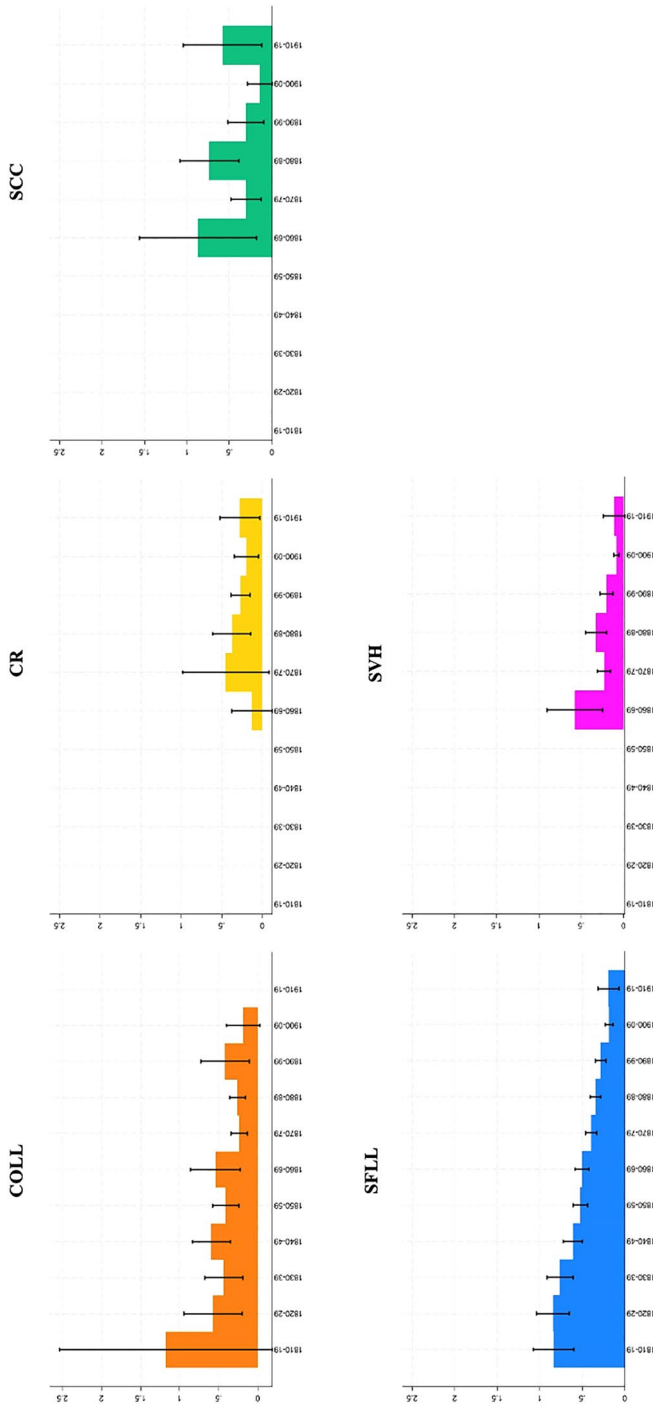
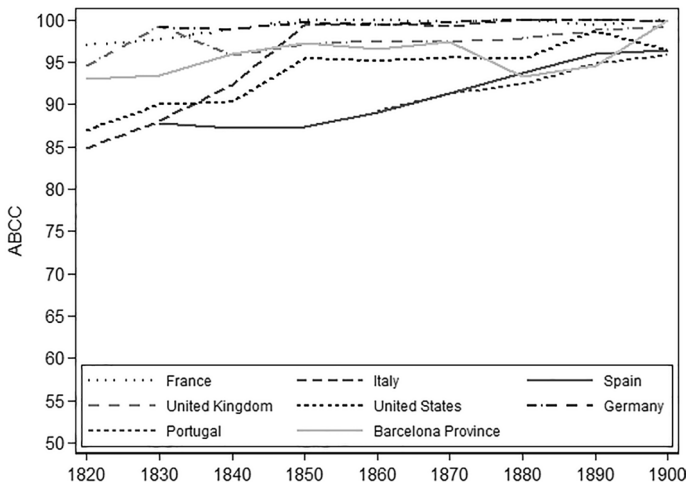


Fig. 7 Mean of individual error by declared year of birth (for cohorts with 10 or more individuals). Source: BALL Database



**Fig. 8** Numeracy in different countries (1810–1900). Source: “Barcelona Province” (Baix Llobregat municipalities) BALL database, “Spain” Beltrán Tapia et al. 2022 and the data for other countries has been obtained from Clio-infra (<https://clio-infra.eu>)

municipalities had a numeracy of 93 while in Italy and the USA, it was 85 and 87, respectively. The Catalan cohort of 1860 was still ahead of countries like Portugal and the USA. Although these comparisons may be somewhat asymmetrical, they confirm the advantage in terms of numeracy of the municipalities analyzed.

## 5 Life course age inconsistencies in Baix Llobregat in the nineteenth and twentieth centuries

To complete the previous descriptive analysis, this section describes the multivariable analyses carried out to show how sociodemographic characteristics and the educational and time contexts determine consistency in individual age declarations over time. An individual perspective would help capture individual’s life-long improvements in terms of numeracy, as such improvements do not necessarily occur at the same time for all individuals in the same cohort. This makes it possible to calibrate the fact that experience in the labor market can be a learning factor in terms of human capital (Camps 2002) or facilitate late literacy, whether through factory programs, military service or in other ways (Puell de la Villa 2001). In addition, the spread of literacy among young people might also have driven a rise in literacy among the adult population (Kuziemko 2014). This improvement can also be identified in the individual tracking conducted in this research.

To do this, each of the five municipalities studied were modeled independently due to their different socioeconomic characteristics, as described earlier, and their population size (Table 4). Note that the numerical variables are scaled (subtracting the mean and dividing by the standard deviation) to prevent problems of convergence, as such variables destabilize the parameter estimation process used by the



**Table 4** Random effect Poisson regression models for the error in consecutive age declarations among the inhabitants of Baix Llobregat

	ALL	COLL	CR	SCC	SFL	SVH
n	19,223	1,889	638	726	13,585	2,385
Id	9,444	665	240	726	8,118	1,695
Id intra var	2946	2,266	1,661	49,385	2,428	7,578
Intercept	-2.464	*** -1.735	*** -2.565	*** -7.696	*** -2.467	*** -2.742
	-0.103	0.259	0.478	1.310	0.118	0.595
Time	0.027	0.233	0.046		-0.022	*** -2.515
Scale (year)	-0.037	0.121	0.237		0.041	0.596
Educational reform (1901)	-0.449	*** -0.883	***		-0.261	**
(ref. No)	-0.074	0.146			0.092	
Age	0.669	*** 0.135	0.543	*	0.723	*** 0.838
(ref. 23—32)	-0.049	0.144	0.268	0.942	0.055	0.262
43—52	0.963	*** 0.251	0.280	-0.102	1.057	*** 1.389
	-0.051	0.146	0.295	1.020	0.057	0.279
53—62	1.200	*** 0.290	0.660	*	1.314	*** 1.647
	0.056	0.165	0.311	0.983	0.062	0.302
63—72	1.368	*** 0.613	*** 0.873	*	1.468	*** 1.513
	0.065	0.189	0.343	1.040	0.073	0.351
Literate	-0.226	-0.029	0.313	-0.331	-0.229	*** -0.754
(ref. No)	0.046	0.144	0.367	0.783	0.052	0.200
Social class (HISCLASS)	-0.946				-0.993	
(ref. Farmers)	0.680				0.668	
High prof (2)	-0.397	* 1.720	* 0.111	1.451	0.601	*** 0.078
	0.158	0.753	0.822	2.196	0.169	0.675
Lower and prof manager (3, 4, 6)	-0.131	-0.612	* -16.652	0.290	-0.134	*** 0.356
	0.081	0.311	3450.849	1.239	0.087	0.432
Lower clerical sales (5)	-0.294	** 0.247	-16.098	0.302	-0.351	-0.517
	0.098	0.331	2986.746	1.366	0.105	0.715

**Table 4** (continued)

	ALL	COLL	CR	SCC	SFLL	SVH
Skilled workers (7)	0.029	-0.476	-0.553	0.272	0.000	0.345
Low-skilled workers (9, 11)	0.058	0.285	1.156	0.965	0.063	0.252
	0.024	-0.309	-0.941	-0.114	0.015	0.218
Low-skilled and unskilled farm workers (10, 12)	0.058	0.403	1.232	0.801	0.063	0.280
	0.014	-0.391	-15.917	0.417	-0.023	-0.027
Sex	0.049	0.305	3172.160	1.858	0.053	0.207
(ref. Male)	0.222	0.098	0.260	0.492	0.210	0.196
Migration status	0.053	0.161	0.247	0.632	0.060	0.208
(ref. Town)	0.028	0.658	0.162	0.174	-0.075	0.516
Municipality of residence	0.040	0.120	0.256	0.810	0.044	0.216
(ref. SFLL)	0.020					
(Sant Feliu de Llobregat)	0.101					
	0.171					
SCC (Santa Coloma de Cervelló)	0.164					
	-0.598					
SVH (Sant Vicenç dels Horts)	0.127					
	-0.803					
	0.088					

Source: BALL Database. (All) All cases, (COLL) Collbató, (CR) Castellví de Rosanes, (SCC) Santa Coloma de Cervelló, (SFLL) Sant Feliu de Llobregat and (SVH) Sant Vicenç dels Horts. N: total cases, id: total individuals, id intra var: intra individual variance. 1st row: estimate, 2nd row: standard deviation of estimate, variation when adding one unit to the variable. Robust *p*-Values are: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.5$ ,  $p < 0.1$

glmer function when they present a disparity of order (age in tens and year in thousands). It should be stressed that this transformation does not affect the results.

In column 1, the results for all municipalities are shown, and in the following columns, the specific results for each of the municipalities are displayed since the Santa Coloma de Cervelló and Sant Vicenç dels Horts have shown statistically significant differences. Specifically, in the case of Sant Vicenç dels Horts, consistency in age declarations increases over time. Moreover, consistency increases in the case of age declarations after 1905, as a proxy for the 1901 educational reform in which primary education was made compulsory until the age of 12, when all the municipalities are taken into account and in the cases of Collbató (drop in age declaration incoherence of 59%) and Sant Feliu de Llobregat (of 23%). Being literate is positively related to better consistency in the age statements of the municipalities of Sant Feliu de Llobregat (20% of decrease) and Sant Vicenç dels Horts (53%). It is also reasonable to suppose that illiteracy dropped after the 1901 reform. In 1900, the male and female illiteracy rates in the total Spanish population were 42.9% and 66.7%, respectively, while in 1910, they were 37.1% and 57.7% (Beltrán-Tapia et al. 2019). We calculated the illiteracy rates of the population aged 10 or more for the province of Barcelona based on data from the 1900 and 1910 censuses. In the first year, the male illiteracy rate was 37%, while in the second, it was 25%. Women's illiteracy also dropped from 56 to 42%. This confirms that the interaction between both variables proving significant in the case of knowing how to write and age declaration occurred after the reform in the municipality of Sant Feliu de Llobregat (model not shown).

With regard to the sociodemographic variables, it is worth stressing how age increases inconsistency, the higher the age of the declarant, higher the inconsistency both in the joint model and in the independent models except for the case of Santa Coloma de Cervelló and for some age groups in the Collbató and Castellví de Rosanes municipalities. Indeed, today in developing countries, older individuals also show a greater tendency to age heap, especially toward ages ending in 0 (Del Popolo 2000). In the realm of social class, the observed inconsistency exhibited a discernible social gradient—the higher the socioeconomic class, the greater the level of consistency observed. The models presented in this paper do not include the variable knowing how to read, as it had no significance in any of the municipalities. Concretely, in the case of the joint model, high professionals exhibit 67% more consistency in their age declarations than farmers, a trend that is also confirmed in the municipality of Sant Feliu de Llobregat. A similar pattern is observed for lower managers, lower professionals, clerical sales, and foremen in both Collbató and Sant Feliu de Llobregat. The same tendency is observed for lower clerical sales in the case of the complete model. This socioeconomic gradient is also found in other studies carried out by calculating the level of numeracy. Tollnek and Baten (2017) showed that for countries like Italy, Germany, Austria, Spain and Uruguay between the eighteenth and nineteenth centuries, professional and skilled workers general had better levels of numeracy than semi-skilled or unskilled workers. In Spain in the eighteenth century, almost 90% of the professional workers analyzed correctly declared their ages, the figure for intermediate and skilled workers and farmers was 80%, around 70% of ages were correctly declared among semi-skilled and unskilled workers. Gómez-i-Aznar (2019) obtained similar results for

Catalonia in around 1720. Álvarez and Ramos Palencia (2018) also confirm that this trend in a number of Castile municipalities around 1750: 72% of unskilled workers, 77% of low and medium skilled workers and 82% of highly skilled workers declared their age correctly. The authors also offer data on literacy through the signing of documents, which mostly match the data for numeracy. In this context, 24% of unskilled workers, 52% of low-skilled workers, 62% of medium skilled and 99% of high skilled workers signed documents.

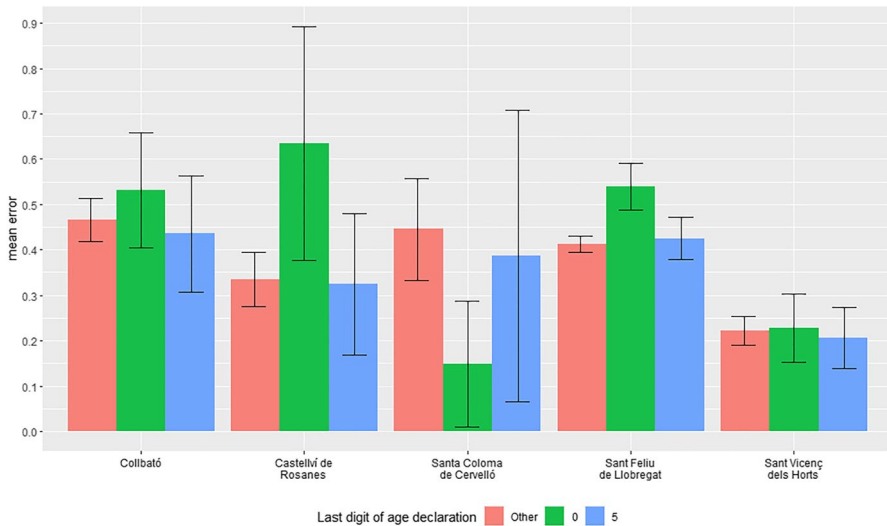
We can also highlight that women tended to state their age more inconsistently than men. In both the joint model and in Sant Feliu de Llobregat, men demonstrated 25% more consistency in their age declarations than women. Two circumstances may explain this pattern. The first is the fact that literacy was lower among women than among men (Núñez 1992; Sarasúa 2002a, b; Beltrán Tapia et al. 2019). The second is related to the fact that the declaration of sociodemographic information of members of the same household was normally provided by the head of the family, which might mean that the declarations for other family members are not as accurate as their own. Unfortunately, individual declaration sheets have not always been kept to check who provided the information on the members of each household. For the municipalities available in the BALL Database, the population register sheets from individual households have been kept for the 1924 and 1936 Santa Coloma de Cervelló population registers. They were signed by 114 men and 12 women heads of family (6 married and 6 widowed) and 225 men and 23 women (5 married, 16 widows and 2 single), respectively. In the case of Castellví de Rosanes in the 1936 population register, the sheets were signed by 37 men and 2 women heads of family, one of whom was married and the other widowed. For the Collbató population register for the same year, the signatories were 95 male heads of family and no women. The variable marital status is also considered, although with no significance in any of the municipalities, hence, it is not included in the models presented here.

Regarding migratory status, immigrants exhibited more accurate age declarations than the local population only in the context of Sant Vicenç dels Horts. However, statistical significance for this variable was not observed in the other municipalities or in the joint model. According to data from Beltrán Tapia et al. (2022), this could be explained by the fact that generally, Barcelona was the Catalan province with the highest numeracy level up to the 1930 census. In 1877, Barcelona was the province with the highest numeracy level, followed by Girona and Tarragona, and then Lleida, much further behind. In 1900, Girona overtook Barcelona and finally, in the 1930 census, all four Catalan provinces had the same numeracy level of 100.<sup>6</sup>

## 6 Relation between the ABCC Index and inconsistency in individual age declarations

Having presented the results using both methods, age heaping and consistency of individual age declarations in different census, in this section, we study whether there is a relation between both proxies. Figure 9 shows the mean error

<sup>6</sup> The exact ABCC value for the Catalan provinces is not given, as it is not provided by Beltrán Tapia et al. (2022).



**Fig. 9** Mean error by declared age (ending in 0, 5 and others). Source: BALL Database

for individuals declaring an age ending in 0, 5 and other digits grouped together and the 95% confidence interval. In the case of Collbató, Castellví de Rosanes, Sant Feliu de Llobregat and Sant Vicenç dels Horts, the error is greater in those declaring ages ending in 0. This corroborates Fig. 3, which shows greater age heaping for ages that are multiples of 10. By contrast, in Santa Coloma de Cervelló, the mean error is greater among those declaring ages that do not end in 0 or 5. These initial findings show that in four of the five municipalities, there is a relation between age heaping and the mean error in individual declarations in the different population registers. For this reason, we next modeled this association with the aim of establishing whether it is statistically significant, controlling for the sociodemographic, economic and educational variables used in the previous section.

On an aggregate basis and considering all municipalities, Table 4 shows a positive and significant association between declaring an age ending in 0 or 5 and a greater mean error in individual declarations controlled for time, educational reform, age, literacy and social group. When analyzing the municipalities individually, the correlation continues to be significant and positive in the municipalities of Castellví de Rosanes and Sant Feliu de Llobregat. The association is not statistically significant for the other municipalities. It is important to highlight that by including the no numerate variable in the regression, there are no important changes in the statistical significance of the other dependent variables in relation to what was seen in the previous model (Table 5). This shows that the results are not distorted, even when controlling for traditional age heaping. In this sense, the 1901 educational reform continues to have an effect in terms of the consistency of age declarations as well as age, sex, social group and migratory status -although the latter only for certain municipalities.

As stated in the introduction, the only study to have tested this correlation is that of Blum and Krauss (2018). The authors found a positive correlation between correctly

**Table 5** Random effect Poisson regression models for the error in consecutive age declarations among the inhabitants of Baix Llobregat (adding the non-numerate variable)

	ALL	COLL	CR	SCC	SFL	SVH
n	19,223	1,889	638	726	13,585	2,385
Id	9,444	665	240	726	8,118	1,695
Id intra var	2,952	2,316	1,691	48.83	2,418	8.171
Intercept	-2.479	***	***	***	***	***
	0.103	-1.737	-2.723	-7.515	-2.478	-2.692
End in 0 or 5	0.090	*	0.479	1.329	0.118	0.598
	0.036	0.010	0.457	-0.560	0.102	-0.165
Scale (year)	0.02699	0.101	0.188	0.854	0.051	0.195
	0.03740	0.233	0.053		-0.021	-2.524
Yes	-0.45378	0.121	0.239		0.041	0.597
Educational reform (1901)	0.07379	***	***		-0.271	**
(Ref. No)		0.146			0.092	
Age	0.66921	0.135	0.557	0.038	0.722	0.837
(Ref. 23—32)	0.04888	0.144	0.268	0.941	0.055	0.262
43—52	0.96072	***	0.264	-0.133	1.054	1.386
	0.05090	0.146	0.296	1.016	0.057	0.279
53—62	1.199	***	0.684	*	1.312	1.635
	0.05612	0.165	0.312	0.982	0.062	0.302
63—72	1.366	***	0.856	*	1.466	1.512
	0.065	0.189	0.343	1.038	0.073	0.351
Yes	-0.226	**	0.348	-0.356	-0.230	-0.759
Literate	0.046	0.144	0.365	0.786	0.051	0.200
(Ref. No)	-0.943				-0.991	
Social class (HISCLASS)	0.680				0.668	0.080
(ref. Farmers)	-0.400	*	-0.109	1.510	-0.602	0.674
High prof (2)	0.158	0.753	0.830	2.210	0.169	0.354

Table 5 (continued)

	ALL	COLL	CR	SCC	SFLL	SVH
Lower and prof manager (3, 4, 6)	-0.132	-0.614 *	-16.359 *	0.365	-0.133	0.431
Lower clerical sales (5)	0.081	0.311	3.069,286	1.241	0.087	-0.528
	-0.291 **	0.245	-16.158	0.322	-0.347 ***	0.716
Skilled workers (7)	0.098	0.332	3.069,286	1.371	0.105	0.341
	0.030	-0.475	-0.574	0.265	-0.001	0.252
Low-skilled workers (9, 11)	0.058	0.285	1.161	0.964	0.063	0.213
	0.023	-0.308	-0.970	-0.107	0.015	0.280
Low-skilled and unskilled farm workers (10, 12)	0.058	0.403	0.240	0.799	0.063	-0.023
	0.014	-0.391	-15.953	0.385	-0.023	0.207
Sex	0.049	0.305	3.001,024	1.856	0.053	
(Ref. Male)	0.222 ***	0.099	0.238	0.460	0.209 ***	0.193
Migration status	0.053	0.161	0.248	0.633	0.060	0.208
(Ref. Town)	0.029	0.657 ***	0.195	0.170	-0.074	0.520 *
Municipality of residence	0.040	0.120	0.257	0.806	0.044	0.216
(Ref. SFLL)	0.020					
(Sant Feliu de Llobregat)	0.101					
	0.171					
SCC (Santa Coloma de Cervelló)	0.164					
	-0.594 ***					
SVH (Sant Vicenç dels Horts)	0.127					
	-0.799 ***					
	0.088					

Source: BALL Database. (All) All cases, (COLL) Collbató, (CR) Castellví de Rosanes, (SCC) Santa Coloma de Cervelló, (SFLL) Sant Feliu de Llobregat and (SVH) Sant Vicenç dels Horts. N: total cases, id: total individuals, id intra var: intra individual variance. 1st row: estimate, 2nd row: standard deviation of estimate, variation when adding one unit to the variable. Robust  $p$  Values are: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.5$ ,  $\cdot p < 0.1$

declaring age in two registers from two different years and the traditional method of measuring numeracy or the attraction to numbers ending in 0 and 5. However, as stated in their work, the sample of 162 individuals might not be fully representative. In our case, we analyzed the whole population aged 23–72 in the municipalities and often crossed two or more individual declarations in different population registers. Although studies have been published showing that numeracy correlates with other proxies for human capital (Tollnek and Baten 2016; Baten et al. 2022), the analysis presented here provides further evidence, although it is still scarce, despite the hundreds of studies published on numeracy levels from a historical perspective.

## 7 Conclusions

This article provides new data on human capital from a sample of municipalities in the current region of Baix Llobregat in the province of Barcelona. This region borders the city of Barcelona and, following in its slipstream, underwent relatively early industrialization in the last third of the nineteenth century preceded by major proto-industrialization. First of all, the basic maths abilities known in the literature as numeracy are estimated by measuring age heaping as a proxy (ABCC Index). This reveals that numeracy levels in the municipalities were high in the second half of the nineteenth century and then dropped in the last two decades. This fact, which matches the literature on numeracy and literacy levels in Catalonia for the period, might be due to child labor in the factories in the zones that underwent the Industrial Revolution. Families would consider that many of their children obtained an effective qualification by working in the factory, which made it helpful to start at a relatively young age. Therefore, there could be the perception at the time that literacy was not essential for everyday life, which hindered schooling (Borrás Llop 2002; Iturralde Valls 2017; Reher 2022). With the start of the new century, especially from 1910 onwards, these levels started to recover. If we look at the average for the years 1900–1915, we see that numeracy levels for these municipalities were 94.6% of correct age declarations, while the Spanish average was 89.9% in 1900, 90.6% in 1910 and 91.2% in 1920 (Beltrán Tapia et al. 2022). As with literacy, this increase in numeracy levels could be partly due to the 1901 reform, which raised the obligatory school leaving age from 9 to 12.

As a novel method to assess numeracy, we developed another proxy to estimate individuals' arithmetic skills by calculating the error in individual age declarations among people identified in two or more consecutive population registers. This was facilitated by the richness of the data in the BALL Database, making it possible to track individuals' declarations at up to eight different moments. Our results indicate that the inconsistency of age declarations increases with age. The observed error in individual declarations in different population registers largely coincides with the results obtained from the ABCC Index. Generally, when the ABCC index is low, there is a greater error in individual declarations. Continuing with the individual-level results, it is worth noting how women show more errors in their individual declarations than men. This is particularly well demonstrated



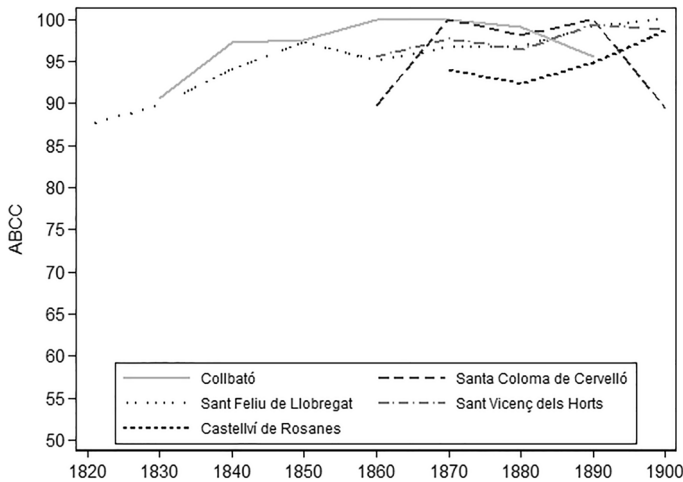
in the municipality of Sant Feliu de Llobregat. With regard to origin, also at an individual level, natives show lower inconsistency in ages only when municipalities are analyzed separately. This same pattern had been observed at an aggregated level with the levels of ABCC. Furthermore, knowing how to write or having an occupation with a higher social status also reduces inconsistency in age declarations in population registers over time, as has been shown in other studies analyzing age heaping. We further find a positive and significant correlation, also on individual basis, between declaring an age ending in 0 or 5 and having a higher error in consecutive age declarations in censuses when considering all the municipalities. This association is maintained for the municipalities of Castellví de Rosanes and Sant Feliu de Llobregat. This result matches the only two studies we know of to date that uses this methodological approach suggesting that the ABCC index is an acceptable indicator for measuring people's numeracy (Blum and Krauss 2018; Baten and Nalle 2022).

This research aims to help fill the historiographical gap regarding the degree of consistency in age declarations by a single individual over their lifetime. Over and above having new databases with indicators of human capital, this new methodology, which registers people with their full name, age and other variables of interest, opens up potential new lines of research. These include: analysis of the accuracy of the traditional age heaping method, the relationship between social mobility and human capital, and the improvement (or otherwise) of arithmetical skills in a single person in relation to socioeconomic context in other Spanish regions or countries that permit comparison. For instance, in the case of Spain, there is still no consensus on the causes of the major regional differences in literacy in the nineteenth century. Likewise, the relationship between human capital and economic evolution is still unclear, given that the territories with the highest levels of literacy in the nineteenth century were not the most economically developed, as is the case of Old Castile (Reher 2022).

Another thought-provoking approach would be to observe whether children of people with fewer errors in individual declarations also show greater consistency in their age declaration in adulthood. This analysis would help corroborate the transmission of human capital in families and whether this is true for all descendants or whether there is a difference based on order of birth or sex. Social mobility could also be analyzed more deeply by observing whether the children of parents showing less inconsistency in age declarations have better jobs than their parents. All this will be answered in future research.

## Appendix

See Fig. 10.



**Fig. 10** ABCC Index by birth decade in Baix Llobregat municipalities (1820–1900). Source: BALL Database

**Acknowledgements** The authors would like to thank the financial support from the Spanish Ministry of Science and Innovation: Project PID2021 -1280100B-I00. Epidemias, estado y desigualdades socio-económicas: predictibilidad y perdurabilidad, siglos XIX y XX (EPI-DESIGUAL). Furthermore, the authors appreciate the comments received at the Research Workshop “Education and development in History: new data and methods” organized in Valencia.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- A’Hearn B, Baten J, Crayen D (2009) Quantifying quantitative literacy: age heaping and the history of human capital. *J Econ Hist* 69(3):783–808. <https://doi.org/10.1017/S0022050709001120>
- Alba Molina R, Aso Pérez C (2008) El Canal de la Infanta: vector de desarrollo económico, social y urbano de la comarca del Bajo Llobregat (Provincia de Barcelona), 1a Parte. *Cimbra* 381:26–34
- Álvarez B, Ramos Palencia F (2018) Human capital and earnings in eighteenth-century Castile. *Explor Econ Hist* 67:105–133. <https://doi.org/10.1016/j.eeh.2017.10.005>
- Barro RJ (1991) Economic growth in a cross section of countries. *Q J Econ* 106(2):407–443
- Barro RJ, Lee JW (1996) International measures of schooling years and schooling quality. *Am Econ Rev* 86(2):218–223
- Baten J, Nalle S (2022) Age heaping and life course reckoning. individual-level evidence from spanish inquisition records. *Vierteljahrsschrift Für Sozial- Und Wirtschaftsgeschichte* 109:337–350. <https://doi.org/10.25162/vswg-2022-0010>
- Baten J, Van Zanden JL (2008) Book production and the onset of modern economic growth. *J Econ Growth* 13:217–235

- Baten J, Crayen D, Voth HJ (2014) Numeracy and the impact of high food prices in industrializing Britain, 1780–1850. *Rev Econ Stat* 96(3):418–430
- Baten J, Benati G, Ferber S (2022) Rethinking age heaping again for understanding its possibilities and limitations. *Econ Hist Rev*. <https://doi.org/10.1111/ehr.13139>
- Becker G (1962) Investment in human capital: a theoretical analysis. *J Political Econ* 70:9–49
- Becker G (1964) [1994] Human capital revisited. In *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. The University of Chicago Press (Third edition)
- Beltrán Tapia FJ, de Miguel Salanova S (2017) Migrants' self-selection in the early stages of modern economic growth, Spain (1880–1930). *Econ Hist Rev* 70(1):101–121. <https://doi.org/10.1111/ehr.12332>
- Beltrán Tapia FJ, Díez-Minguella A, Martínez-Galarraga J, Tirado-Fabregat DA (2021) The uneven transition towards universal literacy in Spain, 1860–1930. *Hist Educ* 50(5):605–627. <https://doi.org/10.1080/0046760X.2021.1890237>
- Beltrán Tapia FJ, Díez-Minguella A, Martínez-Galarraga J, Tirado-Fabregat DA (2022) Two stories, one fate: age-heaping and literacy in Spain, 1877–1930. *Rev De Hist Econ-J Iber Latin Am Econ Hist* 40(3):405–438. <https://doi.org/10.1017/S0212610921000033>
- Beltrán Tapia FJ, Díez-Minguella A, Martínez-Galarraga J, Tirado-Fabregat DA (2019) Capital humano y desigualdad territorial. El proceso de alfabetización en los municipios españoles desde la Ley Moyano hasta la Guerra Civil (Vol. 74). *Estudios de Historia Económica*, Banco de España
- Blum M, Krauss KP (2018) Age heaping and numeracy: looking behind the curtain. *Econ Hist Rev* 71(2):464–479. <https://doi.org/10.1111/ehr.12615>
- Borrás Llop JM (2002) Mercado laboral, escolarización y empleo infantil en una comarca agrícola e industrial (el Vallés Occidental, 1881–1910). *Cuadernos De Historia Contemporánea* 24:233–262
- Brea-Martínez G, Pujadas-Mora JM (2017) Transformación y desigualdad económica en la industrialización en el área de Barcelona, 1715–1860. *Rev Hist Econ/J Iber Latin Am Econ Hist* 36(2):241–273. <https://doi.org/10.1017/S0212610917000234>
- Calvo Calvo A (1995) Les transformacions del sistema de transport al Baix Llobregat (segles XVIII–XIX). In: Calvo Calvo (cur.) *El pas de la societat agrària a industrial al Baix Llobregat*. Agricultura intensiva i industrialització. Publicacions de l'Abadia de Montserrat, Barcelona, pp 653–678
- Camps E (1992) Population turnover and the family cycle: The migration flows in a Catalan town during the nineteenth century. *Contin Chang* 7(2):225–245. <https://doi.org/10.1017/S0268416000001582>
- Camps E (2002) Trabajo infantil y estrategias familiares durante los primeros estadios de la industrialización catalana (1850–1925). Esbozos a partir del estudio de un caso. *Cuadernos De Historia Contemporánea* 24:263–280
- Camps E (1995) La formación del mercado de trabajo industrial en la Cataluña del siglo XIX. *Colección historia social* 33, Madrid
- Carbonell i Porro J (1995) Desenvolupament de la indústria textil a Sant Feliu de Llobregat (1861–1923). In: Calvo Calvo (cur.) *El pas de la societat agrària a industrial al Baix Llobregat*. Agricultura intensiva i industrialització. Publicacions de l'Abadia de Montserrat, Barcelona, pp 405–425
- Carreras, A (1990) Cataluña, primera región industrial de España. In: Nadal y Carreras (dirs.) *Pautas regionales de la industrialización española (siglos XIX y XX)*. Ariel, Barcelona, pp 259–295
- Codina i Vilà, J (1995) El Baix Llobregat, 1750–1850: Un segle transcendent. In: Calvo Calvo (cur.) *El pas de la societat agrària a industrial al Baix Llobregat*. Agricultura intensiva i industrialització. Publicacions de l'Abadia de Montserrat, Barcelona, pp 109–148.
- Colomé Ferrer J, Valls-Junyent F (2012) Las consecuencias demográficas de la crisis filoxérica en la región vitícola del Penedès (Cataluña). *Historia Agraria* 57:47–77
- Crayen D, Baten J (2010) Global trends in numeracy 1820–1949 and its implications for long-term growth. *Explor Econ Hist* 47(1):82–99. <https://doi.org/10.1016/j.eeh.2009.05.004>
- De Gabriel N (1997) Alfabetización y escolarización en España (1887–1950). *Revista De Educación* 34:217–243
- De la Pascua Sánchez MJ (1989) Aproximación a los niveles de alfabetización en la provincia de Cádiz: las poblaciones de Cádiz, El Puerto de Santa María, Medina Sidonia y Alcalá de los Gazules entre 1675–1800. *Trocadero, Revista De Historia Moderna y Contemporánea* 1:51–65
- Gaceta de Madrid, 30/10/1901, N° 303. <https://www.boe.es/diario/gazeta/hemeroteca.php?a=1901&m=10&d=30>
- García Pérez MS (2007) El padrón municipal de habitantes: origen, evolución y significado. *HISPANIA NOVA. Revista de Historia Contemporánea*. Número 7. <http://hispanianova.rediris.es>

- García Ruipérez M (2012) El empadronamiento municipal en España: evolución legislativa y tipología documental. *Documenta & Instrumenta - Documenta et Instrumenta* 10:45–86. [https://doi.org/10.5209/rev\\_DOCU.2012.v10.40485](https://doi.org/10.5209/rev_DOCU.2012.v10.40485)
- Gómez-i-Aznar È (2019). Human capital at the beginnings of the 18th century Catalonia: age-heaping and numeracy in a changing economy (No. 1904; Documentos de Trabajo Asociación Española de Historia Económica). <https://www.aehe.es/wp-content/uploads/2019/02/dt-aehe-1904.pdf>
- Iturralde Valls M (2017) El trabajo infantil en las fábricas de indianas: Barcelona, 1736–1800. *Rev Hist Ind* 26(68):13–46
- Juif D, Quiroga G (2019) Do you have to be tall and educated to be a migrant? Evidence from Spanish recruitment records, 1890–1950. *Econ Hum Biol* 34:115–124. <https://doi.org/10.1016/j.ehb.2018.12.006>
- Juif D, Baten J, Pérez-Artés MC (2020) Numeracy of religious minorities in Spain and Portugal during the inquisition era. *Rev Hist Econ/j Iber Latin Am Econ Hist* 38(1):147–184. <https://doi.org/10.1017/S021261091900034X>
- Kuziemko I (2014) Human capital spillovers in families: Do parents learn from or lean on their children? *J Law Econ* 32(4):755–786. <https://doi.org/10.1086/677231>
- Martínez-Galarra J, Prat M (2016) Wages, prices, and technology. *Econ Hist Rev* 69:548–574. <https://doi.org/10.1111/ehr.12127>
- Mincer J (1974) Schooling, experience, and earnings. National Bureau of Economic Research, New York
- Nadal J (1992) Moler, tejer y fundir. Estudios de historia industrial. Ariel Historia, Barcelona
- Núñez CE (1992) La fuente de la riqueza: educación y desarrollo económico en la España contemporánea. Alianza
- Pérez-Artés MC (2023) Numeracy selectivity of Spanish migrants in colonial America (sixteenth–eighteenth centuries). *Econ Hist Rev*. <https://doi.org/10.1111/ehr.13279>
- Pérez-Artés MC, Baten J (2021) Land inequality and numeracy in Spain during the seventeenth and eighteenth century. *Hist Agrar Rev Agric Hist Rural* 83:7–39. <https://doi.org/10.26882/histagr.083e08p>
- Del Popolo F (2000). Los problemas en la declaración de la edad de la población adulta mayor en los censos. CEPAL.
- Puell de la Villa F (2001) Educación de adultos en el servicio militar español. *Historia De La Educación* 20:307–333
- Pujadas-Mora JM, Brea-Martínez G, Jordà Sánchez JP, Cabré A (2018) The apple never falls far from the tree: siblings and intergenerational transmission among farmers and artisans in the Barcelona area in the sixteenth and seventeenth centuries. *Hist Fam* 23(4):533–567. <https://doi.org/10.1080/1081602X.2018.1426483>
- Pujadas-Mora JM, Fornés A, Ramos Terrades O, Lladós J, Chen J, Valls-Fígols M, Cabré A (2022) The barcelona historical marriage database and the baix llobregat demographic database. from algorithms for handwriting recognition to individual-level demographic and socioeconomic data. *Hist Life Course Stud* 12:99–132
- Recaño Valverde AJ. (1995) Les primeres fases de la transició demogràfica al Baix Llobregat i a l'Hospitalet (1787–1936). In: Calvo Calvo (cur.) El pas de la societat agrària a industrial al Baix Llobregat. Agricultura intensiva i industrialització. Publicacions de l'Abadia de Montserrat, Barcelona, pp 46–105.
- Recaño Valverde AJ (1996) Las migraciones interiores en el área mediterránea, 1860–1960. *Papers de demografia* N°120
- Reher D (2022) Patterns of literacy in historic Spain: an interpretation. *Rev Hist Econ/j Iber Latin Am Econ Hist*. <https://doi.org/10.1017/S0212610922000088>
- Reher D-S, Valero Lobo A (1995) Fuentes de información demográfica en España. Centro de Investigaciones Sociológicas
- Retuerta Jiménez ML (2012) Les inversions barcelonines en els recursos hidràulics a Sant Feliu de Llobregat, Sant Just Desvern i Santa Creu d'Olorda fins a l'arribada del canal de la Infanta. *Estudis D'història Agrària* 24:293–309
- Rocamonde Lourido J, Sabatè J (2019) Paisajes Cooperativos. Una comparativa de dos paisajes industriales vinculados al río Llobregat. *Ciudades* 22:151–172. <https://doi.org/10.24197/ciudades.22.2019.151-172>
- Rodríguez MC, Bennassar B (1978) Signatures et niveau culturel des témoins et accusés dans les procès d'inquisition du ressort du Tribunal de Tolède (1525–1817) et du ressort du Tribunal de Cordoue (1595–1632). *Cahiers Du Monde Hispanique Et Luso-Brésilien* 31:17–46

- Romer PM (1989) Human capital and growth: theory and evidence (No. w3173). National Bureau of Economic Research
- Salas-Vives P, Pujadas-Mora JM (2021) Bottom-up nation-building: national censuses and local administration in nineteenth-century Spain. *J Hist Sociol* 34(2):287–304. <https://doi.org/10.1111/johs.12323>
- Sarasúa C (2002a) Aprendiendo a ser mujeres: las escuelas de niñas en la España del siglo XIX. *Cuadernos De Historia Contemporánea* 24:281–297
- Sarasúa C (2002b) El acceso de niñas y niños a los recursos educativos en la España rural del siglo XIX. In: Martínez Carrión JM (ed) *El nivel de vida en la España rural, siglos XVIII-XX*. Universidad de Alicante, Alicante, pp 549–609
- Schultz TW (1961) Investment in human capital. *Am Econ Rev* 51(1):1–17
- Thorvaldsen G (2018) Censuses and census takers. A global history. Routledge
- Tiana Ferrer A (1987) Educación obligatoria, asistencia escolar y trabajo infantil en España en el primer tercio del siglo XX. *Historia De La Educación: Revista Interuniversitaria* 6:43–60
- Tollnek F, Baten J (2016) Age-heaping-based human capital estimates. In: Diebolt C, Hauptert M (eds) *Handbook of cliometrics*. Springer, pp 131–154
- Tollnek F, Baten J (2017) Farmers at the heart of the ‘human capital revolution’? Decomposing the numeracy increase in early modern Europe. *Econ Hist Rev* 70(3):779–809. <https://doi.org/10.1111/ehr.12382>
- Van Leeuwen MHD, Maas I (2011) HISCLASS: a historical international social class scheme. Leuven University Press
- Viñao Frago A (1999) Alfabetización y primeras letras (siglos XVI-XVII). In: Castillo A (ed) *Escribir y leer en el siglo de Cervantes*. Gedisa Editorial, Barcelona, pp 39–84
- Viñao Frago A (2004) *Escuela para todos: educación y modernidad en la España del siglo XX*. Marcial Pons
- Vincent B (1987) Lisants et non-lisants des royaumes de Grenade et de Valence a la fin du XVI siècle. In: CNRS (ed) *De l’alphabétisation aux circuits du livre en Espagne, XVI-XIX siècle*. CNRS, París, pp 85–104
- Wu W, Little TD (2011) Quantitative research methods. *Encyclopedia of adolescence*. Elsevier Inc., pp 287–297

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Joana María Pujadas-Mora<sup>1</sup>  · María Carmen Pérez-Artés<sup>2</sup> 

✉ Joana María Pujadas-Mora  
jpujadasm@uoc.edu

María Carmen Pérez-Artés  
mcarmenartes@ual.es

<sup>1</sup> Open University of Catalonia & Center for Demographic Studies, Universitat Autònoma de Barcelona, Barcelona, Spain

<sup>2</sup> Department of Economics and Business, Research Group HEDES (SEJ-667) and Mediterranean Research Center On Economics and Sustainable Development (CIMEDES), University of Almería, Almería, Spain