THE
ECONOMIC HISTORY REVIEW

Economic History Review, 75, 1 (2022), pp. 111–137

Rethinking age heaping: a cautionary tale from nineteenth-century Italy†

By BRIAN A'HEARN, ALEXIA DELFINO,
and ALESSANDRO NUVOLARI*

Age heaping is widely employed as an indicator of human capital, more specifically of numeracy. We re-examine the age heaping–numeracy link in the light of evidence from nineteenth-century Italian censuses, in which education explains little of the variation in age heaping. We argue that in general age heaping is most plausibly interpreted as an indicator of cultural, economic, and institutional modernization rather than a straightforward measure of individual cognitive skills. We do not rule out the use of age heaping as a numeracy indicator, but this needs to be done with research designs that are alert to historical specificities of the context under investigation.

Historical research on human capital has faced serious data constraints. Information on both inputs, such as school enrolments, and outputs, such as literacy, is widely available only from the second half of the nineteenth century, and even then is fraught with gaps, inconsistent reporting, and other shortcomings. A promising new measure of human capital that has gained traction among economic historians is age heaping. Irregularities in an age distribution, in particular the over-representation of round numbers such as multiples of 10, may indicate that some individuals lack sufficient understanding of numbers, or 'numeracy', to know their precise age or birth year. Instead, they report a round or otherwise attractive number that is not implausible and has some salience for them. In a histogram, this results in observations being excessively 'heaped' on particular ages, as in figure 1. If age heaping reliably captures numeracy, it has much to recommend it as a measure of human capital: age data are abundant and easy to work with; quantitative reasoning is of interest *per se*; and basic counting abilities may reveal variation even among what might look like an undifferentiated sea of illiterates. On this basis, age heaping has been adopted with some enthusiasm to reconstruct numeracy trends in a wide range of countries and historical periods.[1]

*Authors' Affiliations: Brian A'Hearn, Pembroke College, University of Oxford; Alexia Delfino, IGIER and LEAP, Bocconi University, Milan; Alessandro Nuvolari, Sant'Anna School of Advanced Studies, Pisa.

† We would like to thank Raffaele Danna and Cormac Ó Gráda for helpful comments and Ariell Zimran for sharing data. The suggestions of three anonymous referees and of the editor of this journal are also gratefully acknowledged. The article has benefited from discussions at seminars at the Catholic University of Milan, the University Paris 1 Pantheon-Sorbonne, and the London School of Economics. This research has been supported by the project 'Lost highway. Skills, technology and trade in Italian economic growth, 1815–2018', funded by the Italian Ministry of University and Research, Progetti di Rilevante Interesse Nazionale (PRIN) Bando 2017, n. 2017YLBYZE.

[1] Tollnek and Baten, 'Age heaping'.

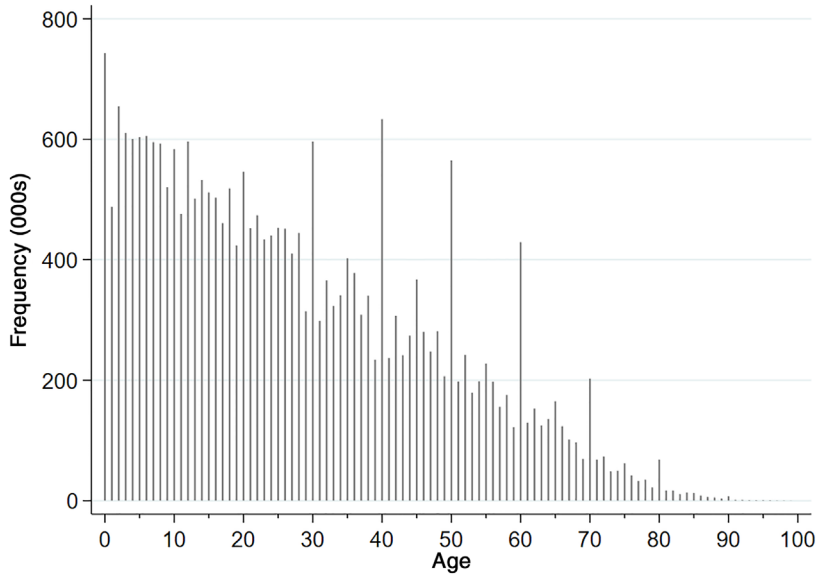

Figure 1. *Declared ages, census of 1871*
[Colour figure can be viewed at wileyonlinelibrary.com]
Note: Frequencies of reported ages in the Italian census of population of 31 Dec. 1871.
Source: MAIC, *Censimento 1871*.

The aim of this article is to reconsider the relationship between age heaping and numeracy in the light of research in other fields of social science and history. While the numeracy interpretation prevails in economic history, in adjacent disciplines, such as political science and social history, factors such as state capacity and culture are regarded as more relevant determinants of age-heaping patterns. We carry out our reappraisal via a case study of nineteenth-century Italy. Italy is a promising test bed for such an inquiry. A pragmatic consideration is that the first Italian censuses reported unusually rich age data, with single-year age distributions by province, gender, marital status, and literacy. But Italy is equally interesting for two further reasons. First, the newly unified kingdom focused considerable attention on the censuses, which were seen as an important component of nation-building; this generated a lively debate about census procedures, results, and their interpretation.[2] Second, Italy was (and remains) a country characterized by major geographic disparities in educational attainment, in economic development, and in institutions and cultural attitudes. This heterogeneity makes Italy a promising case for investigating the determinants of age heaping and their relative importance.

The rest of the article is structured as follows. In the next section we provide a compact survey of the evolution of research on numeracy, age awareness, and age heaping, not only in economic history but also in social history, political science, and cognitive psychology. In section II, we present the Italian census evidence, highlighting a number of anomalies that cannot be reconciled with the simple

[2] Patriarca, *Numbers*.

numeracy interpretation of age heaping now in vogue among economic historians. In section III, we turn to evidence from Italian social history which suggests widespread numeracy, even among illiterates, but limited age awareness, casting further doubt on the indiscriminate use of age heaping as a numeracy indicator. Section IV summarizes our argument and draws conclusions.

I

Age heaping was recognized by nineteenth-century demographers, for whom it complicated tasks like calculating mortality tables. By the early twentieth century, textbooks instructed readers in the use of a standard diagnostic now called the Whipple index.[3] It began to attract the interest of historians from the late 1970s. In an influential study, Fischer used age-heaping trends to study the evolution of age consciousness in the US. In his interpretation, during the nineteenth century an increasing tendency for people to report younger ages at critical junctures—for example, 39 rather than 40—revealed a cultural shift from respect for old age to a growing appeal of youth.[4] Subsequently, Thomas called attention to the social meaning of age in early modern England: 'ages were reported with precision for people under twenty, because differences in age could be of considerable social and administrative importance for young persons. But when people reached adulthood their exact numerical age had much less social meaning; and it was much more vaguely reported'.[5] A similar, cultural interpretation was adopted by Kaiser and Engel who regarded age heaping as a measure of 'time and age-awareness' in eighteenth-century Russia.[6] In the same vein, more recently, Dillon has studied age heaping in the US and Canadian censuses to document variation in age consciousness, emphasizing the cultural and social embeddedness of the very meaning of age.[7]

A slightly different take on age heaping emerges from Herlihy and Klapisch-Zuber's study of tax registers in medieval Tuscany. The authors pointed to a notable improvement in the accuracy of age reporting over the period 1371–1471, which they ascribed to increasing pressure for numerical precision by fiscal authorities.[8] Dealing with the same case, Emigh suggested a more nuanced interpretation, noting that in this period the connection between numerical skills, age awareness, and age reporting was complicated by a number of social and cultural elements such as the expectation that women should marry at a young age.[9]

In perhaps the first use of age heaping by *economic* historians, Mokyr and Ó Gráda suggested that it could serve as a proxy for the degree of 'quantitative

[3] Whipple actually ascribed the introduction of this index, computed as the 'per cent which the number reported as multiples of 5 forms of one-fifth of the total number between ages 23 and 62 years, inclusive', to the US Census Bureau. See Whipple, *Vital statistics*, pp. 166–70. Mathematically, $W = 100 \times (n_{25} + n_{30} \ldots n_{60})/(\frac{1}{5}\sum_{i=23}^{62} n_i)$, where n_i is the number of observations at age i.

[4] Fischer, *Growing old*, pp. 82–6.

[5] Thomas, 'Numeracy', p. 128.

[6] Kaiser and Engel, 'Time- and age-awareness'.

[7] Dillon, *Shady side*.

[8] Herlihy and Klapisch-Zuber, 'Tuscans'.

[9] Emigh, 'Numeracy'.

sophistication' of a population.[10] In advocating this general interpretation, however, Mokyr remained agnostic about the exact meaning of age heaping: 'To the extent, therefore, that the degree of age heaping … [is] … correlated with other qualities such as arithmetical ability ("numeracy"), a respect for accuracy, or a more serious attitude toward time, age heaping measures valuable human attributes which have the potential to create important economic externalities and play a role in development'.[11] Individual cognitive skills ('arithmetical abilities') and a broader cultural outlook ('respect for accuracy' and 'a more serious attitude towards time') are complements in this interpretation. Subsequent work in economic history initially followed this line. Comparing literacy and age heaping, A'Hearn et al. concluded that '[b]oth measures offer a partial view of human capital, and both reflect not only individual but also broader social capabilities—a sort of administrative capital'.[12]

In the last decade, studies of age heaping in economic history have proliferated, covering times and places that range from Roman antiquity to the preconquest Incas and beyond.[13] In comparison to earlier historical studies of age distributions, these studies have in common an exclusive focus on human capital, and are sanguine about age heaping as an unproblematic indicator of basic numerical competences. Tollnek and Baten provide a very useful survey of this literature. Their conclusion is that age heaping 'enables us to assess basic numeracy for a large number of countries over a very long period of time'.[14] In this stream of research, the treatment of contextual factors such as the bureaucratization of everyday life or variation in cultural norms is limited to eliminating possible confounding influences on an estimate of numeracy. Thus, in regression analyses the number of censuses previously conducted might be included as a control variable for state capacity, or country fixed effects included to capture cultural differences. This body of work has established a robust historical correlation between age heaping and variables such as literacy and school enrolments, justifying its employment as an (inverse) index of human capital over time and across space. The success of the age-heaping approach to estimating human capital is not surprising. Compared to direct indicators such as literacy or school enrolments, for which data are scarce before the late nineteenth century, historical data on ages are abundant—and relatively easy to work with. However, notwithstanding its appeal and growing popularity, the numeracy interpretation of age heaping does not enjoy universal consensus. A major dissenting voice is Davids, in whose view a direct link between exact age reporting and numerical problem-solving ability has not been demonstrated.[15]

[10] Mokyr and Ó Gráda, 'Emigration and poverty'. The findings of the study were expanded in Mokyr, *Why Ireland starved*. Our discussion is based on this latter contribution.

[11] Mokyr, *Why Ireland starved*, p. 246.

[12] A'Hearn, Baten, and Crayen, 'Quantifying', p. 805.

[13] Baten and Priwitzer, 'Social and intertemporal differences'; Juif and Baten, 'On the human capital'. Prominent contributions include Crayen and Baten, 'Global trends'; de Moor and van Zanden, '"Every woman counts"'; Hippe and Baten, 'Regional inequality'; Baten, Crayen, and Voth, 'Numeracy'.

[14] Tollnek and Baten, 'Age heaping', p. 135. The characterization of age heaping given in the papers they survey ('useful indicator of basic numeracy', 'proxy for basic numerical skills', 'basic numeracy indicator', 'numeric skills of a population', 'measure of cognitive ability in the past', 'indicator of numeracy', 'basic form of human capital') confirms that the common feature of this research is to consider age heaping as a straightforward indicator of basic numerical skills.

[15] 'No one has as yet demonstrated for any society in the past that an individual with a lack of age awareness has a low degree of ability in quantitative reasoning—and the dearth of historical records may well make that assertion

Beltran Tapia et al. have provided another recent contribution that is sceptical of a simple numeracy interpretation.[16]

Interestingly enough, while economic historians have adopted age heaping as an indicator of numeracy, political scientists are inclined to consider it an indicator of state capacity.[17] Modern states raised age awareness directly via their efforts to ensure that citizens could be reliably identified, located, and monitored—part of a broader programme to make society *legible*, in Scott's phrase.[18] They kept vital records, issued and routinely demanded identity documents, enforced age-based education and military service, and granted age-contingent rights such as the franchise. All this incentivized, facilitated, and even compelled citizens to know their age. Lee and Zhang identify two main factors generating age heaping: first, individuals not knowing their exact age; and second, inaccurate age-reporting by census enumerators. Both factors reflect weak state capacity: scarce age awareness can be accounted for by lack of documentation, poor record keeping, and at best sporadic interaction between citizens and the state; and problems in census taking indicate a limited ability of the state to monitor its territory effectively.[19] The authors confirm this intuition in a large-scale empirical investigation covering 120 countries (with both national and sub-national observations) over the period 1960–2010. They find a systematic correlation between age heaping and indicators of state capacity such as the rule of law, political stability, control of corruption, government effectiveness, and fiscal revenues. Similarly, Suryanaryan and White use census age heaping to measure 'bureaucratic capacity' at the county level in the southern US in the late nineteenth century.[20]

The strong impact that census organization can exert on age-heaping patterns is also found in other studies. Szoltysek et al., in a study of the determinants of age heaping in a database comprising 115 European regions from the seventeenth to the early twentieth century, find that the quality of census procedures (carefully classified as 'pre-modern', 'semi-modern', and 'modern') is a robust predictor of age heaping. Similarly, Spenneman argues that age heaping in late nineteenth- and early twentieth-century Indian censuses was by and large due to an approximate assessment of age by the enumerators, since exact knowledge of age had very limited practical salience for the respondents.[21]

The spread of literacy (and numeracy) and the development of state capacity, with the attendant bureaucratization of everyday life, are interrelated aspects of the process that in the sociological literature has been called modernization. Unsurprisingly, age heaping has also been interpreted as a sign of incomplete or delayed modernization.[22] A further aspect of modernization is cultural change in

impossible to prove'; Davids, *Religion*, p. 62. Consider the recent study of Blum and Krauss, 'Age heaping'. With an unusually detailed (if very small) dataset, the authors are able to show that ages reported as multiples of five are indeed less accurate among German migrants in eighteenth-century Hungary. But a link between age misreporting and numerical abilities can only be established indirectly, via a correlation between 'heaped' ages and lower-skilled occupations.

[16] Beltran Tapia, Diez-Minguela, Martinez-Galarraga, and Tirado-Fabregat, 'Two stories'. For further critical observations on the numeracy interpretation of age heaping, see Fenoaltea, 'Choler', pp. 233–6.

[17] For a useful overview of the historical research on state capacity, see Dincecco, *State capacity*.

[18] Scott, *Seeing like a state*.

[19] Lee and Zhang, 'Legibility'.

[20] Suryanarayan and White, 'Slavery'.

[21] Szoltysek, Poniat, and Gruber, 'Age heaping'; Spenneman, 'Age heaping'.

[22] For example, by demographers Nagi, Stockwell, and Snavley, 'Digit preference', in the 1970s.

the direction of secularism, individualism, and linear time perception. Cultural differences in age awareness are another potentially significant determinant of age heaping. An interesting example comes from twentieth-century China. Age reporting there was 'astonishingly accurate' in censuses from 1953 to 1982, even among largely illiterate, rural groups. Jowett and Li's explanation is based on the practice of *naming* years after the familiar animals of the Chinese zodiac, making birth year easily memorable. In general, this information alone was sufficient for a census enumerator to infer an individual's age, but the calendrical system based on names actually included further elements allowing for a more precise translation into numerical ages and dates (for which purpose officials carried a ready conversion table). The exception proving the rule was the relatively literate, western province of Xianjiang, where age heaping was quite pronounced. The population here was dominated by the Uighur ethnic minority, which alongside smaller Kirgiz and Tajik groups maintained a clear cultural and linguistic separateness from the Han Chinese. Among the Uighur, 'birthdays are not particularly important days and they are not accorded celebrations and special observance'. Jowett and Li do not infer that these groups are innumerate, but rather that they 'neither know, nor care, how old they are'.[23]

By the same token, it is likely that the emergence of birthday celebrations as a significant marker in people's lives increased age awareness in industrializing societies.[24] A modern invention, or rediscovery, the celebration of birthdays has been linked with changing perceptions of time by Schmitt.[25] Time perception has been a staple of sociological theorizing since the work of late nineteenth-century pioneers like Durkheim and Simmel, who argued that the marking of time is a socially embedded process that serves the function of coordinating group activities.[26] In medieval Europe, the Catholic and Orthodox calendars of recurring religious feasts dovetailed with agricultural rhythms underlying social life. With modernization, this equilibrium was disrupted and attitudes towards time changed. In the West, the rise of Protestantism weakened the cult of saints and the general orientation of life around the liturgical calendar. Meanwhile urbanization and industrialization cut links with seasonal cycles in the natural world, and the rise of wage labour further focused attention on the value of time.[27] As Kaiser and Engel note, '[I]n urban industrial society "time rules life" … [a]nd age-awareness, together with numerous other time calculations built into labour and recreation, is very high'.[28] They interpret age heaping in early modern Russia as an indicator of the slow diffusion of the modern calendar, the clock, and 'merchant's time' there.

[23] Quotations from Jowett and Li, 'Age heaping', pp. 440, 427. Wang, Zeng, Jeune, and Vaupel, 'Age validation', is another contribution making an explicit connection between accurate age reporting and the Chinese calendar. Baten, Ma, Morgan, and Wang, 'Evolution', discount the influence of cultural outlook by showing that Chinese immigrants in the US in the period 1850–1910 tend to heap more on multiples of five than on 'dragon-year' births. This suggests only that the Chinese calendar induces no *distortion* in age reporting; it is perfectly consistent with the notion that it promotes more widespread age awareness. In fact, the authors' historical data on China show very low levels of age heaping in international comparative perspective.

[24] Shoham, 'It's about time'.

[25] Schmitt, 'L'invention'. See also Shoham, 'It's about time'.

[26] Sorokin and Merton, 'Social time'.

[27] Schmitt, 'L'invention'; Thompson, 'Time'.

[28] Kaiser and Enegl, 'Time- and age-awareness', p. 836.

With regard to birthdays specifically, the medieval Church had opposed their celebration as a pagan ritual. Calendar age was similarly unimportant in the medieval conception of a life course. The 'ages of man' were developmental rather than strictly chronological, marked by visible signs of maturation and senescence. Only in the modern, linear conception of time did the progressive accumulation of age, or 'capitalization of years', acquire greater meaning.[29] Individualism and, more speculatively, astrology may also have had roles to play in this development. Despite the early origins of such changes, Schmitt finds scant evidence before the late eighteenth century of birthday celebration—even among the rich and well born, even among individuals who knew their birthdate.[30] Birthday celebrations spread among the aristocracy and bourgeoisie only in the nineteenth century, and did not reach the working classes before the twentieth.[31]

If the 'numeracy' interpretation has underplayed the roles of culture and state capacity in the data-generating process, so too has it neglected the development of numeracy skills themselves, which have simply been associated with formal schooling. Yet there is considerable evidence that basic mathematical skills are also acquired outside school. Cognitive psychologists have found that modern children not only learn to count before starting school, but get as far as devising problem-solving algorithms for basic addition and subtraction.[32] Older children whose occupations keep them out of school but involve them in market transactions develop their own 'street math': non-standard methods for solving more complicated problems. In a classic study, unschooled Brazilian street vendors aged 10 to 12 years performed better on several mathematical tasks than their counterparts in urban schools in the 1980s.[33] A few decades earlier, the Soviet psychologist Luria studied cognition among rural Uzbeks in 1931–2. His interviews vividly revealed an unwillingness, or inability, to think outside the range of personal experience. Some of Luria's questions involved numerical problem-solving; for example, computing the time to reach a destination by bicycle, given that it took 30 minutes by foot and the bicycle was five times as fast. Because Luria wished to understand whether difficulty answering such questions was caused by a lack of basic numeracy or the inability to think abstractly, he posed the same question in a different form; for example, 'divide thirty cookies equally among five men'. Even for illiterate individuals who had never been to school, he found, 'simple computations used in everyday practical affairs presented no special difficulties'.[34] Even in the early modern period, numeracy was highly valued in a market economy and 'open to all sorts of people, as computational skills were developed in daily life'. Literacy, by contrast, required formal education and was thus 'more socially exclusive' in eighteenth-century Antwerp, according to

[29] Schmitt, 'L'invention'.
[30] Carl Friedrich Gauss was motivated to devise a method for calculating the date of Easter by a desire to discover his own birthdate. His mother had not recorded the date (which lets us infer his childhood birthdays were not celebrated), but easily remembered where it had fallen in the liturgical calendar: on a Wednesday, eight days before the Feast of the Ascension. His birthdate was 30 April 1777; Dunnington, *Gauss*, p. 69. In the English-speaking world, the diarist Samuel Pepys always noted his birthday, but never mentions any celebration of it.
[31] Schmitt, *L'invention.*
[32] Dehaene, *Number sense*, pp. 122–4.
[33] Saxe, 'Mathematics'.
[34] Luria, *Cognitive development*, p. 126.

Verhoeven.[35] These considerations suggest that, even for illiterates, innumeracy cannot simply be assumed when inaccurate age reporting is observed.

The take-away point of this discussion is that age heaping is the outcome of a complex process featuring many interacting factors. The literature we have just surveyed points to three main determinants: cognitive abilities (numerical fluency), which is the perspective currently adopted in economic history; state capacity, which is the perspective emerging among political scientists; and culture, which is emphasized more by social historians and sociologists. Given the absence of an indicator of numeracy other than age heaping itself and the interactions at play—involving simultaneous causation and feedback—severe problems of 'observational equivalence' can arise. For example, times and places with higher educational attainments will also be characterized by a higher level of state capacity and, possibly, more modern attitudes towards timekeeping, making it difficult to disentangle the various causes of age heaping (especially considering the difficulties involved in measuring culture and state capacity).

II

Italy conducted five censuses between Unification and the First World War.[36] For the first two, in 1861 and 1871, the Ministry of Agriculture, Industry, and Commerce tabulated and published single-year age distributions by province, broken down by literacy, gender, and marital status. In the census of 1881, single-year frequencies were published only for provincial capitals; elsewhere data were aggregated into five-year age groups. In the censuses of 1901 and 1911 only grouped five-year interval data were reported.[37] The national age distribution in 1871 is depicted in figure 1. Among adults, the heaping on multiples of 10 is unmistakable, as is a set of minor peaks at ages ending in 5.

Italians' preference for terminal digits zero and five ('0 and 5 ages') means that we can characterize the extent of age heaping using the Whipple index ('*W*'): the ratio of actual 0 and 5 age shares to the predicted share if terminal digits were uniformly distributed, that is one-fifth.[38] Italy's 1871 *W* value of 152 places it clearly behind northern Europe in accuracy of age reporting, on a par with other nations of Europe's southern and western periphery, but ahead of much of the rest of the world, including eastern Europe, Latin America, and parts of Asia.[39] In this section

[35] Verhoeven, 'Le pays', p. 237. The author accepts age heaping as a measure of numeracy, referring to A'Hearn et al., 'Quantifying', and asserting that it 'seems to be revealing about … basic arithmetic talents' (p. 228).

[36] The census scheduled for 1891 was cancelled due to a fiscal crisis; see Gallo and Paluzzi, 'Trasformazioni', p. 38.

[37] In evaluating census data, economic historians have been mostly concerned with the accuracy of occupational data; Vitali, *Aspetti*; Zamagni, 'Century of change'; Ciccarelli and Missiaia, 'Industrial labour force'.

[38] The formula given in n. 3 is commonly used, sometimes extending the upper age limit to 72. Here we employ a slightly different, two-step procedure: (i) for each 10-year age group {(23–32), (33–42), (43–52), (53–62), (63–72)} we calculate the proportion of 0 and 5 ages and divide by one-fifth; (ii) we then calculate the simple average across the five age groups, multiplying by 100 for ease of presentation. The two methods generally produce similar values, and similar rankings across times and places. For 1871, our two-step *W* is 152, as reported in the text, while the alternative formula yields 147. The advantages of the two-step procedure are: (a) it facilitates comparisons across provinces with different age structures; and (b) it dovetails neatly with our tracking of individual 10-year cohorts across censuses.

[39] Representative values can be taken from the clio-infra database at https://clio-infra.eu/Indicators/NumeracyTotal.html (accessed on 12 Dec. 2020). This source reports values for the ABCC index, which is a

Table 1. *Age heaping by macro-area, gender, literacy, and residence*

	Urban				*Rural*			
	Literate		*Illiterate*		*Literate*		*Illiterate*	
	M	*F*	*M*	*F*	*M*	*F*	*M*	*F*
North	110	127	119	137	109	119	119	130
Centre	117	132	145	154	119	126	146	153
South	129	138	173	185	141	149	191	198
Italy	119	132	147	160	124	132	154	162

Note: Simple averages of provincial Whipple index values, ages 23–72. Urban refers to provincial capital cities, rural to the remainder of the population.
Macro-areas are defined as follows (also in the following tables and figures): north: Piedmont, Liguria, Lombardy, Veneto; centre: Emilia, Tuscany, Marche, Umbria, Lazio; south: Abruzzi, Puglia, Campania, Basilicata, Calabria, Sicily, Sardinia.
Sources: Authors' own calculations; MAIC, *Censimento 1871*.

we examine four anomalies in census age heaping which cannot be explained in terms of numeracy: variation by literacy, over time, by gender and marital status, and across alternative sources.

The key evidence supporting age heaping as a proxy for individual cognitive ability is its correlation with illiteracy.[40] With separate age distributions for those who could and could not read, the census of 1871 allows us to investigate this correlation for Italy. Before proceeding, we should ask whether census information on literacy, which was self-reported, can be trusted. Comparisons with estimates from alternative sources suggest the answer is yes. Province-level correlations between census literacy and marriage register signature rates, school enrolments 10 years earlier, and conscript literacy test results are in the range +0.93 to +0.96 for both men and women.[41]

W values are presented for literates and illiterates in table 1, by 'macroarea', and in figure 2, by province. Both presentations show systematically greater age heaping among illiterates; this is consistent with age heaping representing a failure to learn numeracy skills in school. Yet a closer look reveals patterns that do not sit comfortably with this interpretation. There is, in particular, huge variation in age heaping even among the educated. In figure 2 more than 50 points separate the minimum and maximum *W* values for literates. In table 1, the gap is 40 points between educated rural men in the north and educated rural women in the south.

simple transformation of the Whipple index; see A'Hearn et al., 'Quantifying', p. 788. We convert these back to the original *W* values. For the 1830s birth decade (aged in their 40s in 1871), we have Sweden (102), France (110), Italy (148), Spain (149), Poland (190), Argentina (214), and India (364). Another way to contextualize Italy's age heaping is the 1955 UN *Demographic Yearbook* (pp. 19–20) classification of census data: less than 105 'highly accurate', 105–109.9 'fairly accurate', 110–124.9 'approximate', 125–174.9 'rough', 175 and more 'very rough'. In this perspective, Italy's 1871 census would be in the rough data category, alongside countries such as Brazil (1950), Venezuela (1950), and the Philippines (1948).

[40] Similar to the analysis developed here, Beltran Tapia et al., 'Two stories', document several anomalous correlations between age heaping and literacy in Spanish census data.

[41] Marriage register signature rates in 1872 are from MAIC, *Annuario statistico italiano 1905–7*, p. 255. For comparison, census literacy is calculated for men aged 21–32 and women aged 17–26 in 1871. Enrolments in academic year 1862–3 are from MIP, *Istruzione*, and refer to both public and private schools. Provinces of the Veneto and Lazio regions, not yet part of Italy in 1862, are excluded. Conscript literacy test results are from MAIC, *Annuario statistico italiano 1905–7*, p. 252. For a comprehensive overview of trends in literacy during the nineteenth century, see Ciccarelli and Weisdorf, 'Pioneering into the past'.

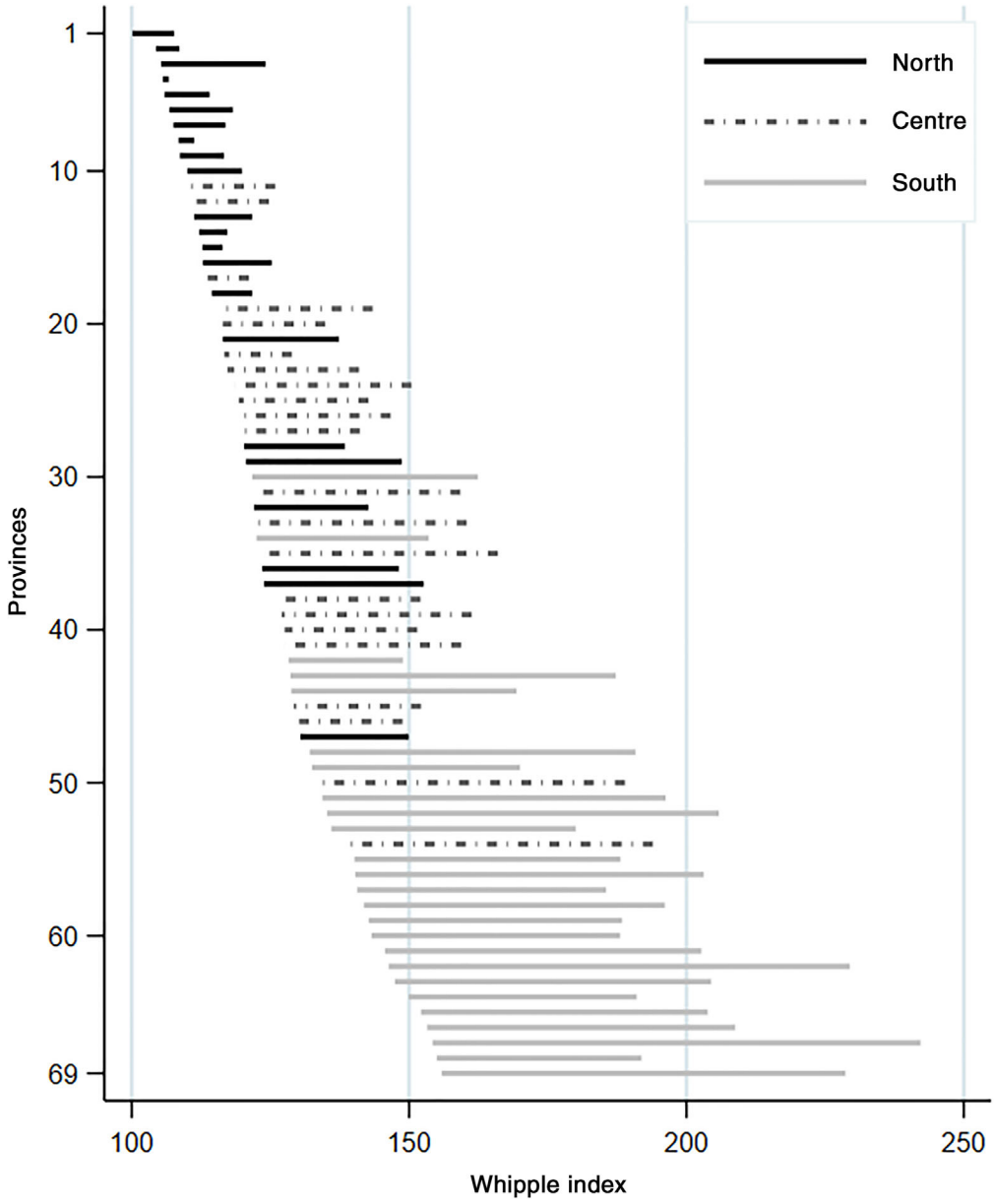

Figure 2. *Age heaping by province, literate and illiterate, 1871*
[Colour figure can be viewed at wileyonlinelibrary.com]

Note: Whipple index for literates (lower/left value) and illiterates (higher/right value). Ages 23–72. Provinces ranked 1–69 by age heaping among literates: (1) Belluno, Udine, Sondrio, Treviso, Rovigo, Verona, Bergamo, Padova, Vicenza, (10) Mantova, Reggio Emilia, Modena, Como, Venezia, Cremona, Brescia, Bologna, Milano, Ascoli Piceno, (20) Parma, Novara, Ferrara, Ravenna, Siena, Piacenza, Massa-Carrara, Forlì, Porto Maurizio, Pavia, (30) Chieti, Ancona, Torino, Pesaro, Teramo, Arezzo, Alessandria, Cuneo, Lucca, Grosseto, (40) Pisa, Macerata, L'Aquila, Catania, Palermo, Firenze, Livorno, Genova, Agrigento, Campobasso, (50) Perugia, Messina, Trapani, Benevento, Roma, Lecce, Caltanissetta, Salerno, Potenza, Caserta, (60) Avellino, Bari, Cagliari, Foggia, Catanzaro, Napoli, Cosenza, Sassari, Reggio Calabria, Siracusa.
Sources: Authors' own calculations; MAIC, *Censimento 1871*.

It could be argued that the quality of instruction in numeracy skills specifically varied across provinces, and was lower on average in the south; or, notwithstanding the correlations just reported, perhaps self-reported literacy was exaggerated in some, mainly southern, provinces. However, if either of these conjectures were valid, we would expect a smaller gap between illiterates and literates in the south (since many purported literates there would be either mathematically weak or entirely uneducated). In fact just the opposite is true: the age-heaping gap between literates and illiterates is by far the greatest in the south—as much as 50 points for rural men. If we restrict attention to illiterates, neither such complication arises. Self-declared illiterates had little or no mathematical instruction of any quality, no matter where they were; nor had they exaggerated their literacy. Yet among illiterates too there is tremendous variation in age heaping: more than 100 points (!) between the minimum and maximum provincial values in figure 2; and 79 points between illiterate rural men in the north and illiterate rural women in the south in table 1.

The variation we have just documented *within* the categories of literate and illiterate dwarfs the variation *between* them. On average, W is 126 for literates and 158 for illiterates in the census data.[42] An informative counterfactual is obtained by imputing these national averages to the literates and illiterates of each province. The resulting counterfactual W, in which differences stem exclusively from differences in literacy rates, ranges quite narrowly: from 137 (Bergamo) to 154 (Caltanissetta), with a standard deviation of 4.8. The true range of W is dramatically greater: from 105 (Belluno) to 230 (Sassari), with a standard deviation of 31.9.

Clearly, differences *within* literacy categories—by province, by gender, and by residence—are decisive in forming the observed geographic patterns in age heaping. Average differences *between* literates and illiterates—which could plausibly reflect numeracy skills learned in school—do not get us very far. We need to understand how it could be that masses of illiterate peasant women in the north ($W = 130$) reported their ages just as accurately as the few educated men of the cities in the south ($W = 129$). Plainly, an explanation in terms of individual numeracy does not suffice here.

If age heaping primarily reflects numeracy skills, then a gradual improvement in educational attainment should produce a gradual decline in age heaping as younger, better-educated cohorts gradually displace earlier generations in the population. Tracking any specific cohort, instead, age heaping should *rise* as individuals who have long since completed their education grow older and report their age less accurately.[43]

Charting age heaping across Italian censuses is complicated by the switch to tabulation by five-year age groups mentioned earlier: this aggregation masks the pattern of heaping in the underlying data. Because that pattern was dominated by heaping on multiples of 10, however, it is still detectable: frequencies are systematically greater in age ranges with a multiple of 10, such as 30–4, compared

[42] These are simple averages of province-specific W values for total literates and illiterates (not broken down by gender or urban/rural residence).

[43] On the tendency for older individuals to report their age less accurately, see Crayen and Baten, 'Global trends', app. A.

Figure 3. *Age heaping, 1861–1911, by macro-area (T index)*
[Colour figure can be viewed at wileyonlinelibrary.com]

Note: Simple averages of provincial T index values as described in section II. The provinces of Lazio and Veneto are omitted throughout for comparability with 1861. The 20–9 age group is omitted throughout for comparability with 1901.
Source: Authors' own calculations based on MAIC, *Censimento generale* (1865); idem, *Censimento generale* (1875); idem, *Censimento della popolazione* (1883); idem, *Censimento della popolazione* (1904); idem, *Censimento della popolazione* (1914).

to age ranges without one, such as 35–9. This regularity allows us to construct a workable alternative to the Whipple index.

We begin by calculating ratios for each age group in the intervals {(30–9), (40–9), (50–9), (60–9), (70–9)} which take the form $\tau_{30-39} = 100 \cdot (f_{30-34})/(f_{35-39})$, where f stands for frequency and the subscripts denote the age range.[44] Because of heaping on multiples of 10, these ratios tend to be greater than 100. Of course, they will also exceed 100 even in the absence of heaping, if the age distribution of the population has the classic pyramid shape. Accordingly, we normalize the τs by expressing them relative to a prediction based on the overall age structure of the population.[45] A simple average of the normalized τs across the five age groups gives us an overall index that we call T (for tens). Figure 3 plots the results for the north, centre, south, and islands.

In the north and centre, the T index declines steadily, in the way that might be produced by a steady increase in schooling among younger generations. The south follows a very different trajectory: much higher and quite stable across the first three censuses, then dropping sharply almost to the levels of the north–centre in 1901. In figure 4 we track individual cohorts through time, restricting attention to the south. We expect age heaping to rise as a cohort grows older, a near-universal feature of historical age distributions, and that is exactly what we see across the first three

[44] The 1901 census tabulation does not permit the calculation of τ for the (20–29) age group.
[45] More specifically, we non-parametrically estimate a smoothed age distribution for each province, compute from it a set of predicted $\hat{\tau}$ ratios, then calculate the actual/predicted $\tau/\hat{\tau}$ ratio to obtain a normalized value. For more details, see online app. S1.

Figure 4. *Age heaping in the south, 1861–1911, by cohort (τ values)*
[Colour figure can be viewed at wileyonlinelibrary.com]

Note: Simple averages of normalized τ values, calculated as described in section II, for southern provinces. Lines connect the values for a particular 10-year cohort across successive censuses. Labels give the age the youngest members of the cohort reached in a given census.
Source: As for fig. 3.

censuses. Consider, for example, the cohort born 1831–40: $\tau = 106$ in 1861, when they are in their 30s; $\tau = 121$ in 1871, when they are in their 40s; and $\tau = 143$ in 1881 when they are in their 50s. From 1901 to 1911, similarly, τ increases for each of the cohorts initially aged 40–9, 50–9, and 60–9. The year 1901 stands out as a stark anomaly. Not only does heaping fail to increase for any of the cohorts we can track from 1881; it actually *decreases*.

Could it be that the numeracy of southern adults improved decisively between 1881 and 1901? On the face of it, this seems quite implausible. There is no record of a successful mass education initiative targeting adults of the relevant ages in the south, and self-reported *literacy* for the cohorts observed in both censuses hardly changed between 1881 and 1901. The only plausible interpretation is that data-gathering procedures changed between these censuses, and there is plenty of independent evidence that this did indeed happen.

The basic element of the census was a family questionnaire (*scheda di censimento*) to be completed by heads of household. In a world of pervasive illiteracy, it was often a neighbour who wrote out the family's responses—a potential source of inaccuracies right from the start. In the censuses of 1861–81, it was local governments (*comuni*, or municipalities) that were charged with collecting data and conducting basic counts. Their tasks included generating a complete list of households; distributing census questionnaires; collecting the completed forms and, where these were not forthcoming, sending an interviewer to record verbal responses; checking the forms for errors; transcribing from the family return a set of individual records; tabulating from the individual records the population

totals by age, sex, literacy, and marital status; and forwarding the results to the Direzione Generale della Statistica in Rome. To carry out these tasks *comuni* were directed to engage temporary clerks and enlist the voluntary efforts of local notables, professionals, schoolteachers, and government employees. Crucially, no funding for these operations was provided by the central government. Therefore, the success of the census depended on the active cooperation of both ordinary people and local elites. Neither could be assumed.

In the run-up to the census of 1871, Luigi Bodio, head of the Direzione Generale, worried that the situation was not propitious. The clergy had been alienated by the recent annexation of Rome, and:

> the suspicion that has grown among the people that the census must serve fiscal purposes. And they are not wrong … The criteria for distributing taxes, conscription, etc. are based on the number of inhabitants. And the poor multitudes feel the burdens, more than the honours, of being citizens … With the census, it is a question of fighting a great battle: a battle against ignorance, against the schemes and mechanisms of concealment of taxpayers …[46]

Appealing to mayors for information on, *inter alia*, anthropometric measurements of the local population, the Italian Society for Anthropology and Ethnology in 1872 found it necessary to reassure them that 'these statistical data cannot appear to citizens as having any relation with public finance'.[47] Mistrust also characterized relations within government, between centre and periphery. During preparations for the census of 1881 Bodio lamented the performance of under-resourced and unmotivated local governments, focusing specifically on age data:

> I could give the names of any number of *comuni* whose tabulations of the population by age were written from the imagination; but they were invented poorly, because in the gradation by age the number of elderly was greater than that of the middle-aged. If the municipal employees charged with making those counts had invented with artfulness, perhaps none of us could have detected the falsification. With all this, and in spite of numerous cases of negligence and bad faith being reported to the Ministry of the Interior, it was never possible to ensure the exemplary punishment of those who should have been held responsible.[48]

What could be done? The census of 1901 saw several changes implemented. Households now completed both a detailed family form *and* summary individual forms, eliminating the transcription process formerly carried out by the *comune*, and the individual forms were now forwarded directly to Rome for processing, eliminating local tabulation of basic statistics.[49] A technical adjustment aimed at improving the accuracy of age distributions (indispensable for estimating mortality tables)[50] was to drop the age question from the family questionnaire and replace it

[46] Quoted in Favero, 'Dibattito', p. 125, our translation.
[47] Raseri, 'Materiali', p. 3, our translation.
[48] Favero, *Le misure*, p. 138, our translation.
[49] Gallo and Paluzzi, 'Trasformazioni', pp. 38–9. The census of 1901 was also the first in which mechanical computation (Hollerith machines) was employed.
[50] See Rameri, 'Legge statistica', for a coeval analysis of how age heaping affected mortality statistics.

with a question about birth year and month.[51] Local census workers were exhorted to ensure birthdates were recorded 'with scrupulous precision', even checking them against birth records when comparison revealed an inconsistency with declared age (still reported on the individual forms).[52] Meanwhile outreach efforts aimed at overcoming the indifference, or, in some locations, outright hostility of the population were stepped up. In preparation for the census of 1901, schoolteachers were urged both to explain to their pupils the nature and importance of the census, and to convene meetings of local heads of household for the same purpose, while also advising on how to fill out the census forms.[53] The central government's budgetary provision for the census of 1901 was increased by more than 40 per cent relative to 1881.[54]

The decrease in age heaping in 1901 thus illustrates the state's increased allocation of resources to census operations, its enhanced technical competence, its increasing success in overcoming the suspicions and enlisting the cooperation of its citizens, and its growing ability to monitor and control the actions of local government. These are all aspects of state capacity, which explains not only variation in census age heaping over time, but also across regions. In the south, it was difficult to induce *comuni* to comply with basic legal requirements such as keeping proper accounts and publishing budgets, to say nothing of effectively providing infrastructure, health services, and education.[55] At the other extreme were towns in recently annexed Veneto pleading to be allowed to carry out *ad hoc* local censuses as a foundation for municipal population registers.[56] Persistent variation in the competence and commitment of local officials had historical roots in the pre-Unification states; the same was true of attitudes toward the state among local citizens.[57]

In table 2 several proxy indicators of state capacity in Italian states just prior to Unification are presented: they align well with age heaping in 1871. In the Kingdom of the Two Sicilies, where age heaping was highest, official statistical publications per capita were lowest. Its land register—vitally important not only for property transfers but also for tax purposes—was the only one still based exclusively on verbal declarations of landowners rather than an expert survey and valuation.[58] Progress in unifying the system of weights and measures, measured by the number of pages required to list all the relevant conversions into the decimal metric system, was slowest here and in the former Papal State. Mass conscription made considerable demands on state capacity at both the central and local level (where pre-Unification attitudes and institutions persisted), suggesting a fourth measure: draft evasion. The data in table 2 show that in 1863 there was a large gap

[51] MAIC, *Relazione sul metodo*, p. LVIII, 193/6. Here Italy followed international best practice, as per the recommendations of the International Statistical Conference of 1872; Dillon, *Shady side*, p. 42. A discussion of age questions in the early censuses can be found in ISTAT, *Censimenti*, pp. 28–36.
[52] MAIC, *Relazione sul metodo*, p. 209.
[53] Ibid., pp. 216, 230.
[54] From 600,000 (720,000 after cost overruns) in 1881 to 845,000 lire in 1901, while prices fell roughly 5%. Local government spending on the census remained steady at roughly 2,000,000 lire; MAIC, *Relazione generale*, p. xcviii; idem, *Relazione sul metodo* pp. cxli, 166/7; idem, 'Risultati', pp. 18–19.
[55] Randeraad, *Authority*, ch. 6.
[56] Favero, *Le misure*, p. 57.
[57] For an intriguing recent interpretation of the development of state capacity in Italy's pre-Unification states, see Dincecco, Federico, and Vindigni, 'Warfare'.
[58] Zamagni, *Dalla periferia*, pp. 70–3.

Table 2. *Indicators of state capacity, 1850s*

	Age heaping	*Draft evasion*	*Statistical publications*	*Cadaster/land registry*	*Weights and measures*	*Tax revenue*
Piedmont	133	3.8	100	Mixed	3.1	100
Lombardy-Venetia	115	1.8	102	Survey	2.9	83
Parma	134	1.8		Survey	2.0	59
Modena	126	1.3		Mixed	2.0	63
Tuscany	146	1.7	64	Survey	2.9	63
Papal State	153	17.0	49	Survey	4.0	70
Two Sicilies	181	7.2	41	Descript.	4.1	55

Notes: Tax revenue: total ordinary revenue of central government p.c., relative to Piedmont's 30.46 lire; estimate for Kingdom of the Two Sicilies for continental provinces only. Draft evasion: percentage of draft dodgers (class of 1843 called to arms in 1863–4) relative to men declared fit to serve and not granted an exemption. Publishing: number of statistical works published 1850–9 per million population, relative to Piedmont's 8.0. Cadaster: basis of the land register as described in section II. Weights and measures: diversity of units, measured as pages per *circondario*—a sub-provincial administrative unit, larger than the *comune* (municipality)—in the territories of the pre-Unification states in the 1877 table of conversion factors. Age heaping: Whipple index; Piedmont refers to Piedmont and Liguria only.
Sources: Authors' own calculations; MAIC, *Censimento 1871*; MG, *Della leva*; Patriarca, *Numbers*, p. 241; Zamagni, *Dalla periferia*, p. 73; MAIC, *Tavole di ragguaglio*; Romani, *Storia economica*, p. 404.

between the north the and centre–south in compliance with the draft. A final index of state capacity is tax revenue per capita. Here too the north–south gradient is hard to overlook. Individually, these indicators are only suggestive, but they all point in the direction of a significant lag in state capacity in those parts of Italy where age heaping was most pronounced. Overall, these results are in line with the recent research agenda of political scientists who have argued for a close link between age heaping and state capacity. Not surprisingly, in the context of nineteenth-century Italy this connection took the form of a north–south divide.

Women exceeded men in both age heaping (157 vs. 144) and illiteracy (81 vs. 63) in 1871.[59] However, women's excess age heaping is *not* well explained by their relative lack of education: table 1 reveals a gender gap even among the educated: 13 points among urban literates; 8 among rural literates. The gender gap in age heaping is largest in the urban north, precisely where women's education was most advanced.[60]

Further insight into the gender gap in age heaping can be gained by examining estimates by marital status (table 3). Two striking patterns emerge. First, age-heaping levels are much higher among widows than married women. Second, age-heaping gender gaps are much the smallest in the married category. In the south this effect is truly dramatic: the gender gap is large among single persons (+26) and bigger still among the widowed (+44), but *negative* (–6) among the married. These effects are *not* driven by age differences between the categories; for example, by widows being older than married women and reporting age inaccurately for that reason only. Our calculations give equal weight to all age groups precisely in order to minimize such effects. In any case, in the data underlying table 3 these same patterns show up within every age group.

[59] The same pattern is evident in 1861. The illiteracy rates have been calculated just as the Whipple values: calculated for each age group in each province, averaged within provinces, and finally averaged across provinces.
[60] Women's illiteracy rates, calculated again as simple averages across age groups and provinces, were 49, 65, and 79% in the provincial capitals of the north, centre, and south, respectively; in rural areas the corresponding rates were 71, 85, and 95%.

Table 3. *Age heaping by macro-area, gender, and marital status, 1871*

	Male			*Female*			*Gap (female–male)*		
	Single	*Married*	*Widowed*	*Single*	*Married*	*Widowed*	*Single*	*Married*	*Widowed*
North	115	113	120	132	121	139	16	8	20
Centre	135	137	147	148	141	171	13	4	25
South	149	184	189	175	171	233	26	−6	44
Italy	134	146	153	153	148	184	19	1	30

Note: Simple averages of provincial Whipple index values, ages 23–72. Gaps may not equal female–male differences due to rounding.
Sources: Authors' own calculations; MAIC, *Censimento 1871*.

If age heaping primarily reflected human capital, these differences across marital status would be evidence of selection effects. Consider the contrast between married and single persons. The estimates in table 3 would imply no meaningful selection into marriage among men in the north and centre, and some positive selection among women. In the south, by contrast, table 3 would suggest no meaningful selection among women, and powerful *negative* selection into marriage among men: *W* is 35 points higher for married than single men there. This last is difficult to rationalize.

The situation is even more perplexing when we consider the contrast between widows and wives. The individual numeracy interpretation would imply powerful negative selection into widowhood, especially in the south, where heaping was fully 62 points higher among widows than married women. It is not hard to imagine explanations for negative selection into *early* widowhood: women with low human capital might marry less healthy or older men, or might have less success in remarrying, for example.[61] There are both empirical and conceptual issues with such explanations, but the fundamental problem is that negative selection into early widowhood implies, *ipso facto*, positive selection into later widowhood. Among progressively older individuals, then, the difference between widowed and married women should grow progressively smaller. In fact the *opposite* is true, especially in the south: here the widowed–married difference is +31 points for women aged 23–32 and +67 points for women aged 63–72.

If differences in numeracy are implausible, what *does* explain differences in age heaping by marital status? Note first that ages were normally reported by men. Census officials intended the head of household to fill out the family questionnaire, delegate the job to a trusted person, or report verbally to a census interviewer. Only rarely were women heads of household. Even among widows, only 16 per cent were heads of household in one study; the vast majority lived as dependents in families where a relative—typically a married son—was the head of household.[62] Consider also that the social convention in Italy was for wives to be somewhat younger than their husbands. Rettaroli reports a four-year gap in mean age at first marriage in the second half of the nineteenth century.[63]

[61] Livi Bacci, 'Remarriage', shows that in the 1880s remarriage was common. Among individuals losing a spouse before the age of 50, 40% of women and 73% of men remarried.
[62] Kertzer and Karweit, 'Widowhood'. The study refers to Casalecchio di Reno, near Bologna, in 1881.
[63] Rettaroli, 'Age at marriage'; see also MAIC, *Movimento*, p. xvii. Corsini, 'Remarriage', referring to the charivari (*scampanata*), reminds us that local social norms regarding age differences were enforced.

It appears that some men invented an age for their wives consistent with a conventional, respectable gap relative to their own age. This had the effect of lowering age heaping among married women. In cases where the men's ages were not heaped, their wives inherited the same low degree of heaping. In cases where the men's ages were highly heaped, their wives' ages would actually be *less* heaped, unless the conventional age gap were exactly five years. These two situations characterized the north and south, respectively.[64] Turning to widows, in cases where wives' ages were adjusted to their husbands', there would be a natural 'rebound' in age heaping when women began once again to report their own age. Additional factors working in the same direction were, first, understatement of age, usually involving rounding, by widows with an eye on remarriage; and second, inaccuracy in ages reported indirectly by heads of household such as adult children, not influenced (as husbands had been) by social convention.

Such patterns are not unique to Italy. Several recent studies have pointed to a possible role of cultural expectations concerning spousal ages in shaping age-heaping patterns. France's census of 1851, for example, shows the Italian pattern of lower age heaping among married women, compared with singles and widows, but not among married men.[65] For nineteenth-century Ireland, Blum et al. compare age heaping among married women in the census and alternative sources in which they reported their ages directly, finding that census age heaping was biased down.[66] In both British (1851) and American (1880) censuses, married women's ages are significantly less heaped than singles and widows, which Földvári et al. attribute to adjustment of wives' reported ages to those of their husbands.[67] At present, we lack the sort of microdata that would provide more definitive evidence on why age heaping in Italy varied so much with marital status. What we can say at this stage is that in Italy, women's census ages were distorted by being reported by others and through the influence of social norms.

Selection effects are a central concern in the literature on emigration from Italy, one of Europe's principal sending countries.[68] Gomellini and Ó Gráda explored whether age heaping could resolve the ambiguities and controversies in the debate. Comparing Italian-born individuals in the US census with the population in the Italian census, they found greater age heaping among emigrants, implying negative selection.[69] Remarkably, when they investigated a different source they found a completely different picture: Italian passengers arriving in New York on the steamer *SS Roma* between 1902 and 1905 reported their ages with near-perfect accuracy.[70] In table 4 we explore such comparisons further.

[64] A complete account requires explaining why married men in the South reported their ages less accurately than single men. It is likely that this resulted from exaggerating their age, associating greater age with the gravitas befitting a paterfamilias.

[65] Perrin, 'Gender gap'.

[66] Blum, Colvin, McAtackney, and McLaughlin, 'Women of an uncertain age'. They conclude by suggesting particular caution when dealing with age heaping data in traditional societies: 'we recommend that female age heaping estimates should be used with caution for patriarchal societies where the census is the only available source' (p. 219).

[67] Földvári, van Leeuwen, and van Leeuwen-Li, 'How did women count?'.

[68] As reported by Gomellini, Ó Gráda, and Vecchi, 'Migration', p. 222, there were 12,452,285 departures between 1869 and 1910 alone.

[69] Gomellini and Ó Gráda, 'Migrations'; Gomellini et al., 'Migration', p. 228.

[70] *W* values for the *Roma* passengers are 102 for men and 94 for women. We thank Cormac Ó Gráda for sharing these unpublished results.

Table 4. *Age heaping among Italian emigrants in the US, 1898–1912*

	Literate	*Illiterate*	*Male*	*Female*
US census, 1910	145	157	150	149
NY arrivals, 1898–1912	100	101	99	105
South	*100*	*100*	*99*	*103*
Italy census, 1871	123	164	144	159
South	*141*	*191*	*177*	*190*

Note: Whipple index calculated for macro-area and national aggregates rather than at provincial level.
Sources: Authors' own calculations; MAIC, *Censimento 1871*; Ruggles et al., 'IPUMS USA'; Battery Conservancy, 'Castle Garden database'.

The top row of table 4 presents *W* values for Italian immigrants in the US census of 1910.[71] We restrict attention to individuals arriving in 1898 or later, aged 23–72 at the time of the census. We know from other sources that this group consisted overwhelmingly of southern Italians, who accounted for 81 per cent of registered departures for the US and Canada in the years 1898–1910.[72] Age heaping in the US census is surprisingly high. In Italy, by this time, we know it was minimal, even in the south (figure 3). The Italian-American *W* values for 1910 look more like Italy in 1871 than Italy in 1911. This suggests strong negative selection.

The second row in table 4 shows an alternative set of *W* values based on passenger lists from Italian emigrant ships arriving in New York City between 1898 and 1912, aged 23–72.[73] This group should comprise approximately the same Italian-born individuals reported in the census of 1910, but the Whipple index looks completely different. There is essentially no age heaping at all: not for literates, not for illiterates, not for men, and only a hint for women. This suggests strong *positive* selection, if age heaping measures numeracy.

This example calls attention once again to the importance of the historical data-generating process. American census data were self-reported, or in some cases estimated by an enumerator. Passenger list data were compiled by onboard clerks, among whom 'sloppiness was extraordinarily rare'.[74] The information, for which the ship's commanding officer was responsible, was used in a high-stakes, potentially adversarial interview with immigration officials. Meanwhile, from 1901, Italian legislation obliged overseas emigrants to obtain a passport, in which their birthdate was recorded. Everything about this process generated the perception that accuracy was important; and in these circumstances, accuracy was forthcoming.

[71] Ruggles, Flood, Goeken, Grover, Meyer, Pacas, and Sobek, 'IPUMS USA'. Selecting all individuals born in Italy from the complete count census data, we have 1,351,055 observations. Narrowing to those aged 23–72 arriving in 1898 or later, we have 673,250 cases.
[72] Commissariato Generale dell'Emigrazione, *Annuario statistico*, p. 149.
[73] Battery Conservancy, 'Castle Garden Database'. We thank Ariell Zimran for sharing the data with us. Castle Garden was New York's main immigrant processing centre until 1891, after which it moved to the better-known Ellis Island. The full sample numbers 2,013,269; this falls to 827,727 when we impose the restrictions that age is in the range 23–72 (for consistency with other estimates presented here) and year of arrival is 1898 or later.
[74] Spitzer and Zimran, 'Migrant self-selection', p. 230. The authors, who use height as a measure of selectivity and find evidence of positive selection, carefully examined original passenger lists.

III

Beyond the quantitative anomalies emerging from census data, qualitative evidence on Italian social history during the nineteenth century suggests caution in assuming that individuals reporting heaped ages were severely innumerate. Numbers were everywhere in the worlds of both peasants and city dwellers, even in their games. In the ubiquitous game of *morra* the object is to guess a sum quickly: the total number of fingers extended simultaneously by the players.[75] Meanwhile, the state lottery (*lotto*) ensured that from the late seventeenth century onward all Neapolitans knew their numbers, at least from 1 to 90.[76] A passion for the *lotto* extended from the Bourbon King himself to the denizens of the city's poorest slums, who gambled on a regular basis. The legal minimum wager was only one *grano*, a fraction of the price of a loaf of bread; in practice, shares as small as a twelfth of the minimum were available on a clandestine basis from the neighbourhood *lotto* shop operator, often on credit.

Printed lottery tickets specifying the numbers chosen by the bettor and the potential winnings, as well as the public display of the week's winning numbers at *lotto* shops, presumed widespread recognition of written numbers. So too did the popular books on the numerological interpretation of dreams (*smorfia*). Some of these *smorfia* manuals were intended for illiterates and matched illustrations of dreams with the associated lucky numbers written in Arabic numerals. Writing in the 1880s, the celebrated journalist Matilde Serao noted that the *smorfia* was so ingrained in Neapolitan habits of thought that a colloquial expression for crazy was to call someone '*nu vintiroie*'—a '22'; to call a woman a '78' was something worse. In Serao's time, illegal private lotteries with minimum wagers as low as a couple of *centesimi* flourished alongside the Italian state *lotto.* Receipts, though written in pencil on dirty scraps of paper, had the canonical form of old, specifying the bettor's wager, his chosen numbers, and the associated payoffs.[77]

Of course, superstitions like the *smorfia* are often cited as evidence of a pre-modern, non-rational attitude, just as addiction to the *lotto* is supposed to be

[75] Dickens vividly described the game in his *Pictures from Italy* (1846): 'The most favourite game is the national one of Mora, which they pursue with surprising ardour, and at which they will stake everything they possess. It is a destructive kind of gambling, requiring no accessories but the ten fingers, which are always—I intend no pun—at hand. Two men play together. One calls a number—say the extreme one, ten. He marks what portion of it he pleases by throwing out three, or four, or five fingers; and his adversary has, in the same instant, at hazard, and without seeing his hand, to throw out as many fingers, as will make the exact balance. Their eyes and hands become so used to this, and act with such astonishing rapidity, that an uninitiated bystander would find it very difficult, if not impossible, to follow the progress of the game … It is never the quietest game in the world; for the numbers are always called in a loud sharp voice, and follow as close upon each other as they can be counted. On a holiday evening, standing at a window, or walking in a garden, or passing through the streets, or sauntering in any quiet place about the town, you will hear this game in progress in a score of wine-shops at once; and looking over any vineyard walk, or turning almost any corner, will come upon a knot of players in full cry. It is observable that most men have a propensity to throw out some particular number oftener than another; and the vigilance with which two sharp-eyed players will mutually endeavour to detect this weakness, and adapt their game to it, is very curious and entertaining'; Dickens, *Pictures*, pp. 45–6. The game was especially popular in Naples in the first half of the nineteenth century according to Rocco, 'Il gioco'; on its continuing popularity in the southern region of Abruzzo today, see Spitilli, 'Di passata'.

[76] Macry, *Giocare*. In the *lotto*, a random number between 1 and 90 was drawn from each of five urns. The gambler's wager was that one or more of three numbers he or she had chosen would be drawn. Winnings were proportional to the amount wagered, with the ratio lowest for just one of the bettor's numbers appearing (*c.* 12:1) and highest for a triple in which all three were drawn (as high as 5000:1, depending on the period).

[77] Serao, *Il ventre*, pp. 57, 64.

evidence of a passive, fatalistic belief that one's future is a matter of chance, impervious to control or plan (though possibly susceptible of divination). We do not challenge that interpretation, though it is worth noting that playing numbers suggested by dreams or current events was a perfectly reasonable strategy given the structure of *lotto* payoffs.[78] Our argument is only that the popularity of the game suggests the ability to count and to recognize and name numbers was widespread.

There were matters more serious than gambling that required basic numeracy. Near the subsistence minimum, there is no room for errors such as a sharecropper failing to recognize that his share of 86 sacks of grain is 43, a shepherd leaving behind uncounted sheep in a mountain pasture, or an urban household being cheated by the baker or grocer. Federico has shown that Italian agriculture, notwithstanding its low productivity, was quite market-orientated, with on-farm consumption accounting for only about 30 per cent of total production.[79] In this context, peasant families developed 'survival tactics' based on rudimentary cost–benefit analyses, which surely required a modicum of numerical skill.[80] To make this essentially evolutionary argument is not to claim that the average Italian landless labourer understood compound interest, or to deny recent findings that poverty-induced stress impairs cognition.[81] It is merely to assert that there were strong incentives for a family to acquire a minimal competence in numerical reasoning.[82]

The tools of basic numerical competence were there to be acquired, though not necessarily in forms familiar to us today. Methods of counting with one's fingers to 100 and beyond, though no longer set out in Italian mathematics textbooks in the nineteenth century as they had been earlier, remained common in practice.[83] Even if few ordinary Italians were proficient with quill and ink (skills requiring years of practice), recording numbers temporarily by means of traditional systems of marks made with a piece of charcoal or a sharp object was a daily practice. So too was use of the abacus.[84] For record-keeping, tally sticks and their like were often used. Shepherds had employed these for centuries throughout Europe, and they continued in use into the nineteenth century. Tally sticks recording a transaction and split down the middle into a matched pair, one half for each party, were recognized as legal documents in Napoleon's Civil Code. The system was still common in the twentieth century in Puglia, where illiterate shepherds not only recorded commercial transactions in this way, but numbered their sheep in the same fashion, making incisions along the upper and lower edges of the ears

[78] The *lotto* did not split a fixed total award among all holders of winning tickets. In that case, it would have been wise to avoid popular numbers chosen by other gamblers. Instead, the *lotto* paid a *fixed multiple* of the amount wagered to *every* winning player. The operation made big losses when popular numbers were drawn from the urns.

[79] Federico, 'Mercantilizzazione'.

[80] Federico, 'Contadini'.

[81] Mani, Mullainathan, Shafir, and Zhao, 'Poverty'.

[82] Dealing with an earlier period, Emigh is impressed by the 'bottom-up' numeracy of ordinary Tuscan farmers, who could estimate average yields and knew the value of their assets. At the same time, they reported their ages inaccurately. She concludes that 'inaccuracies in age reporting may not necessarily reflect lack of numeracy per se; it may reflect a somewhat narrower lack of knowledge of the biological age'; Emigh, 'Numeracy', p. 670.

[83] Roggero, 'Conti'.

[84] Here 'abacus' means placing tokens on lines drawn on paper or in the sand, or incised in a board. Roggero, 'Conti', notes that by 1800 abacus methods had mostly disappeared from textbooks and schools, but must have remained in common use.

according to rules for denoting units, tens, hundreds, and thousands.[85] Written record-keeping using non-standard, ideographic systems of numeration has been documented for illiterate Umbrian peasants and Venetian fishermen in the late nineteenth century (see online appendix S2).[86] These studies suggest that, even among unskilled occupations, some basic numeric knowledge was widespread.

On reflection, this survey of numbers in everyday Italian life is entirely in line with the studies on the acquisition of numerical competences in informal contexts reported in section I. In fact, it would be astonishing if ordinary Italians of the nineteenth century failed to acquire this minimal sort of numerical competence. The limits of finger-counting, street-math, and the tally stick are clear; they are not good substitutes for proper education in basic arithmetical operations. The claim for numeracy made here is only that the typical Italian had the cognitive capacity required to recognize, comprehend, state, and record the numbers relevant for human age.

As we observed in section I, an important cultural factor potentially impinging on age awareness is the celebration of birthdays, which remind a person on a regular basis of his or her age, as well as the ages of family members and friends. There is no evidence that birthdays were commonly celebrated in ordinary Italian families in this period. Italy seems to have lagged behind northern Europe in adopting the birthday. Impressionistic evidence is provided by the frequency of birthday words in printed works. In the upper panel of figure 5 the trends are rising in Italian, English, and German, but latest and most weakly for the Italian *compleanno*.[87]

The lower panel of figure 5 shows that '*onomastico*'—celebration of one's name day, the feast day of the saint after whom one is named—appeared more commonly than *compleanno* until relatively recently. In southern Italy, at least in Naples, the *onomastico* was more commonly celebrated than the birthday as late as 1960, as we are reminded in a recent novel: 'Although custom had it that it was the *onomastico* that should be celebrated—birthdays were then considered irrelevant—the Sarratores and Nella insisted on arranging a little party'.[88] The celebration of name days, which need not coincide with birthdays and, in the case of popular names, may be shared with other family members or religious holidays, is not focused on the individual's age. To the extent that celebration of individual age achieved on birthdays was a modern cultural practice diffusing in the nineteenth century from north to south, from rich to poor, and from city to country, we have a natural explanation for temporal, occupational, and geographic differences in age heaping in the Italian context.

IV

The findings reported in this article suggest particular caution in interpreting age heaping as a reliable 'general-purpose' measure of numeracy. The evidence from

[85] La Sorsa, 'Storia'.
[86] Nicasi, 'Segni numerici'; Ninni, *Segni.*
[87] The evidence is drawn from Google's Ngram Viewer. It should be borne in mind that the size and composition of the underlying corpora of digitized books vary across languages and over time.
[88] Ferrante, *L'amica geniale*, p. 221.

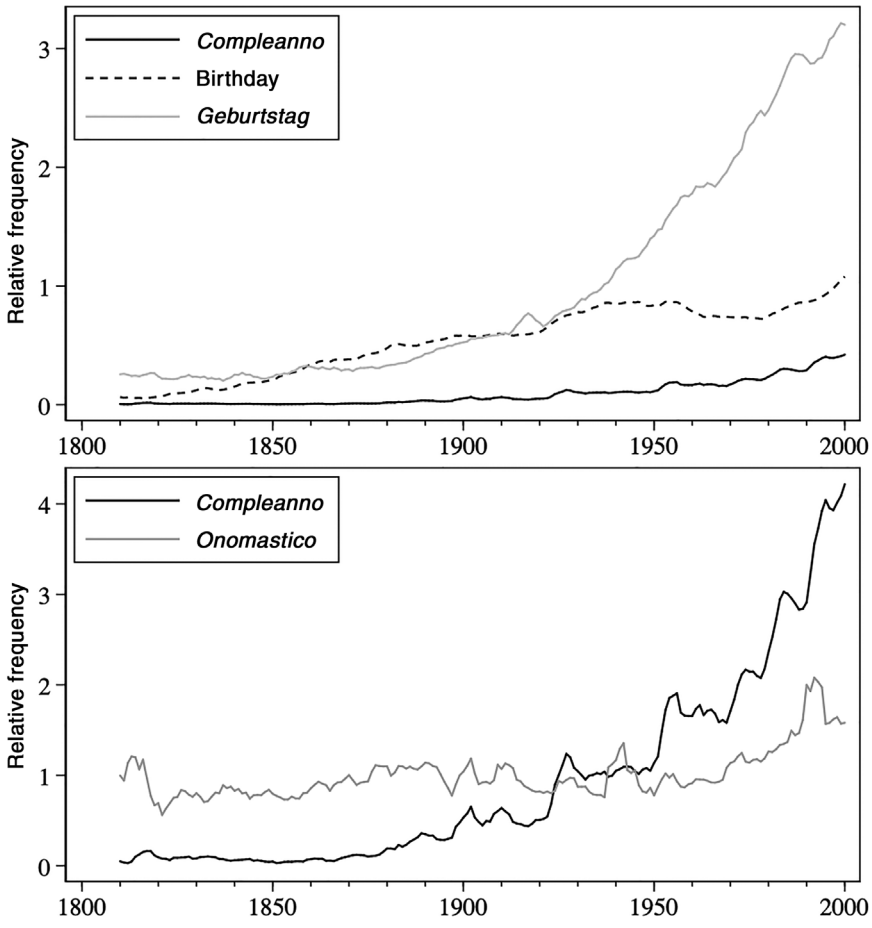

Figure 5. *Frequency of birthday words*

Note: Relative frequency of birthday and nameday words appearing in scanned English, Italian, and German publications. Five-year moving averages. Upper panel frequencies per 10^{-5}; lower panel per 10^{-6}.
Source: Google Ngram Viewer (accessed on 13 Jan. 2021).

other disciplines reviewed in section I indicates two ways in which the link between age heaping and numeracy can break down. First, age heaping in historical sources may not represent individuals' knowledge of their age faithfully. Second, ignorance of age—even correctly measured—cannot be assumed to mean ignorance of basic numerical skills. If these problems are conceptually distinct, they are in practice difficult to disentangle. The typical Italian of 1871 was simultaneously poorly educated, immersed in a market economy and surrounded by numbers in daily life, unlikely to care about age in the first place and sometimes culturally motivated to distort it, disengaged from a mistrusted state, and surveyed by a government with limited capacity to gather and process data. The Italian evidence of sections II and III reveals that all of these factors influenced census age heaping. Education, and with its school-inculcated arithmetic skills, surely has a role to play, but cannot explain most of the variation in age heaping over time, between groups, and across places.

In this perspective, age heaping should be more properly seen as an outcome of modernization, a complex historical process featuring institutional change (the emergence and consolidation of state capacity), cultural shifts (secularism, individualism, and respect for precision), and the spread of basic education (literacy and numeracy).[89] It remains, on this basis, an important indicator for diagnosing and predicting economic development.[90] The use of age heaping as indicator of numerical skills is not completely unwarranted, but it cannot be based simply on the broad correlation between age heaping, literacy, and other schooling indicators. Rather, it must be grounded in research designs that are alert to the specificities of context, that are able to control effectively for the various factors other than numerical skills that shape age heaping in a particular historical instance.[91]

DOI: 10.1111/ehr.13087

Footnote references

A'Hearn, B., Baten, J., and Crayen, D., 'Quantifying quantitative literacy: age heaping and the history of human capital', *Journal of Economic History*, 69 (2009), pp. 783–808.
Baten, J., Crayen, D., and Voth, H.-J., 'Numeracy and the impact of high food prices in industrializing Britain, 1780–1850', *Review of Economics and Statistics*, 96 (2014), pp. 418–30.
Baten, J., Ma, D., Morgan, S., and Wang, Q., 'Evolution of living standards and human capital in China in the 18th–20th centuries: evidence from real wages, age-heaping and anthropometrics', *Explorations in Economic History*, 47 (2010), pp. 347–59.
Baten, J. and Priwitzer, S., 'Social and intertemporal differences in basic numeracy in Pannonia (first century BCE to third century CE)', *Scandinavian Economic History Review*, 63 (2015), pp. 110–34.
Beltrán Tapia, F., Diez-Minguela, A., Martinez-Galarraga, J., and Tirado-Fabregat, D., 'Two stories, one fate: age-heaping and literacy in Spain, 1877–1930', *Revista de Historia Economica* (forthcoming).
Blum, M., Colvin, C. L., McAtackney, L., and McLaughlin, E., 'Women of an uncertain age: quantifying human capital accumulation in rural Ireland in the nineteenth century', *Economic History Review*, 70 (2017), pp. 187–223.
Blum, M. and Krauss, K.-P., 'Age heaping and numeracy: looking behind the curtain', *Economic History Review*, 71 (2018), pp. 464–79.
Cafagna, L., 'Modernizzazione attiva e modernizzazione passiva', *Meridiana*, 2 (1988), pp. 229–40.
Ciccarelli, C. and Fachin, S., 'Regional growth with spatial dependence: a case study on early Italian industrialization', *Papers in Regional Science*, 96 (2017), pp. 675–95.
Ciccarelli, C. and Missiaia, A., 'The industrial labor force of Italy's provinces: estimates from the population censuses, 1871–1911', *Rivista di Storia Economica*, 29 (2013), pp. 141–91.
Ciccarelli, C. and Weisdorf, J., 'Pioneering into the past: regional literacy developments in Italy before Italy', *European Review of Economic History*, 23 (2019), pp. 329–64.
Corsini, C. A., 'Why is remarriage a male affair? Some evidence from Tuscan villages during the eighteenth century', in J. Dupâquier, E. Hélin, P. Laslett, M. Livi Bacci, and S. Sogner, eds., *Marriage and remarriage in populations of the past* (1981), pp. 385–95.
Crayen, D. and Baten, J., 'Global trends in numeracy 1820–1949 and its implications for long-term growth', *Explorations in Economic History*, 47 (2010), pp. 82–99.
Davids, K., *Religion, technology, and the great and little divergences: China and Europe compared, c. 700–1800* (Leiden, 2013).

[89] We use the term 'modernization' in a descriptive way, without endorsing any specific variant of sociological theory. For a useful recent take on modernization, firmly locating its origins in economic change, see Inglehart and Welzel, *Modernisation*. Influential interpretations in Italian economic history depict two qualitatively different processes of modernization taking place in the North and in the South; see Cafagna, 'Modernizzazione'; Felice and Vasta, 'Passive modernization?'.

[90] We read the contribution of Ciccarelli and Fachin, 'Regional growth', in this light. Interpreting Italian age heaping in the 1871 census as a measure of numeracy, the authors include it in a composite human capital index that has substantial explanatory power for industrial productivity growth.

[91] Baten et al., 'Numeracy', is an important example of a research design formulated along these lines. The authors' focus on short-term changes, rather than long-run trends, allows them to argue that increased age heaping in early nineteenth-century Britain reflected decreased individual cognitive ability (linked with deteriorating early childhood nutrition) rather than cultural or administrative change.

Dehaene, S., *The number sense. How the mind creates mathematics* (Oxford, 1997).
de Moor, T. and van Zanden, J. L., '"Every woman counts": a gender-analysis of numeracy in the Low Countries in the early modern period', *Journal of Interdisciplinary History*, 41 (2010), pp. 179–208.
Dickens, C., *Pictures from Italy* (1846).
Dillon, L., *The shady side of fifty. Age and old age in late Victorian Canada and the United States* (Montreal, 2008).
Dincecco, M., *State capacity and economic development. Present and past* (Cambridge, 2017).
Dincecco, M., Federico, G., and Vindigni, A., 'Warfare, taxation, and political change: evidence from the Italian Risorgimento', *Journal of Economic History*, 71 (2011), pp. 887–914.
Dunnington, G. W., *Carl Friedrich Gauss: titan of science* (New York, 1955).
Emigh, R. J., 'Numeracy or enumeration? The use of numbers by states and societies', *Social Science History*, 26 (2002), pp. 653–98.
Favero, G., 'Il dibattito sull'organizzazione dei primi censimenti unitari negli atti della Giunta e del Consiglio Superiore di Statistica', *Annali di Statistica*, XII, 2 (2012), pp. 121–45.
Favero, G., *Le misure del regno: direzione di statistica e municipi nell'Italia Liberale* (Padua, 2001).
Federico, G., 'Mercantilizzazione e sviluppo economico italiano (1861–1940)', *Rivista di Storia Economica*, 3 (1986), pp. 149–86.
Federico, G., 'Contadini e mercato: tattiche di sopravvivenza', *Società e Storia*, 4 (1987), pp. 877–913.
Felice, E. and Vasta, M., 'Passive modernization? The new Human Development Index and its components in Italy's regions (1871–2007)', *European Review of Economic History*, 19 (2015), pp. 44–66.
Fenoaltea, S., 'Choler: the failure of the Cliometric School', *Annals of the Fondazione Luigi Einaudi*, 54 (2020), pp. 231–42.
Ferrante, E., *L'amica geniale* (Rome, 2011).
Fischer, D. H., *Growing old in America* (Oxford, 1977).
Földvári, P., van Leeuwen, B., and van Leeuwen-Li, J., 'How did women count? A note on gender-specific age heaping differences in the sixteenth to nineteenth centuries', *Economic History Review*, 65 (2012), pp. 304–13.
Gallo, G. and Paluzzi, E., 'Le trasformazioni del censimento della popolazione in Italia: i cambiamenti della macchina organizzativa censuaria negli ultimi 150 anni', *Annali di statistica*, XII, 2 (2012), pp. 31–75.
Gomellini, M. and Ó Gráda, C., 'Migrations', in G. Toniolo, ed., *The Oxford handbook of the Italian economy since Unification* (Oxford, 2013), pp. 271–302.
Gomellini, M., Ó Gráda, C., and Vecchi, G., 'Migration', in G. Vecchi, ed., *Measuring wellbeing. A history of Italian living standards* (New York, 2017), pp. 215–54.
Herlihy, D. and Klapisch-Zuber, C., *Tuscans and their families: a study of the Florentine Catasto of 1427* (New Haven, Conn., 1985 edn.).
Hippe, R. and Baten, J., 'Regional inequality in human capital formation in Europe, 1790–1980', *Scandinavian Economic History Review*, 60 (2012), pp. 254–89.
Inglehart, R. and Welzel, C., *Modernization, cultural change, and democracy. The human development sequence* (Cambridge, 2005).
Jowett, A. J. and Li, Y.-Q., 'Age heaping: contrasting patterns from China', *GeoJournal*, 28 (1992), pp. 427–42.
Juif, D. and Baten, J., 'On the human capital of Inca Indios before and after the Spanish Conquest. Was there a "pre-colonial legacy"?', *Explorations in Economic History*, 50 (2013), pp. 227–41.
Kaiser, D. H, and Engel, P., 'Time- and age-awareness in early modern Russia', *Comparative Studies in Society and History*, 35 (1993), pp. 824–39.
Kertzer, D. I. and Karweit, N., 'The impact of widowhood in nineteenth-century Italy', in D. Kertzer and P. Laslett, eds., *Aging in the past: demography, society, and old age* (Berkeley, Calif., 1995), pp. 229–49.
La Sorsa, S., 'Storia e folclore della mena delle pecore in Puglia', *Archivio Storico Pugliese*, 6 (1953), pp. 487–99.
Lee, M. M. and Zhang, N., 'Legibility and the informational foundations of state capacity', *Journal of Politics*, 79 (2017), pp. 118–32.
Livi Bacci, M., 'On the frequency of remarriage in nineteenth century Italy: methods and results', in J. Dupâquier, E. Hélin, P. Laslett, M. Livi Bacci, and S. Sogner, eds., *Marriage and remarriage in populations of the past* (1981), pp. 347–62.
Luria, A. R., *Cognitive development. Its cultural and social foundations* (Cambridge, Mass., 1976).
Macry, P., *Giocare la vita. Storia del lotto a Napoli tra Sette e Ottocento* (Rome, 1997).
Mani, A., Mullainathan, S., Shafir, E., and Zhao, J., 'Poverty impedes cognitive function', *Science*, 341 (2013), pp. 976–80.
Mokyr, J., *Why Ireland starved: a quantitative and analytical history of the Irish economy* (Boston, Mass., 1983).
Mokyr, J. and Ó Gráda, C., 'Emigration and poverty in prefamine Ireland', *Explorations in Economic History*, 19 (1982), pp. 360–84.
Nagi, M. H,, Stockwell, E. G., and Snavley, L. M., 'Digit preference and avoidance in the age statistics of some recent African censuses: some patterns and correlates', *International Statistical Review*, 41 (1973), pp. 165–74.
Nicasi, G., 'Dei segni numerici usati attualmente dai contadini della valle di Morra nel territorio di Città di Castello', *Bollettino della Regia Deputazione di Storia Patria per l'Umbria*, 12 (1906), pp. 267–76.
Ninni, A. P,. *Sui segni prealfabetici usati anche ora nella numerazione scritta dai pescatori clodiensi* (Venice, 1889).
Patriarca, S., *Numbers and nationhood. Writing statistics in nineteenth-century Italy* (Cambridge, 1996).

Perrin, F., 'Gender gap in numeracy and the role of marital status. Evidence from 19th century France', *Revue d'économie Politique*, 130 (2020), pp. 51–76.
Rameri, L., 'Legge statistica della influenza del sesso sulla durata della vita umana in Italia', *Annali di Statistica*, ser. 2, 10 (1879), pp. 1–174.
Randeraad, N., *Authority in search of liberty. The prefects in Liberal Italy* (Amsterdam, 1993).
Raseri, E., 'Materiali per l'etnologia italiana', *Annali di statistica*, ser. 2, 8 (1879), pp. 1–206.
Rettaroli, R., 'Age at marriage in nineteenth-century Italy', *Journal of Family History*, 15 (1990), pp. 409–25.
Rocco, E., 'Il gioco della mora', in F. de Bourcard, ed., *Usi e costumi di Napoli e contorni*, vol. 1 (Naples, 1857), pp. 67–70.
Roggero, M., 'Conti sulle dita, calcoli a penna. L'aritmetica elementare a fine Settecento', *Studi Storici*, 35 (1994), pp. 1039–60.
Romani, M., *Storia economica dell'Italia nel secolo XIX*, 1815–1882 (Bologna, 1982).
Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M., 'IPUMS USA: version 10.0' (2020), https://doi.org/10.18128/D010.V10.0.
Saxe, G. B., 'The mathematics of child street vendors', *Child Development*, 59 (1988), pp. 1415–25.
Schmitt, J. C., 'L'invention de l'anniversaire', *Annales. Histoire, Science Sociales*, 62 (2007), pp. 793–835.
Schmitt, J. C., *L'invention de l'anniversaire* (Paris, 2012).
Scott, J. C., *Seeing like a state. How certain schemes to improve the human condition have failed* (New Haven, Conn., 1998).
Serao, M., *Il ventre di Napoli* (Milan, 1884).
Shoham, H., 'It's about time: birthdays as modern rites of temporality', *Time and Society* , 30 (2020), pp. 78–99.
Sorokin, P. A. and Merton, R. K., 'Social time: a methodological and functional analysis', *American Journal of Sociology*, 42 (1937), pp. 615–29.
Spenneman, D. H. R., 'Age heaping among Indian hawkers in south-eastern Australia and their source communities in the Punjab', *Journal of Punjab Studies*, 24 (2017), pp. 149–202.
Spitilli, G., 'Di passata in passata. Il gioco della morra nell'Abruzzo settentrionale', in A. M. Pioletti, ed. *Giochi, sport tradizionali e società. Viaggio tra la Valle d'Aosta, l'Italia e l'Unione Europea* (Aosta, 2012), pp. 104–21.
Spitzer, Y. and Zimran, A., 'Migrant self-selection: anthropometric evidence from the mass migration of Italians to the United States, 1907–1925', *Journal of Development Economics*, 134 (2018), pp. 226–47.
Suryanaryan, P. and White, S., 'Slavery, reconstruction and bureaucratic capacity in the American South', *American Political Science Review*, 115 (2021), pp. 568–84.
Szoltysek, M., Poniat, R., and Gruber, S., 'Age heaping patterns in Mosaic data', *Historical Methods*, 51 (2018), pp. 13–18.
Thomas, K., 'Numeracy in early modern England: the Prothero Lecture', *Transactions of the Royal Historical Society*, 37 (1987), pp. 103–32.
Thompson, E. P., 'Time, work-discipline, and industrial capitalism', *Past and Present*, 38 (1967), pp. 56–97.
Tollnek, F. and Baten, J., 'Age heaping based human capital estimates', in C. Diebolt and M. Haupert, eds., *Handbook of cliometrics* (Heidelberg, 2016), pp. 131–54.
Verhoeven, G., '"Le pays où on ne sait pas lire": literacy, numeracy and human capital in the commercial hub of the Austrian Netherlands (1715–75)', *European History Quarterly*, 44 (2014), pp. 223–43.
Vitali, O., *Aspetti dello sviluppo economico italiano alla luce della ricostruzione della popolazione attiva* (Rome, 1970).
Wang, Z., Zeng, Y., Jeune, B., and Vaupel, J. W., 'Age validation of Han Chinese centenarians', *Genus*, 54 (1998), pp. 123–41.
Whipple, G. C., *Vital statistics. An introduction to the science of demography* (New York, 1919).
Zamagni, V., 'A century of change: trends in the composition of the Italian labour force, 1881–1981', *Historical Social Research*, 44 (1987), pp. 36–97.
Zamagni, V., *Dalla periferia al centro. La seconda rinascita economica dell'Italia (1861–1990)* (2nd edn., Bologna, 1993).

Official publications

Battery Conservancy, 'Castle Garden Database' (2009), http://www.castlegarden.org (accessed on 5 Dec. 2020).
Commissariato Generale dell'Emigrazione, *Annuario statistico della emigrazione italiana dal 1876 al 1925* (Rome, 1926).
ISTAT (Istituto nazionale di Statistica), *Censimenti e società. Mutamenti sociodemografici della Sicilia in 150 anni di storia* (Rome, 2014).
MAIC (Ministero di Agricoltura, Industria e Commercio), *Censimento generale della popolazione del 31 dicembre 1861*, vol. II (Turin, 1865).
MAIC (Ministero di Agricoltura, Industria e Commercio), *Censimento generale al 31 dicembre 1871 della popolazione del Regno d'Italia*, II: *Popolazione classificata per età sesso, stato civile e istruzione elementare* (Rome, 1875).
MAIC (Ministero di Agricoltura, Industria e Commercio), *Tavole di ragguaglio dei pesi e misure già in uso nelle varie provincie del regno col peso metrico decimale* (Rome, 1877).
MAIC (Ministero di Agricoltura, Industria e Commercio), *Popolazione. Movimento dello stato civile. Anno XX –1881* (Rome, 1882).

MAIC (Ministero di Agricoltura, Industria e Commercio), *Censimento della popolazione del Regno d'Italia al 31 dicembre 1881*, II: *Popolazione classificata per età sesso, stato civile e istruzione elementare* (Rome, 1883).
MAIC (Ministero di Agricoltura, Industria e Commercio), *Censimento della popolazione del Regno d'Italia al 31 dicembre 1881, Relazione generale* (Rome, 1885).
MAIC (Ministero di Agricoltura, Industria e Commercio), Censimento della popolazione del Regno d'Italia al 10 febbraio 1901, *II:* Numero delle famiglie e numero degli abitanti classificati secondo la qualità della dimora, il luogo di nascita, il sesso, l'età, lo stato civile e l'istruzione (Rome, 1904).
MAIC (Ministero di Agricoltura, Industria e Commercio. Direzione Generale della Statistica), *Relazione sul metodo di esecuzione e risultati del censimento, raffrontati con quelli dei censimenti italiani precedenti e di censimenti esteri*, 5: *Censimento della popolazione del Regno d'Italia al 10 febbraio 1901* (Rome, 1904).
MAIC (Ministero di Agricoltura, Industria e Commercio. Direzione Generale della Statistica), *Annuario statistico italiano 1905–1907* (Rome, 1908).
MAIC (Ministero di Agricoltura, Industria e Commercio), 'Risultati di un'inchiesta eseguita dalla Direzione Generale della Statistica circa l'ordinamento degli Uffici Centrali di Statistica dei principali stati', *Bollettino del Ministero di Agricoltura, Industria e Commercio*, anno VIII, vol. II, fascicolo 5 (Rome, 1909).
MAIC (Ministero di Agricoltura, Industria e Commercio), *Censimento della popolazione del Regno d'Italia al 10 giugno 1911*, II: *Popolazione presente classificata per sesso, età, stato civile e istruzione* (Rome, 1914).
MIP (Ministero dell'Istruzione Pubblica), *Istruzione Pubblica e Privata. Anno Scolastico 1862–3. Parte Prima. Istruzione Primaria* (Turin, 1865).
MG (Ministero della Guerra. Direzione Generale delle Leve, Bassa-Forza e Matricola), *Della leva sui giovani nati nell'anno 1843 e delle vicende dell'esercito dal 1o ottobre 1863 al 30 settembre 1864. Relazione del Maggior Generale Federico Torre al Sig. Ministro dellaGuerra* (Turin, 1865).
UN (United Nations), *Demographic Yearbook* (New York, 1955).

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

S1. Computing the T index
S2. Numeracy among the Chioggia fishermen