LEM
WORKING PAPER SERIES

Rethinking Age-Heaping. A Cautionary Tale from Nineteenth Century Italy

Brian A'Hearn [§]
Alexia Delfino *
Alessandro Nuvolari °

[§] Pembroke College, University of Oxford, UK
* London School of Economics, UK
° Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy

RETHINKING AGE-HEAPING
A CAUTIONARY TALE FROM NINETEENTH CENTURY ITALY

Brian A'Hearn
(Pembroke College, University of Oxford)

Alexia Delfino
(London School of Economics)

Alessandro Nuvolari
(Sant'Anna School of Advanced Studies, Pisa)

Abstract

A growing literature employs age-heaping as an indicator of human capital, more specifically of numeracy. We re-examine this link in light of a range of evidence from nineteenth century Italy, which suggests that age-heapers were most likely ignorant of their age, not their numbers. Though it can stand in as an acceptable proxy for literacy, our findings indicate that age-heaping is most plausibly interpreted as an indicator of cultural and institutional modernisation rather than a measure of cognitive skills.

1. Introduction

It is more than half a century since the concept of human capital crystallized in the work of pioneers like Mincer, Schultz, and Becker. Spreading from development to labour economics, economic history and most recently growth theory, human capital remains central to how we think about both individual and aggregate economic performance. *Historical* research on human capital has faced serious data constraints, for information on school enrolments, educational attainment, literacy rates, and (even more so) skills is widely available only from the late-nineteenth century, and even then typically fraught with gaps, inconsistent reporting, and other shortcomings. For earlier periods, researchers have employed considerable ingenuity in distilling proxies from the existing sources. For health status, they have made use of information on heights, which reflect net nutrition in the growing years. For education, they have turned to signature rates on marriage registers and other documents as an indicator of literacy (Cipolla 1969). For the supply and demand of artisanal skills, they have studied the skill premium, the ratio of the daily wages of carpenters or masons relative to unskilled construction labourers (van Zanden 2009).

Recently, a new human capital proxy has gained traction among economic historians: age-heaping. Irregularities in an age distribution, in particular the systematic overrepresentation of – or heaping upon – popular ages, indicate that people do not know, have reasons not to reveal, or are not able to compute exactly their true age. Instead they report a round or otherwise attractive number. When age-heaping results from an inability to understand numbers, to count accurately, or to undertake elementary calculations, it can be taken as an indicator of basic numerical skills, or "numeracy". Beginning with a paper by A'Hearn, Baten and Crayen (2009), the recent age-heaping literature has expanded to cover a wide array of countries and historical periods, from Roman antiquity (Baten and Priwitzer 2015) to the preconquest Incas (Juif and Baten 2013) and beyond. Prominent contributions include Crayen and Baten (2010), de Moor and van Zanden (2010), Manzel, Baten and Stolz (2012), Hippe and Baten, (2012), and Baten, Crayen and Voth (2014).

Age-heaping had long been recognized as a problem by demographers; textbooks of the early twentieth century illustrated how its severity could be measured by calculating an "index of concentration" that is still in use today (Whipple, 1919; 196-70). The interest of historians in age-heaping may date to the late 1970s. Studying the U.S. from colonial times to the mid-twentieth century, Fischer (1977; 82-86) detected a change in popular attitudes towards youth in the increasing tendency of people to report younger ages at critical junctures, e.g. 39 rather than 40. Closer to home, from the perspective of this study, Herlihy and Klapisch-Zuber (1985; 159-82) documented age-heaping in medieval Tuscany in considerable detail. In that context, a notable improvement in the accuracy of age-reporting was attributed to the pressure exerted by fiscal authorities collecting the data for tax purposes. In a later study focused specifically on age-heaping, Kaiser and Engel (1993) employed it as a measure of "time- and age-awareness" in the pre-industrial society of early eighteenth century Russia.

In perhaps the first use of age-heaping by an *economic* historian Mokyr (1983) suggested that it could serve as a proxy for the degree of "quantitative sophistication"

of a population. In advocating this general interpretation, 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 (Mokyr 1983; 246).

In Mokyr's view, then, age-heaping could be both an index of the distribution of individual skills ("arithmetical abilities"), and in this way a plausible proxy for individual human capital, and a reflection of factors with a broader, contextual dimension such as "respect for accuracy" or "a more serious attitude towards time".

Subsequent studies have documented the existence of a robust correlation between age-heaping and human capital indicators such as a literacy and primary schooling enrolments, at both the individual and aggregate level (A'Hearn, Baten and Crayen, 2009; Crayen and Baten, 2010; Hippe and Baten, 2012).[1] Though a direct link between historical age-heaping and numerical problem solving ability has not been demonstrated, recent contributions increasingly elide this connection and interpret inaccurate age reporting as a straightforward and unambiguous indicator of math skills, numerical cognitive ability, or numerical education (e.g. Baten and Juif 2014; see also Tollnek and Baten 2016).[2] The success of the age-heaping approach to estimating human capital is not surprising. Historical data on ages are abundant and relatively easy to handle. Moreover, age-heaping's potential insights into basic computational ability accord well with the emphasis in recent research on cognitive skills and abilities, rather than "inputs" such as school enrolments (Hanushek and Woessman, 2008).

In this paper, we reappraise the relationship between age-heaping and human capital using evidence from nineteenth century Italy. Nineteenth century Italy is a promising test bed for such an inquiry. The Kingdom of Italy was a newly created country, in which the national census acquired important political connotations in terms of nation building. Alongside the data of the censuses, therefore, there arose an intriguing literature commenting and discussing the results and their interpretation (Patriarca 1996). Furthermore, Italy was (and remains) a country characterized not only by major geographic differences in economic performance and living standards (Vecchi 2011), but also in institutions and cultural attitudes, so much so that influential interpretations depict two qualitatively distinct historical processes of modernisation

---

[1] A'Hearn, Crayen and Baten (2009) studied the connection between literacy and age-heaping at the individual level using data from the 1850, 1870, and 1900 US censuses. Besides a negative effect of individual literacy on the probability of reporting a popular, rounded age, their regressions also revealed a significant role for the literacy rate of a person's birth region. Hence their study supports both an "individualist" and a broader, "contextualist" interpretation. Most recent contributions seem to lean more towards the former. For example, in the much-used clio-infra data bank (https://www.clio-infra.eu/), age-heaping is assuredly labelled as "numeracy" and placed alongside average years of education.

[2] An exception is Baten, Crayen and Voth's (2014) study of age-heaping among the English cohorts affected by high grain prices during the Napoleonic wars. The authors acknowledge the potential influence of cultural norms and administrative practices before dismissing such factors for their case. They proceed to interpret age-heaping as a measure of cognitive ability.

in the North and the South (Cafagna 1988; Felice and Vasta 2015). This heterogeneity along several dimensions makes the Italian case particularly interesting for probing further the nature and meaning of age-heaping. Finally, the early Italian censuses report unusually rich age data, with single-year age distributions by province, gender, marital status, and literacy.
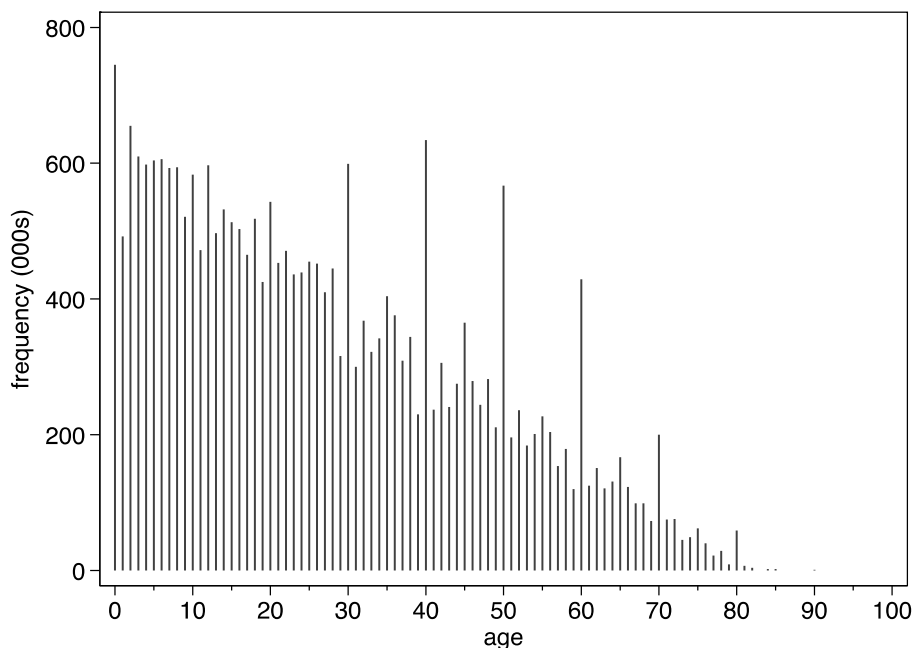
The remainder of the paper is structured as follows. In the next section we examine age-heaping patterns in successive Italian censuses, and investigate their correlation with age, literacy, gender, marital status. In Section 3, using micro data from death registers, we explore similar issues at the individual level in three provinces of Southern Italy. In Section 4, we discuss our findings in the light of evidence of a more qualitative nature that sheds light on the numerical skills and age-awareness of Italians in the second half of the nineteenth century. Section 5 summarizes our argument and draws conclusions for research on age-heaping.

2. Evidence from the Italian census

2.1. *Measuring age-heaping*

The Italian age distribution in 1871 is depicted in Figure 1. Among adults, the heaping on multiples of ten is unmistakable, as is a set of minor peaks associated with ages ending in 5. More subtle but quite regular is a preference for even numbers.

Fig. 1. Declared ages, census of 1871

Note: Frequencies of reported ages in the Italian census of population of 31.12.1871.

Italians' preference for terminal digits zero and five ("0/5 ages") means that we can characterize the extent of age-heaping using the Whipple index. This index is defined

as the ratio of the actual share of 0/5 ages to the predicted share in a uniform distribution across terminal digits, i.e. 20 per cent.[3] Multiplying Whipple index values ("W") by 100 for ease of presentation, we have a measure that is 100 when 0/5 ages just equal the expected proportion of one-fifth, and 500 when all ages are 0/5's. Five Whipple points are equivalent to 1 percentage point in the share of 0/5 ages. For example, if the 0/5 share increases from 20 to 21 per cent, W rises from 100 to 105.

Complications arise due to a tendency, found in widely varying contexts, for age-heaping to rise with age itself.[4] To facilitate comparisons of populations with different age structures, we employ a two step procedure, first calculating W separately for each ten-year age-group, then computing their simple average as our overall measure of heaping. Think of this version of W as measuring the age-heaping of the average cohort rather than the average individual. For comparison we occasionally report the alternative in which terminal digit frequencies for all age groups are pooled, which we call a "one-shot" procedure.[5]

Table 1. Age-heaping by age group and gender, census of 1871

age group	women W	women n	men W	men n	total W	total n
13-22	110	2480	105	2442	107	4922
23-32	131	2130	118	2076	125	4206
33-42	154	1747	142	1758	148	3505
43-52	167	1436	155	1467	161	2903
53-62	177	973	157	995	167	1967
63-72	169	565	148	596	158	1161
73-82	178	189	154	204	166	393
average 2-step	159	6850	144	6892	152	13742
average 1-shot	154		140		147	

Notes: W is the age-group and gender specific value of the Whipple index, calculated using national population totals; 100 = no heaping on 0/5 ages. n = number of observations, in thousands. "1-shot" average pools all ages and calculates one W value; "2-step" figure calculates W separately by age-group, then takes the simple average of these figures. Age groups 13-22 and 73-82 excluded from both calculations and from the total n.

The W values reported in Table 1 show clearly how age-heaping is almost absent among young people and rises with age, at least up to the 53-62 group. Also evident is a systematic gender gap indicating less accurate age reporting by women in every age group. Note that for comparability with other studies we restrict attention to the five central cohorts spanning ages 23-72 in calculating the averages reported in the final rows of Table 1.[6] The overall figure of 152 places nineteenth century Italy clearly

---

[3] The Whipple is just one of many possible measures of the deviation from an expected smooth distribution. An obvious disadvantage is the inability to capture any form of heaping that does not take the 0/5 form. On its advantages – chiefly but not only simplicity, see A'Hearn et al. (2009).
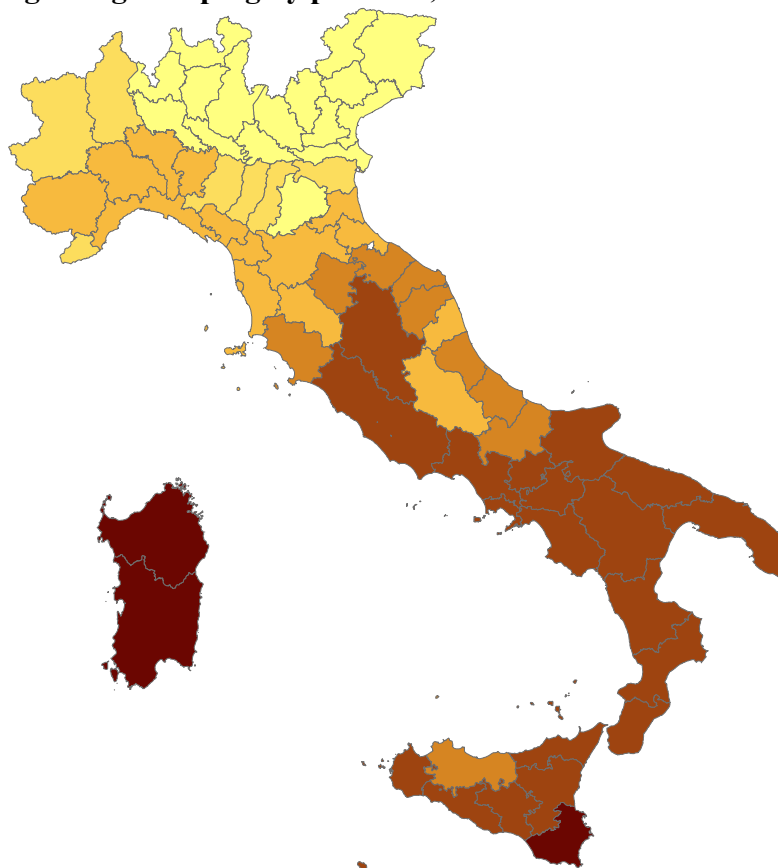
[4] See Baten and Crayen (2010) for a discussion of age-effects on heaping propensity.

[5] On a related problem *within* ten-year age groups see A'Hearn, Baten, and Crayen (2009), who explain the choice of age intervals such as 23-32 rather than the more natural 20-29.

[6] The 13-22 group is excluded because some samples, particularly those from military sources, have a great concentration of individuals at a particular age such as 18, 20, or 21. As for the 73+ range, the concern is with the reliability of the data. In many studies sample sizes in this age range are small.

behind Northern and Central Europe (in the sense of having less accurate age reporting), on a par with other nations of Europe's Southern and Western periphery, and ahead of much of the rest of the world including Eastern Europe, Latin America, and parts of Asia.[7]

Fig. 2. Age-heaping by province, census of 1871

Note: Whipple index class by province, darker colours indicating higher *W* values. Classes are 105-19, 120-31, 132-45, 146-65, 166-202, and 203-30.

Extreme geographic variation in age-heaping is one of the interesting features of the Italian case (Fig. 2). With *W*=105, Belluno province in the Northeast displays only one percentage point of excess 0/5 frequency, a value as low as anywhere in Europe at the time. In Sardinia, at the same time, Sassari province has *W*=230, which would make it an outlier in Europe, if it were a country. In Sassari, at least 26% of individuals reported an incorrect age.[8] We say "at least" 26% reported their age

---

There is also evidence of a tendency for the ages of the elderly to be not merely rounded but also exaggerated.

[7] Comparative international age-heaping estimates can be found in Crayen and Baten (2010), Hippe and Baten (2012), and Baten and Juif (2014). The Table 1 estimate for the 43-52 age group neatly matches Baten and Juif's figure for the 1820s birth cohort (when their ABCC index is converted back to its original Whipple index value). Our figures are somewhat higher than those reported in Baten and Crayen (2010); this likely results from their averaging estimates for the same cohort from later censuses.

[8] Sassari's *W* of 230 is the ratio of 0.46 actual to 0.20 expected 0/5 ages; 26 percentage points is thus the excess frequency of 0/5s.

inaccurately because there is no way of knowing whether ages with other terminal digits are accurate; they are merely "unheaped".

There is a clear North-South gradient in Figure 2, but more subtle patterns can also be discerned. The former Kingdom of Lombardy-Venetia in the Northeast stands out as an area of uniformly low age-heaping, clearly contrasting with its immediate geographic neighbours. Meanwhile though age-heaping is clearly much more prevalent in the South (the former Kingdom of the Two Sicilies) than elsewhere, its northernmost provinces in Abruzzi-Molise would not look at all out of place across the border in Tuscany. And within the former Papal State there is almost as much variation as in Italy overall: from 119 in Bologna to 178 in Perugia.

2.2. *Age-heaping, literacy, and gender*

Is age-heaping a good indicator of the cognitive abilities that we might call numeracy? The census data give us no *direct* evidence on this, but do allow us to examine age-heaping's correlation with an alternative measure of human capital, namely self-reported literacy. Table 2 presents Whipple index values separately for literates and illiterates. *W* is higher among illiterates in every age-group and for both genders. The same regularity is clearly evident in the underlying provincial data: *W* is higher among illiterates in 340 of 345 age-group- and province-specific comparisons.[9] It seems clear that the tendency to report a rounded age is linked to an individual's human capital. At the same time, there are one or two surprises in Table 2. One is non-trivial heaping of ages among *literate* individuals, who, in addition, show the same tendency as illiterates to heap more as their age increases. And the presence of gender gaps among both literates and illiterates, in all age-groups, suggests that age-heaping is about more than just education.

Table 2. Age-heaping by age group and literacy, census of 1871

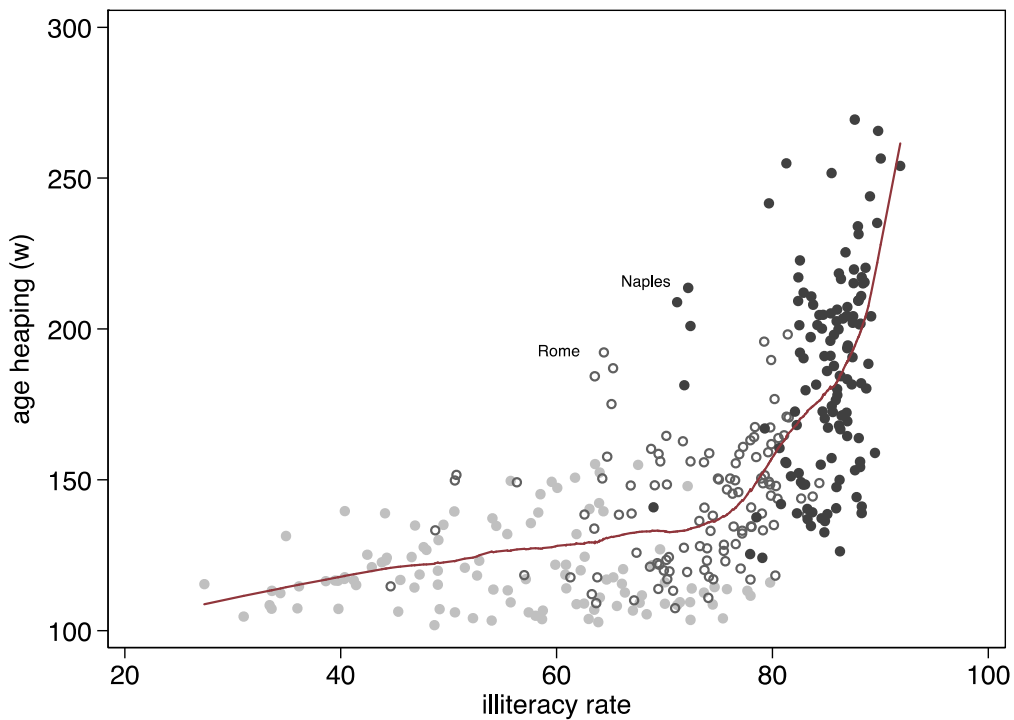|	women		men		total	
age group	literate	illiterate	literate	illiterate	literate	illiterate
13-22	105	113	101	107	103	110
23-32	116	137	107	127	110	132
33-42	129	162	119	156	122	159
43-52	138	174	126	173	130	173
53-62	138	185	125	177	129	182
63-72	131	176	120	165	123	171
73-82	133	185	125	169	127	178
average	130	167	119	160	123	164

Note: *A*ge-group and literacy specific Whipple index values calculated using national population totals. The overall *W* reported in the final row is computed with the two-step procedure, excluding age groups 13-22 and 73-82.

We can learn more by disaggregating, exploiting Italy's remarkable geographic variability. Figure 3 plots the two indicators by age-group and province. An overall positive relationship is immediately obvious. Equally apparent is that this relationship is somewhat noisy, and highly nonlinear. This is not what we would expect based on

---

[9] The number 345 results from 5 age-groups observed in each of 69 provinces.

a simple model in which basic education endows individuals with both literacy and numeracy. There is a mechanical effect at work here, with W running into its practical lower bound of 100 in the North while illiteracy presses up against its ceiling of 100% in the South. Whatever the reason, this nonlinearity makes both prediction and interpretation difficult. Do the southern provinces have just a bit less human capital than the less-educated provinces of the North-Centre, as illiteracy would suggest, or is the disparity dramatic, as age-heaping would indicate? Also catching the eye in Figure 3 are two outliers; the important provinces of Rome and Naples both have W values at least 50 points above what would be predicted based on literacy.

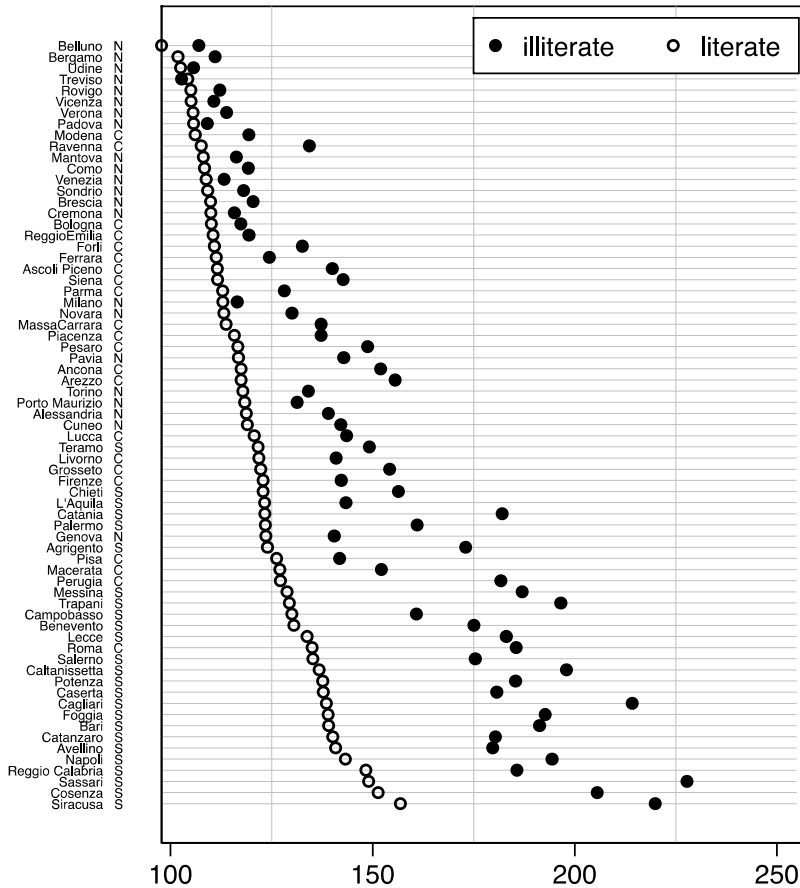**Fig. 3. Age-heaping and illiteracy, by province and age-group**

Notes: Points represent Whipple index and illiteracy pairs for a specific province and age-group. Light grey markers indicate North provinces, hollow markers Centre, dark markers South. Ages 23-72. Line based on locally weighted regression with a bandwidth of 30% of sample.

Average provincial W values for literates and illiterates are plotted in Figure 4. Age-heaping is greater among illiterates every province but Treviso. What is surprising is the variation in W within the two groups. If a few years of school inculcated basic literacy and numeracy, we would expect similar degrees of age-heaping among literates throughout Italy. Instead, W ranges from 100 to 150 among literates. We would expect a higher average, but again limited variation, among illiterates. Instead the range among illiterates is from 105 to 230! Clearly, it is not education that explains these within-group differences. Toward the bottom of Figure 4 there are 23 mostly Southern provinces with W values in excess of 125 among literates. Towards the top there are 19 mostly Northern provinces with rates *below* 125 among *illiterates*. Remarkably, the sparse educational elite of the South in provinces like Napoli,

Reggio Calabria, and Bari reported age less accurately than the unschooled peasants and workers of the North.

Fig. 4. Age-heaping among literates and illiterates

Note: Whipple index values, ages 23-72, by self-reported literacy, census of 1871.

Disaggregating by gender yields further insights. First, the gender gap in age-heaping is barely correlated with the gender gap in illiteracy (Figure 5). Though statistically significant, the correlation coefficient is only +0.16, and the outliers are many. Second, we also find differences by marital status (Table 3). Married women display the same degree of age-heaping as married men, with respective W values of 149 and 147. But as soon as women become widows, their age reporting changes dramatically and W shoots up to 191. (Among widowed men there is no similar change.) This is not due to widows on average being older than married women; we see large gaps between widowed and married women in every age group but the youngest. A plausible explanation could be that proposed by Földvári, van Leeuwen and van Leeuwen-Li (2012), namely that married women (or their husbands when reporting for them) adapted their declared age to that of their spouse. After being widowed, they would have reverted to directly reporting their inaccurately-known age. Blum, Colvin, McAtackney and McLaughlin (2017) find similar patterns in nineteenth century Ireland comparing census data with the self-declared ages of female prisoners.
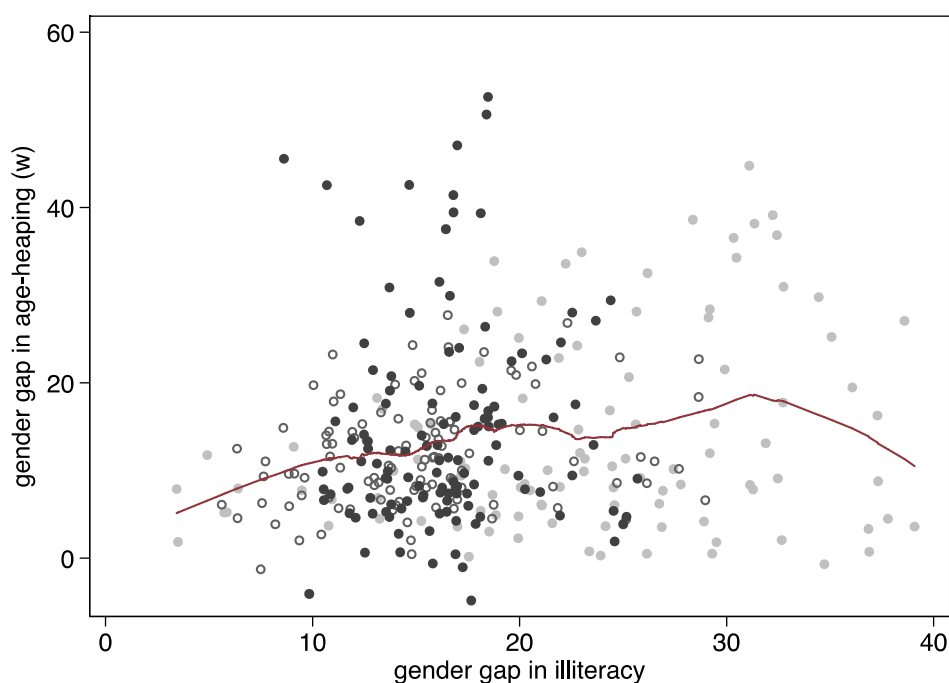
9

An unexplained feature of the Italian data is that widows heaped ages much more than never-married women. Perhaps widows either understated their ages with an eye on the (second-) marriage market or exaggerated them to create an aura of dignity or authority, choosing in both cases round numbers. Whatever the explanation, women's age-heaping patterns clearly reflect more than just their human capital.

Table 3. Age-heaping by age group, marital status, and gender

age group	Women never-married	married	widowed	Men never-married	married	widowed
13-22	110	116	116	105	77	96
23-32	121	135	161	108	131	129
33-42	157	150	190	133	143	145
43-52	168	157	202	146	156	158
53-62	168	157	213	144	158	166
63-72	161	146	189	139	148	152
73-82	160	150	191	137	158	154
average	155	149	191	134	147	150

Note: Marital status, age-group and gender specific Whipple index values calculated using national totals.

Fig. 5. Gender gaps in illiteracy and age-heaping

Notes: Gender gaps defined as female minus male values of *W* and illiteracy. Points represent a specific province and age-group. Light grey markers indicate Northern provinces, hollow markers Centre, and dark markers South. Ages 23-72. Line based on locally weighted regression with a bandwidth of 30% of sample.

2.3. *Age-heaping across five censuses*

A further dimension along which we can explore age-heaping in the Italian censuses is time. We can track the evolution of age-heaping across five censuses from 1861 to 1911.[10] A significant complication in this exercise is that the censuses of 1881 and later report adult age distributions only in terms of five-year age-groups such as 30-34 and 35-39. Though such data do not allow us to compute Whipple values, we can still construct a crude measure of heaping on multiples of ten. This alternative "R" index is based on the ratio of the frequency of the age-group with terminal digits 0-4 ($f_{.0-.4}$) to the frequency of the successive age-group with terminal digits 5-9 ($f_{.5-.9}$), for example (f_{30-34})/(f_{35-39}). Such ratios exceed one if there is significant heaping on multiples of ten (significantly more than on terminal digit 5, that is). Details of the R index's construction are set out in Appendix 1.
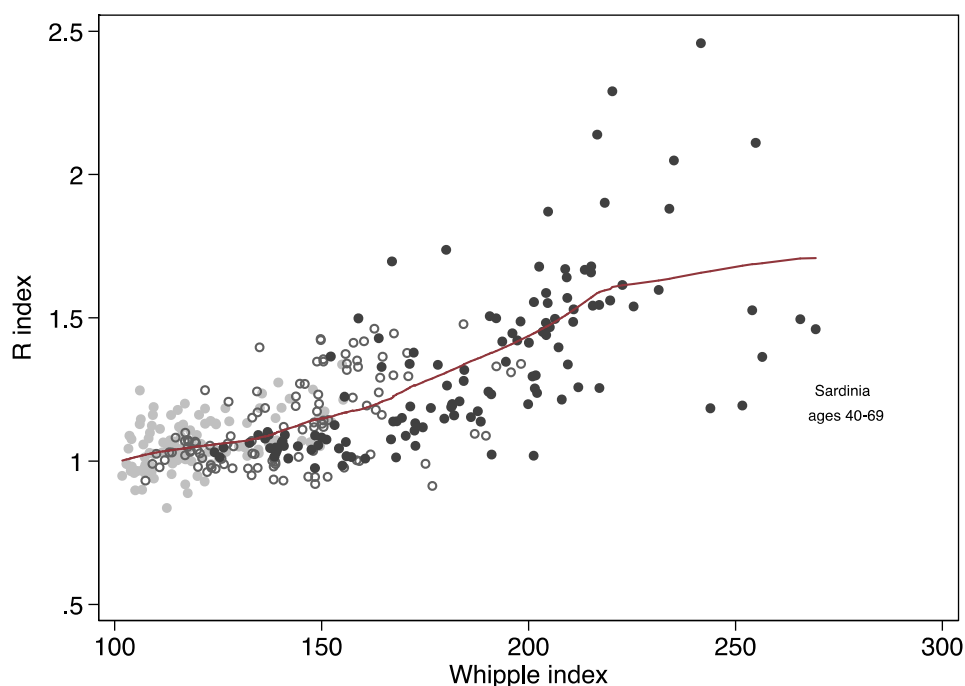
The correlation of R with W across all province–age-group pairs is strong (+0.73) in the census of 1871 (Figure 6). The relationship would be both tighter and more linear were it not for the Sardinian provinces, whose R values for individuals in their 40s, 50s, and 60s are far lower than would be predicted based on W. This illustrates a weakness of R, which fails to fully reflect the extreme age-heaping in Sardinia because terminal digit 0 does not dominate terminal digit 5 on the island as much as elsewhere.[11] Table 4 presents average R values by age interval and macro-area across five censuses.

Considering first the North, we see that overall age-heaping diminishes fairly steadily in each successive census. Tracking individual cohorts through time – moving down along a diagonal in Table 4, that is – we see that the aging effect is at work in a regular way, at least when the initial W value indicates non-trivial age-heaping. The situation in the South and Islands is quite different. Overall age-heaping alternately rises and falls, rather than diminishing steadily, and there is a discontinuous drop between 1881 and 1901, without parallel elsewhere in the table. Across the first three censuses from 1861 to '81, the aging effect is strong and regular in the South, but this changes in 1901. Every cohort that we can track across the 20-year gap shows the same or even *lower* age-heaping two decades later. Afterwards, the usual aging effect appears to be at work again, with R increasing for three of four cohorts between 1901 and 1911. Following the generation in their twenties in 1861, we see R increments of +8 as they reach their thirties, +22 in their forties, *-20* as they age across two decades into their sixties, and +17 as they enter their seventies. 1901 stands out as a complete anomaly. The only plausible explanation is a substantial change in census-taking procedures, or compliance with them, in the South. We return to this issue in Section 4.

---

[10] The missing census in this period is that of 1891, cancelled due to a fiscal crisis (Gallo and Paluzzi, 2012; 38).

[11] Nationwide, a nontrivial number of R values (53/345) are below 100. These are mostly quite close to 100 and reflect our crude index's difficulty in distinguishing the effects of heaping from the effects of mortality (Appendix 1). There are a few more extreme cases, some of which may result from the migration of working-age males distorting provincial age structures.

Fig. 6. Two measures of age-heaping

Notes: Each point represents a province – age-group pair. Light grey markers indicate Northern, hollow markers Centre, and dark markers Southern provinces. Whipple age-groups paired with *R* age-groups following the pattern (23-32, 20-29). Line based on locally weighted regression with a bandwidth of 50% of sample.

3. Micro evidence from Southern death records

Census age-heaping seems to have been influenced by human capital, but also by data collection procedures, social norms, and as yet unidentified local factors. In an effort to sharpen our focus and assess the relative importance of these diverse influences, we turn now to individual-level data. This section reports the results of an exploratory study of microdata from the high age-heaping regions of the Italian South.

3.1. *Death in Naples*

In 1808 Gioacchino Murat, installed as King of Naples by Napoleon, instituted a system of vital record keeping in southern Italy. Information on births, deaths, and marriages was collected by local civil authorities, recorded in duplicate, and retained in archives. The system was maintained by the restored Bourbons throughout the decades leading up to Italian unification. The Neapolitan death records in particular lend themselves to the sort of analysis we wish to conduct. First, they contain both direct declarations of age from the witnesses and indirect reports regarding the deceased. Second, though the dead tend to be either elderly or children (who are not included in the study), the witnesses cover quite a wide range of ages. Third, the format of the records, handwritten but on a single-page printed form, facilitates data collection. Digital images of individual records have been made available on-line via the *Antenati* ("ancestors") website maintained by the *Direzione Generale degli*

Archivi of the *Ministero dei beni e delle attività culturali*.[12] We study 1861, to coincide with the first Italian census.

Table 4. *R* age-heaping values by census and cohort

	1861	1871	1881	1891	1901	1911
North						
20s	109	100	104			107
30s	100	104	98		101	102
40s	112	103	107		98	98
50s	113	115	104		96	97
60s	134	118	119		109	99
70s	136	145	119		123	125
*average**	*119*	*117*	*109*		*105*	*105*
decennial chg.		*-2*	*-8*		*-2*	*-1*
Centre						
20s	95	106	110			103
30s	97	96	103		97	101
40s	123	106	104		104	93
50s	119	132	109		103	105
60s	136	125	135		97	104
70s	146	142	122		118	111
*average**	*124*	*120*	*115*		*104*	*103*
decennial chg.		*-4*	*-5*		*-5.5*	*-1*
South-Islands						
20s	105	106	103			107
30s	106	113	118		100	102
40s	135	121	135		103	98
50s	156	155	143		107	108
60s	178	167	185		115	119
70s	170	175	182		121	132
*average**	*149*	*146*	*153*		*109*	*112*
decennial chg.		*-3*	*+7*		*-22*	*+3*

Note: R index calculated as described in text. Reported figures are simple averages of provincial values. Averages in italicized rows exclude 20-29 age-group for comparability with the census of 1901. Veneto and Lazio regions excluded in all years for comparability with 1861.

It was a legal requirement to declare a death in order to receive permission for burial, and no fee was charged. Two adult males 21 or older were to be found among the relatives or neighbours of the deceased to declare his or her death and play the role of witness. The regulations specified that names, ages, addresses, sex, and occupation of both witnesses and the deceased be recorded, along with information on the parents of the deceased. If able, witnesses were to sign the death certificate, which provides us with a crude proxy for literacy.

Data were drawn from three locations: metropolitan Naples, Bari province, and Reggio Calabria province. These were among the highest census age-heaping provinces in Italy, where even literate individuals reported age less accurately than the uneducated of the north. A total of 4,353 records of unique individuals aged 22 or more were obtained. They refer predominantly to witnesses rather than the deceased, not only because there were two witnesses for each death, but also because a
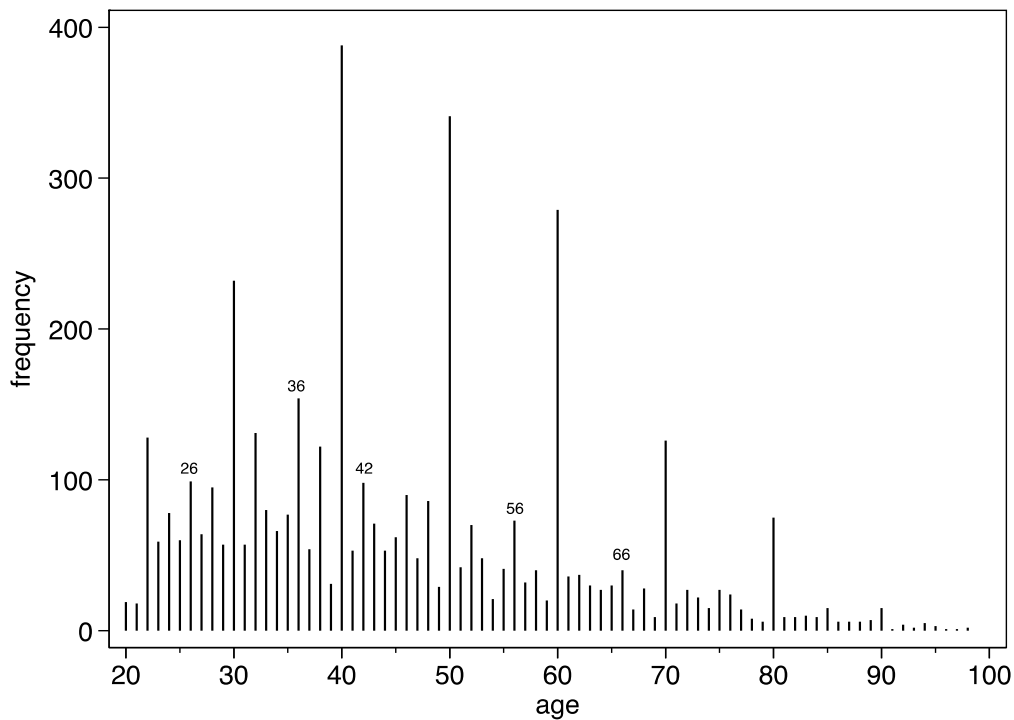
---

[12] www.antenati.san.beni-culturali.it

tragically large share of the dead were infants and young children, while our study requires data on adults. Over 500 occupations were found in the death records, which have been coded into seven categories: unskilled workers (the most numerous group, including most rural occupations), enlisted soldiers, artisans, distinct categories of dependent and autonomous service sector workers, "elite" occupations such as the liberal professions, and a residual group. Further details of dataset construction can be found in Appendix 2.

3.2. *Death record age-heaping*

Figure 7 depicts the age distribution among adults in our sample. Extreme heaping on multiples of ten is immediately obvious; they are 30% of all recorded ages. A more subtle pattern is that, unlike the census data, the second-most-favoured digit is *not* 5. Instead, 6, 8, and 2 are all preferred to 5, a pattern we discuss further in Section 4.[13]

Fig. 7. Age frequencies in Southern death records

Note: age frequencies of witnesses and deceased aged 20 or more; sample of death records in three Southern provinces.

Aggregating towns and neighbourhoods into what we can loosely call districts, we calculate Whipple index values following the procedures described in Section 2 (Table 5). We see the familiar pattern of age-heaping rising with age. The district average W values range widely, from a low of 166 in central Naples to 258 in Bari and its surroundings. They are in general higher than the provincial W values from the census of 1861. It is interesting to note that the periphery of Naples looks more like

---

[13] Also evident is the effect of the restriction that witnesses be "at least 21 years of age", which seems to have been typically interpreted as "22 or more"

distant rural provinces than the nearby city. The *pattern* of heaping also varies. The popularity of 6 evident in Figure 7 derives from Bari and Reggio provinces, not Naples, for example.[14] Because the Whipple index is insensitive to the preference for sixes, eights, and twos, we present an alternative age-heaping index in Table 5: the sum of the absolute deviations from ten per cent of each terminal digit's share. This index can vary between 0 and 180, and is calculated using a one-shot procedure due to small sample sizes. The proportional variation in the alternative index is greater, but the rank order of districts by age-heaping is almost identical.

Tab. 5. Age heaping and signature rates, ages 23-72, by district

	Napoli		Bari			Reggio		All
	centre	periphery	Altamura	Bari	Barletta	Palmi	Reggio	
23-32	119	159	213	148	214	183	197	157
33-42	178	230	225	196	170	241	250	207
43-52	167	214	271	231	208	272	271	226
53-62	199	322	352	220	310	271	272	255
63-72	165	277	229	240	214	235	257	223
73-82	181	346	71	245	250	182	250	244
average	166	240	258	207	223	240	249	214
census	192		192			202		
altern. index	35	51	86	50	63	73	71	52
n	*940*	*622*	*310*	*1,012*	*301*	*397*	*341*	*3,923*
signature	*83*	*61*	*14*	*15*	*10*	*3*	*3*	*34*

Notes: Whipple index values calculated as described in the text; witnesses and deceased combined. "Census" is the *W* value from the 1861 census for the relevant province. "Alternative index" is the sum of the absolute deviations from 10% for each terminal digit's share of all ages. "*n*" refers to ages 23-72. "Signature" is the share of witnesses who signed the death record.

The final row of Table 5 gives the signature rates for our adult male witnesses. They range from an extraordinary 83% in Naples city to an equally extraordinary 3% in the Calabrian districts. There is a general alignment of low age heaping with high signature rates, but the relationship is certainly not monotonic. Naples' immediate hinterland stands out with both high signature rates and high age-heaping. Altamura has easily the greatest age-heaping, but its signature rate is the median.

3.3. *Individual age reporting*

Heaping on multiples of 10 in our sample is extreme. Since the majority of these observations must be incorrect, predicting which individuals report "heaped" ages is tantamount to identifying who did not report their age accurately. We do this by means of logit regressions in which the dependent variable is a dummy for ages ending in 0 or 5 (where 5 is included for comparability with other studies although it is not popular in our sample) and the explanatory variables are age-group, sex, occupation, geographic region, and signature ability. Table 6 presents our estimates of the marginal effects of these regressors in several specifications. The reference group is an illiterate, unskilled, male witness aged 22-31, from Bari or Barletta district.

---

[14] At the level of individual towns, we find the share of terminal digit 6 to be as high as 30% or more of all ages in five cases, as low as 0% in three others.

Tab. 6. Marginal effects on the probability of reporting a 0/5 age

	all		witnesses		signature		town fe	
unskilled	*reference*							
artisan	-0.070	***	-0.076	***	-0.042	*	-0.053	**
enlisted	-0.077		-0.095		-0.051		-0.044	
other/miss.	-0.087	*	-0.055		-0.021		-0.035	
elite	-0.113	***	-0.121	***	-0.053		-0.061	**
service dep.	-0.049		-0.042		-0.004		-0.014	
service indep.	-0.065	***	-0.061	**	-0.021		-0.030	
fem. craft	0.074							
fem. elite	-0.176	***						
fem. missing	0.015							
fem. serv.	0.277							
fem. unsk.	0.021							
age 22-31	*reference*							
age 32-41	0.076	***	0.049	**	0.044	*	0.031	
age 42-51	0.111	***	0.090	***	0.089	***	0.082	***
age 52-61	0.156	***	0.115	***	0.109	***	0.118	***
age 62-71	0.099	***	0.048		0.046		0.077	*
age 72+	0.066	*	-0.193	***	-0.207	***	-0.165	**
Bari-Barletta dist.	*reference*							
Naples city	-0.047	**	-0.037		0.020			
other rural	0.073	***	0.082	***	0.094	***		
deceased	0.010							
signature					-0.117	***	-0.062	**
N	4,353		3,140		3,140		3,140	

Note: estimated marginal effects, evaluated at zero for all variables, from logit regressions predicting an age ending in 0 or 5. Stars indicate statistical significance at 1, 5, and 10%. The χ^2_{88} test statistic for the null that the town fixed effects in Col. 4 are jointly zero is 225.6, p-value 0.0000.

Columns 1 and 2 report estimates of a baseline model without signature ability, referring to the full sample of 4,353 and the narrower subset of witnesses, respectively. In both cases we observe the same pattern of age, geographic, and occupational effects. Age-effects follow a hump-shaped pattern with the highest probability of reporting a "heaped" age in the 52-61 age range. Geographic effects are also strong and significant. Relative to the reference district of Bari-Barletta, the inhabitants of other agricultural districts, including peripheral Naples, are 7 percentage points more likely to report a 0/5 age. Meanwhile, residents of Naples city are four and half points *less* likely to report such ages. These effects are meaningfully large compared to the variation across districts in raw probabilities of reporting a heaped age: 31% in Naples, 40% in Bari-Barletta, and 47% in the other rural districts. Turning to occupational effects, all groups have markedly lower probabilities of reporting a heaped age than unskilled labourers. The effects are large and mostly statistically significant: 5-6 percentage points for service workers, 7 points for artisans, 11 points for the socio-economic elite. Gender differences can be read from the female occupational effects, which indicate that, outside the social elite, women were more likely to have 0/5 ages reported for them (recall that all women in the sample are deceased) than the lowest-skilled men.

The models in Columns 3 and 4 add signature ability to the regressors. Signature ability itself has a powerful and statistically significant effect, lowering the

16

probability of a heaped age by 12 percentage points in Column 3. It also lowers the size of several other effects. Occupational effect estimates are diminished by about half, and the Naples effect reverses sign. On the other hand, the regional effect of agricultural districts other than Bari-Barletta strengthens to more than 9 percentage points. Whatever is going on in these rural areas, it is not explained (or not *fully* explained) by either occupational structure or literacy. And artisans remain a statistically significant 4 points less likely to report a heaped age than unskilled workers, even after controlling for signature ability. For the socio-economic elite, the effect is 5 points (the associated p-value is 0.104). Column 4 adds town fixed effects to the model, treating the neighbourhoods of greater Naples as separate towns. The town effects themselves are large and highly significant. Their inclusion reduces the predictive power of signature ability by half in Column 4, and causes most occupational effects to strengthen. Those for artisans and the elite are statistically significant at 5%.

Table 7 presents estimates of the "signature" model of Table 6 for three sub-samples. In the agricultural regions, signature ability has a powerful and statistically significant effect, reducing the probability of reporting a 0/5 age by roughly 16 percentage points. Occupational effects are somewhat inconsistent and never statistically significant. Naples city looks completely different. Here signature ability has much less predictive power and is not statistically significant. It is instead occupational effects that are strong in Naples. For craftsmen and the elite, the effect is statistically significant and, at negative 15-16 points, as powerful as is signature ability in agricultural districts. The same patterns remain when town/neighbourhood effects are added to the model. (The local fixed effects themselves are jointly highly significant except within Naples city.) Where signature ability was super-abundant, in central Naples, it has little predictive power; occupation and neighbourhood matter more. Where it was very scarce, in the provinces, signature ability is the most reliable predictor.

4. Of numbers and numeracy, birthdays and bureaucracy

Several of our findings sit uncomfortably with the idea of age-heaping as a measure of numeracy. In this section we present evidence supporting the alternative argument that the correlation of age-heaping and illiteracy in Italy arose from common roots in a process of administrative and cultural modernisation.

4.1. *The State*

Census officials were well aware of age-heaping – and not very happy about it. It was an instantly recognizable symptom of a general weakness in census procedures, and a specific problem for the development of mortality tables.[15] One response was to fine-tune census questions and exhort local census workers to greater efforts. The logistics of the census involved distribution of forms to each family, to be filled out by the head of household for later collection. In practice, as the head of household was often illiterate, it was frequently another household member or neighbour who wrote out

---

[15] Numerous efforts were made to circumvent the problem by means of interpolation, e.g. Italy (1885). Useful overviews of the history of nineteenth century Italian censuses are Gallo and Palluzzi (2012), Mastroluca and Verrascina (2012) and Favero (2012).

the family's responses, a potential source of inaccuracy right from the start. In the event that a family failed to complete the form properly, census workers were to call back , interview the family, and record answers in their own hand. The instructions of 1881 specifically directed local enumerators to check reported birth year and age for consistency. In 1901 instructions to ensure age data were recorded with "scrupulous precision" were reiterated. Moreover, the census question was changed with the specific goal of eliciting more accurate responses; age was no longer asked about at all, only the month and year of birth.[16] Only as a last resort should age in years be recorded, from which enumerators were to calculate birth year, which they were further instructed to check against local birth records.

Tab. 7. Marginal effects on probability of reporting a 0/5 age

	Naples		Bari-Barletta		Other agricult.	
artisan	-0.146	*	-0.043		-0.009	
enlisted	-0.149				0.079	
other/miss.	-0.090		-0.044		-0.001	
elite	-0.161	*	-0.014		-0.034	
service dep.	-0.084		0.136		-0.100	
service indep.	-0.078		0.043		-0.057	
fem. craft						
fem. elite						
fem. missing						
fem. serv.						
fem. unsk.						
age 32-41	0.167	***	-0.025		0.028	
age 42-51	0.143	**	0.034		0.095	**
age 52-61	0.172	**	0.016		0.150	***
age 62-71	0.032		-0.058		0.154	**
age 72+	0.126				-0.233	**
signature	-0.045		-0.167	***	-0.158	***
N	722		1,105		1,303	

Note: the table presents estimated marginal effects, evaluated at zero for all variables, from logit regressions predicting an age ending in 0 or 5. Stars indicate statistical significance at 1, 5, and 10%.

A second response was an attempt to centralize data processing. In the censuses of 1861 and 1871, it was local government (the *comune*) that distributed census forms to each household, collected completed forms, checked them for errors and omissions, remedied these, transcribed from the family form a set of individual records, tabulated from the individual records the distribution of ages (and other desired statistics), and forwarded the results up the administrative chain of command to the *Direzione generale della statistica* in Rome. To carry out these tasks *comuni* were to engage temporary clerks and rely on the unremunerated voluntary efforts of local notables, professionals, schoolteachers, and government employees. Importantly, no funding was provided by the central government to finance these operations.

The results were highly unsatisfactory from the point of view of officials in Rome. In the preparations for the census of 1881, Luigi Bodio, head of the Statistical Directorate, complained about age distributions in earlier censuses specifically.

---

[16] Census of 1901, Volume 5, *Relazione sul metodo,* pp. LVIII, 193/4, 196.

"I could cite the names of numerous *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 appeared greater than that of the middle-aged. If the municipal employees charged with carrying out that enumeration had invented with artfulness, perhaps none of us could have detected the falsification." (Favero 2001; 138).

In the census of 1901 *individual* (rather than *family*) records were filled out by household heads and forwarded by the *comune* directly to Rome.[17] In this way, the operations of information processing *and their costs* were centralized in a dedicated government office for the first time (Gallo and Paluzzi, 2012; 38-39).

It was not just local government but also the cooperation of local populations, notables and common people alike, that determined the success of data collection. Writing about the upcoming census of 1871, Bodio worried that the situation was not propitious with the clergy alienated by the recent annexation of Rome, and because of

"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 of the state. (…) With the census, it is a question of fighting a great battle: a battle against ignorance, against the devices and mechanisms of concealment of taxpayers…" (Favero 2001; 94).

It was hoped that the statistics of large cities might be more trustworthy due to the greater average education of urban populations. In the census of 1881, these hopes were disappointed, at least in age reporting.[18] What is interesting is the interpretation about education by census officials. They associated it not with cognitive skills, but with acceptance of, even identification with, the nation-building purposes of the state's intrusion into private lives (Favero 2001; 98).

We call the state's ability to monitor activity, gather and process information, implement policy, and elicit the cooperation of local institutions and individuals "administrative capacity". The best explanation for the discontinuous improvement in census age-heaping in 1901 is an improvement in the Italian state's administrative capacity.[19] And this varied as much geographically as temporally. The prefects, representatives of the central government tasked with monitoring and controlling the

---

[17] 1901 was also the first census in which mechanical data processing and calculation aids (Hollerith machines) were employed (Gallo and Paluzzi, 2012; 39 and Favero, 2012;132-133).

[18] In the census of 1881, separate age-distributions were published for the 69 provincial capitals. These had the single-year format of earlier censuses, which was otherwise discontinued and would disappear altogether in 1901. A comparison of the capital cities with the relevant provincial totals is possible on the basis of the R index of Section 2. This shows that the expected urban effect on age-heaping was significant in some provinces, particularly in Sicily and Sardinia, but on average small, and not infrequently of the "wrong" sign.

[19] Crayen and Baten (2010; 90) find that variables likely to correlate with administrative capacity (state antiquity, the number of censuses previously conducted, and school enrollments) lower age-heaping in a large, international sample.

actions of local government in the provinces, encountered difficulties everywhere, but especially in the South. Here it proved difficult to induce many *comuni* to comply with basic legal requirements such as keeping proper accounts and publishing budgets, to say nothing of effectively providing public services such as infrastructure, health, and education (Randeraad 1993; ch 6). At the other extreme are cases from the north in the late 1860s, in which towns in recently-annexed Veneto requested permission to carry out *ad hoc* local censuses as a foundation for the continually updated population registers required under Italian administrative law but not actually implemented by many towns elsewhere in Italy (Favero 2001; 57).

Persistent local variation in the competence of peripheral officials and their commitment to the priorities of the central administration derived from historical traditions in the pre-unification states; the same was true of the attitudes toward the state among local citizens. Tracing the history of public statistical inquiries, Patriarca (1996) finds some common features in the efforts of statesmen and scholars throughout pre-unification Italy: Enlightenment-era traditions of political arithmetic, a common experience of Napoleonic administrative bureaucracy, and retention of much of this apparatus by restoration governments. [20] "(E)very state … had kept statistics and had its own heroes" (Randeraad 2010; 129). Yet there were also real differences.

The primacy of Habsburg Lombardy-Venetia was generally acknowledged. The most prominent statisticians of the early nineteenth century, Melchiorre Gioia and Gian Domenico Romagnosi, were both active in Milan, as was Pietro Maestri, author of the first *Annuario statistico italiano* and later the founding head of Italy's Statistical Directorate. And it was here that the greatest number of statistical works were published per capita (Table 8). In terms of practical results, the land register (*catasto*) of Lombardy was considered a triumph, both at the time and in retrospect. Completed in 1760, the *catasto teresiano* (after Empress Maria-Teresa) was based on a comprehensive and detailed mapping of landholdings conducted by Dutch surveyors under the supervision of imperial officials from other provinces. This sort of survey was, by a margin of decades, an absolute first in Italy. The Austrian administration also collected data on population, crime, and the economy. Originally reserved for internal use, by 1846 they were being regularly published in "monumental compilations in strictly tabular form" (Patriarca 1996, p. 119).

Regional case studies reveal the difficulties experienced by statistical offices and commissions. Local authorities were reluctant to provide information, while higher authorities were at times reluctant to divulge it, or to indulge the liberal or Italian-nationalist sympathies of their statisticians. Nonetheless important studies were undertaken and at least some statistics regularly published in Lombardy-Venetia, Piedmont, and Tuscany. Things were different in the South. In Sicily the formal institutional set-up appeared ideal. The restored Bourbons assigned statistical assistants to their intendants (i.e. prefects) in separatist-minded Sicily, and in 1832 created a *Direzione generale di statistica* in Palermo, staffed by an extraordinarily ambitious and active group of individuals including Francesco Ferrara, later one of Italy's most famous economists. But the actual results were meagre indeed. The Directorate's statisticians encountered lack of cooperation or outright resistance at

---

[20] These two paragraphs are largely based on Patriarca's *Numbers and Nationhood*, particularly Chapter 4, "Official Numbers".

every level, from the intendants, from local authorities, from the social elite, and from the people. Without authority over intendants and mayors, their efforts were doomed (Patriarca 1996; 88-95).

Table 8. Indicators of administrative capacity, 1850s

	population	*revenue*	*publishing*	*cadastre*	*age-heaping*
Piedmont-Sardinia	4,600	100	100	Mixed	133
Lombardy-Venetia	5,600	83	102	Survey	115
Parma	500	59		Survey	134
Modena	600	63		Mixed	126
Tuscany	1,800	63	64	Survey	146
Papal State	3,100	70	49	Survey	153
Two Sicilies	9,300	55	41	descriptive	181

Notes: Publishing – number of statistical works published 1850-59 per million population, relative to Piedmont's 8.0. 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. Cadastre – basis of the land register as described in text. Age-heaping – Whipple index; *W* of Piedmont-Sardinia for Piedmont and Liguria provinces only; Sardinian *W*=224. *Sources*: population – Correnti and Maestri (1858, p. 381); revenue – Romani (1982, p. 404); publishing – Patriarca (1996, p.241); cadastre – Zamagni (1993, p. 73); age-heaping – our calculations.

There is evidence of something similar in the death records analysed in Section 3, where the extent and pattern of heaping varied from one town to the next. Ferdinando Maria Jusseppi was mayor of three small towns in Reggio Calabria province (district of Palmi) in 1861. Fully 63 per cent of recorded ages in "his" towns were 0/5 ages. In addition, every one of 13 deceased women had the occupation *filatrice* (spinster), regardless of age, marital status, or other circumstance. Similarly, of 51 males, both witnesses and deceased, 48 had the occupation *bracciale* (labourer). Spinster and labourer seem to have been automatic defaults. Meanwhile, the signature rate among 47 witnesses was identically zero – even for the lone individual identified as a property owner. Minimal information was recorded about the deceased and their relatives. One has the impression that Jusseppi, his deputy, and the clerk (both of whom shared his surname) did not invest much effort in accurate gathering and recording of data.

Turning to the indicators of administrative capacity in Table 8, we see that the Bourbon kingdom was last among the major pre-unification states by statistical publications, while its *catasto* in the 1850s was the only one still based exclusively on the verbal declarations of landowners rather than a survey valuation. A crude summary measure of administrative capacity is total tax revenue per capita. It is crude both because the structure or prosperity of some economies allows for high tax revenues even in the absence of sophisticated administrative capabilities, and because taxes are a policy choice. Still, the existence of a North-South gradient, with revenues per capita distinctly lowest in the Kingdom of the Two Sicilies, is hard to overlook.

To be sure, our evidence is flawed and incomplete, but it all points in the direction of lower administrative capacity in the parts of Italy where age-heaping was greatest.[21]

4.2. *The People*

There are both a priori reasons and a few scraps of direct evidence to suggest that most nineteenth century Italians, north and south, possessed rudimentary mathematical knowledge. Numbers were everywhere in the world of both peasants and city dwellers, even in their games. In the ubiquitous game of *morra*, once a popular motif for painters of common life in Italy, the object is to quickly guess a sum: the total number of fingers extended simultaneously by the players.[22] Meanwhile, the state *lotto*, or lottery, ensured that from the late seventeenth century onward all Neapolitans knew their numbers from 1 to 90 (Macry 1997).[23] A passion for the *lotto* extended from the Bourbon King himself to the denizens of the city's poorest slums. The poor made micro-wagers and experienced micro-gains and -losses on a regular basis. The 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 understanding of written numbers. So too did the popular books illustrating 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 twenty-two; to call a woman a seventy-eight was something worse (Serao 1885; 57). In this period, 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, chosen numbers, and associated payoffs (Serao 1885; 64).

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 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

---

[21] Recent contributions in political science adopt a similar approach, taking age-heaping in *contemporary* national censuses as a suitable proxy for state capacity, see Lee and Zhang (2016) and Driscoll (2016).

[22] *Morra* is similar to the game *rock, paper, scissors* widely played in the English-speaking world. The game, which can involve more than two players, was very popular in Naples in the first half of the nineteenth century according to Rocco (1857). De Sivo (1865) goes so far as to claim *morra* as the root of word *camorra*, i.e. the Neapolitan mafia. On the game's popularity in the southern region of Abruzzo today, see Spitilli (2012).

[23] 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 a single number appearing (ca. 12:1) and highest for a triple (as high as 5000:1, depending on the period). In the years leading up to 1861, *lotto* drawings took place weekly.

interpretation here, though it is worth noting that playing numbers suggested by dreams or recent events was *not* irrational given the structure of *lotto* payoffs.[24] Our argument is only that the popularity of the game suggests the ability to count and to recognize and name numbers was widespread.

And of course 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. To make this essentially evolutionary argument is not to claim that the average Italian landless labourer understood compound interest, nor to deny modern findings that poverty-induced stress impairs cognition (Mani et al. 2013). It is merely to assert that there were strong incentives for a family to acquire a minimal competence in numerical reasoning.[25]

The tools required for basic numerical competence were there to be acquired, though not necessarily in forms familiar to us today. Methods of counting to 100 and beyond on one's fingers, though no longer set out in Italian math textbooks in the nineteenth century as they had been earlier, remained common in practice (Roggero 1994). And traditional currency and measurement systems were designed to facilitate quick mental calculations based on halving and doubling.[26] In Italy as elsewhere, the monetary system had been based on 12 and 20, numbers twice-divisible by two: 12 *denari* per *soldo*, and 20 *soldi* per *lira*.[27] Twelve was also the basis of various physical measures: in southern Italy 1 *palmo* (.26m) was 12 *once*; 1 *botte* (524l) was 12 *barili*; and 1 *libbra* (.32kg) was 12 *once*. Attempts to introduce decimalization beginning in the Napoleonic period were resisted. As late as 1888 the school inspector for the *circondario* of Lagonegro (south of Naples, in Basilicata), wrote that teaching in mathematics was producing the most satisfactory results of all disciplines, perhaps because the subject suited the "speculative and calculating nature" of locals. Practical instruction in using the metric system, however, was being held back by the persistence of ancient Neapolitan measures in daily usage, he claimed (Italy 1888). Part of the explanation for this persistence is likely to be reliance on established routines for mental calculations. Traditional measurement systems may also explain

---

[24] The *lotto* did not pay out a fixed total shared among all holders of winning tickets. Had that been the case, it would have been wise to avoid popular numbers chosen by other gamblers, given that the probabilities of all numbers being extracted were equal. Instead, the *lotto* paid a *fixed* multiple of the amount wagered to *every* winning player (making a big loss when popular numbers were extracted from the urn).

[25] Tollnek and Baten (2012) make a similar argument for farmers, who not only had the incentive to develop numeracy skills, but were able to pass these on to their children, who benefited from an abundant diet that relaxed nutritional constraints on cognitive development. A student of medieval Tuscany's fiscal records is impressed by the "bottom-up" numeracy of ordinary 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, 2002; 670).

[26] The so-called "Russian peasant multiplication" method reduces *any* multiplication problem to a series of halving, doubling, and summation operations.

[27] In the Kingdom of Naples 12 *cavalli* per *grano* and 20 *grana* per *tarì* had played equivalent roles in the monetary system, though there were other denominations in circulation such as the *carlino* of 10 grana and the *piastra* of 12 *carlini*.

some details of the heaping patterns in reported ages, such as the preference for sixes, i.e. half-twelves (Thomas 1987). A similar holdover was the use of twenties in oral statements of age, e.g. "three-twenties" for sixty, especially in the south (Menninger 1977, 49; Price 1991, 465).

Pen and paper were not common in nineteenth century Italy, nor the years of practice required to achieve proficiency with quill and ink, but recording numbers temporarily by means of traditional systems of marks made with a piece of charcoal or a sharp object was common. So too was use of the abacus in the nineteenth century (Roggero 1994).[28] For record-keeping, recourse was had to tally sticks and their like. 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 *code civil*, and the system was still common in Puglia in the twentieth century. Illiterate shepherds there not only recorded commercial transactions in this way, but numbered their sheep analogously, following rules for recording units, tens, hundreds and thousands by making incisions into the upper and lower lobes of both ears (La Sorsa, 1953).

We know from modern studies that children learn to count readily, largely figuring it out for themselves before starting school. They even devise strategies for basic addition and subtraction without any instruction, such as breaking 7+8 into 7+(3+5) and then rearranging this into the two easier sums (7+3)=10 and 10+5=15 (Dehaene 1997, Ch. 5). Children of the late twentieth century whose occupations keep them out of school but involve them in monetary transactions develop their own "street math": non-standard methods for solving more complicated problems. In a classic study, unschooled Brazilian street vendors aged 10-12 performed better on several mathematical tasks than their counterparts in urban schools (Saxe 1988).[29] A few decades earlier, the Soviet developmental psychologist Luria had studied cognition among rural Uzbeks in 1931-32. His interviews vividly revealed an inability or unwillingness to think abstractly and outside the range of personal experience among the uneducated members of this population. Some 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 problems answering such questions were caused by a lack of basic numeracy or the inability to think abstractly, he posed the same question in different form, e.g. "divide thirty cookies equally among five men". Even for illiterate, never-schooled individuals, he found, "simple computations used in everyday practical affairs presented no special difficulties" (Luria 1976; 126). 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, and the claim for numeracy made here is only that the typical Italian had the cognitive capacity required to comprehend, to state, to record, and to recognize the numbers relevant for human age.

---

[28] Here "abacus" refers to placing tokens on lines drawn on paper or in the sand, or incised in a board, rather than moving beads along wires in a frame. Roggero (1994) notes that by 1800 abacus methods had mostly disappeared from textbooks and schools, but must have remained in common use.

[29] See also Carraher, Carraher and Schlieman (1985) for another influential study of the numerical abilities of working children in Recife.

Contemporaries took Italy's high rates of illiteracy very seriously. It was a source of national embarrassment. Though well-aware of age-heaping, as we have seen, they did not deplore it as further evidence of the population's tragically underdeveloped cognitive abilities. They reported instead, in a matter-of-fact way, that it resulted from ignorance of precise age. "The relative maxima … at the ages indicated by numbers which end with 0 or 5, are due to the ignorance or the negligence of the individuals being surveyed, who in large part do not remember their age exactly, but when asked about it round the numbers," we read in one report (Italy 1885b; 87). Inaccuracies, according to the general report on the census of 1881, derived from rounding, deliberate understatement (among women), exaggeration (among the elderly), and inconsistent understandings of whether current age meant years completed or the year currently underway. In general, people simply did not think of age in terms of single years, "relatively small units of measure not used in practice by many of those interviewed" (Italy 1885a; xxxix).

Today, at least in rich countries, birthdays and bureaucracy offer constant reminders of our age and birth year. In nineteenth century Italy, especially in the South where age-heaping was so prevalent, neither played an important role. The absence of the state from much of daily life – no compulsory schooling, no voting rights, no regular policing, no old age pensions, no taxation of income – limited the average person's interactions with the administrative apparatus, and meant that proof of identity was never required, files of individual data never compiled. When interaction with government officials was required, as in the case of vital records declarations, the authorities were not able (or not interested) to detect inaccuracy (Section 3). The state generated no incentive to know one's age accurately.

What of birthdays? There is no evidence that they were celebrated in ordinary Italian families in this period. A modern invention, or rediscovery, the celebration of birthdays has been linked with changing perceptions of time by Jean-Claude Schmitt (2007). A staple of sociological theorizing since late nineteenth century pioneers such as Durkheim and Simmel, the argument is that the marking of time is a socially embedded process that serves the function of coordinating group activities (Sorokin and Merton 1937). In medieval Europe, the Catholic and Orthodox churches' calendars of recurrent religious feasts dovetailed with agricultural rhythms underlying social life. In the West, the rise of Protestantism weakened the cult of saints and the general orientation of life around the liturgical calendar, while urbanization and industrialization cut links with natural cycles, created coordination problems, and focused attention on the value of time (Schmitt 2007, Thompson 1967). As Kaiser and Engel (1993; 836) 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". 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.

With regard to birthdays specifically, the medieval Church had opposed their celebration as a pagan ritual. It was either baptism or death – either marking the start of one's true life in God, that mattered. 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 maturity and eventually senescence. In the modern, linear conception of time, the progressive accumulation of

age, or "capitalization of years", acquired greater meaning too (Schmitt 2007). Individualism and, more speculatively, astrology may also have had roles to play. Despite such changes, Schmitt finds scant evidence of the celebration of birthdays even among the rich and well-born, even among individuals who definitely knew their exact birth date and year, before the late eighteenth century.[30] Birthday celebrations spread among the aristocracy and bourgeoisie only in the nineteenth century, and did not reach the working classes before the twentieth (Schmitt 2012, 73-92).

Table 9. Frequency of birthday words in published work in three languages

	1700	*1750*	*1800*	*1850*	*1900*	*1950*	*2000*
*compleanno**	0	0	1.0	1.5	8.3	14.1	39.0
Geburtstag	0	6.4	12.8	19.7	43.3	122.8	218.8
birthday	2.4	3.1	8.9	31.7	77.5	98.9	103.5

Notes: relative frequency of the phrases "compleanno" and "anniversario della nascita", "Geburtstag", and "birthday" in the Google Books Italian, German, and British English corpora; 21-year moving averages; case-insensitive searches using Google's N-gram Viewer on 27/12/16; raw values multiplied by 100,000 for ease of interpretation; "Geburtstag" value for 1700, determined by a single instance of the word in 1692, is suppressed.

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. Among books published circa 1850, "*Geburtstag*" occurs more than ten times as frequently as "*compleanno*" and "*anniversario della nascita*" together in the relevant German and Italian collections of digitized historical texts. "Birthday" occurs more than twenty times as often in the British English collection (Table 9).[31] The trends are rising in all three languages, but latest and most weakly in Italian. And in southern Italy, at least in Naples, it was celebration of the saint's day (*onomastico*) that remained the norm 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" (Ferrante 2011, 221; Galasso 2012). Saint's days did not coincide with birthdays ordinarily; in the case of popular given names they were shared with countless others and fell on religious holidays. The life of the individual and his or her age were less in focus than in a birthday celebration. To the extent that celebration of individual age achieved on birthdays was a modern cultural practice diffusing from north to south, from rich to poor, and from city to country in the nineteenth century, we have a natural explanation for occupational and geographic effects on age-heaping.

---

[30] Goethe is a well-known exception, whose birthdays became something of a public event as his celebrity grew. A well-known case which is more typical is Pepys, who always noted his birthday in his diary but never mentions any celebration of it.

[31] The evidence is drawn from Google's "Ngram Viewer". Two sources of uncertainty or slippage mean this evidence must be taken with a grain of salt. The first is that common cultural practices need not be reflected in published books. It is easy to imagine publishing being dominated by novels in one country and by religious books in another. The second is uncertainty about what works are included in Google's language-specific collections of digitized books. The Italian and German collections are much smaller than the English, with the result that a single instance of a birthday word can have a substantial impact on the relative frequency series in the early years.

An interesting parallel 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 (1992) explanation is based on the practice of *naming* years after the familiar animals of the Chinese zodiac, making birth year an indelible part of individual identity. 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).[32] 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 Uygur ethnic minority, which alongside smaller Kirgiz and Tajik groups maintained a clear cultural and linguistic separateness from the Han Chinese. Among the Uygur, "birthdays are not particularly important days and they are not accorded celebrations and special observance" (Jowett and Li, 1992; 440). Jowett and Li (1992; 427) do not infer that these groups are innumerate, but rather that they "neither know, nor care, how old they are."

5. Conclusions

The promise of insights into the distribution of human capital in the past has prompted a profusion of age-heaping studies.[33] The evidence of nineteenth century Italy suggests it may be misleading to interpret age-heaping as a direct measure of numerical cognitive skills, however.

Italian census data exhibit the same broad correlation of illiteracy and age-heaping that has been found in many other studies. But they also throw up a number of anomalies that are not so easily reconciled with a cognitive ability interpretation: i) age-heaping varies quite widely even among *literates*, as well as among illiterates; ii) age-heaping varies by gender in a way that is poorly correlated with literacy and clearly tied up with marital status; iii) age-heaping varies with age in a way inconsistent with a straightforward human capital interpretation (a common finding treated as a mere nuisance in the literature); and iv) age-heaping varies across censuses and regions in a way that can only be explained in terms of the efficacy of data gathering procedures. An exploratory study of micro-data from southern Italy shows much the same: witnesses who did not sign death records were more likely to report 0/5 ages, but even controlling for this measure of literacy other factors, such as occupation and place, had a significant role – in some cases more important role – in predicting inaccurate age reports.

Parallels in analogous contexts, together with shreds of direct evidence gleaned from Italian social history, cast doubt on the assumption that most Italians, even the large

---

[32] The twelve-year cycle of zodiac animals ("earthly branches") interacted with a ten year-cycle of "heavenly stem" names, creating a 60-year set of unique birth-year identifiers. Within years, 24 "solar seasons" named after prevailing weather conditions were distinguished. "Thus, someone born at the time of the white dew in a geng shen [monkey] year, has a date of birth in early September 1920 or 1980" (Jowett and Li, 1982; 429, "[monkey]" added).

[33] By our count there have been at least fifteen publications on age-heaping since 2009, in addition to a number of working papers.

pool of illiterates, lacked basic familiarity with numbers and the most elementary arithmetical operations. More plausible than ignorance of numbers is simple ignorance of age among Italians, or a reluctance to make the effort to report it accurately. And there is evidence to support both hypotheses. Age-awareness was limited in much of Italy by pre-modern perceptions of time and attitudes regarding the celebration of birthdays. And evidence on the capacity of pre-unification Italian states to gather and process information, eliciting the cooperation of their citizens, corresponds well to regional variation in age-heaping. Writing nearly forty years ago, Herlihy and Klapisch-Zuber (1985, orig. 1978; 164) argued for an interaction between popular culture and the state's demand for information. From 1371 onwards "the people of Tuscany, including its humblest members, underwent an apprenticeship in measuring time… Through the Catasto of 1427, we can glimpse the means by which the commune's authority encouraged this cultural revolution, and we can discern the kinds of conscious or unconscious resistance which opposed the official demand for accuracy."

Cultural change in the direction of secularism, individualism, and linear time-perception, and institutional change in the direction of administrative capacity building by the state are both aspects of what used to be called modernisation.[34] The spread of literacy too forms part of this bloc of interrelated social, economic, cultural and political changes. Age-heaping, on this Italian-inspired reading, correlates with illiteracy because both are reflections of slow and incomplete modernisation. Just as its advocates have contended, age-heaping is therefore a reasonable *proxy* for education, literacy, and perhaps numeracy. What it *directly* measures, we argue, is not numeracy but a broader mix of contextual factors that are just as interesting for diagnosing and predicting economic development.

---

[34] We use the term "modernisation" in a descriptive way, to characterize a multidimensional, mutually reinforcing process of change, without endorsing any specific variant of sociological theory. For a useful recent take on modernisation, firmly locating its origins in economic change, see Inglehart and Welzel (2005).

References

A'Hearn, Brian, Jörg Baten, and Dorothee Crayen, "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital," *Journal of Economic History*, vol. 69 no. 3 (2009), pp. 783-808.

Baten, Jörg, Dorothee Crayen, and Hans-Joachim Voth, "Numeracy and the Impact of High Food Prices in Industrializing Britain, 1780-1850" *Review of Economics and Statistics*, vol. 96 no. 3 (2014), pp. 418-30.

Baten, Jörg and Dácil Juif, "A story of large landowners and math skills: Inequality and human capital formation in long-run development, 1820–2000", *Journal of Comparative Economics*, vol. 42 (2014), pp. 375-401.

Baten, Jörg and Stefan Priwitzer, "Social and intertemporal differences in basic numeracy in Pannonia (first century BCE to third century CE)", *Scandinavian Economic History Review*, vol. 63 no. 2 (2015), pp. 110-34.

Blum, Matthias, Christopher Colvin, Laura McAtackney and Eoin McLaughlin, "Women of an uncertain age: quantifying human capital accumulation in rural Ireland in the nineteenth century", *Economic History Review,* vol. 70 no. 1 (2017), pp. 187-223.

Cafagna, Luciano, "Modernizzazione attiva e modernizzazione passiva", *Meridiana*, vol. 2 (1988), pp. 229-40.

Carraher, Teresinha N., Carraher, David W. and Analucia D. Schlieman, "Mathematics in the streets and in the schools", *British Journal of Development Psychology*, Vol 3 (1985), pp. 21-29.

Cipolla, Carlo M.. *Literacy and Development in the West*. Harmondsworth: Penguin, 1969.

Crayen, Dorothee and Jörg Baten, "Global trends in numeracy 1820–1949 and its implications for long-term growth," *Explorations in Economic History*, vol. 47 (2010), pp. 82-99.

Dehaene, Stanislas. *The Number Sense. How the Mind Creates Mathematics*. Oxford: OUP, 1997.

de Moor, Tine and Jan Luiten van Zanden, "'Every woman counts': a gender-analysis of numeracy in the Low Countries in the early modern period", *Journal of Interdisciplinary History*, vol. 61 (2010), pp. 179-208.

De Sivo, Giacinto, *Storia delle Due Sicilie dal 1847 al 1861*. Verona, 1865.

Driscoll, Jesse, "Seeing like a subaltern: State-building, Census-taking and the political economy of population counts", mimeo (2016), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2825378

Emigh, Rebecca, "Numeracy or enumeration: the use of numbers by states and societies", *Social Science History,* vol. 26, no. 4 (2002), pp. 653-698.

Favero, Giovanni. *Le misure del regno: Direzione di statistica e municipi nell'Italia liberale*. Padova: Il Poligrafo, 2001.

Favero, Giovanni, "Il dibattito sull'organizzazione dei primi censimenti unitari negli atti della Giunta e del Consiglio Superiore di Statistica", *Annali di Statistica*, Series XII, Vol. 2 (2012), pp. 121-145.

Felice, Emanuele and Michelangelo Vasta, "Passive modernisation? The new Human Development Index and its components in Italy's regions (1871-2007)", *European Review of Economic History,* Vol. 19 (2015), pp. 44-66.

Ferrante, Elena. *L'amica geniale*. Rome: Edizioni e/o, 2011.

Fischer, David Hackett. *Growing Old in America.* Oxford: Oxford University Press, 1977.

Földvári, Peter, Bas van Leeuwen, and Jieli van Leeuwen-Li, "How did women count? A note on gender-specific age heaping differences in the sixteenth to nineteenth centuries," *Economic History Review*, vol. 65 no. 1 (2012), pp. 304-13.

Galasso, Giuseppe, "E il compleanno uccise l'onomastico," *Corriere della Sera*, 19 March, 2012.

Gallo, Gerardo and Evelina Palazzi, "Le trasformazioni del censimento della popolazione in Italia: i cambiamenti della macchina organizzativa censuaria negli ultimi 150 anni", *Annali di Statistica*, Series XII, Vol. 2 (2012), pp. 31-75.

Hanushek, Eric and Ludger Woessman, "The Role of Cognitive Skills in Development", *Journal of Economic Literature,* vol. 46 (2008), pp. 607-68.

Hippe, Ralph and Jörg Baten, "Regional inequality in human capital formation in Europe, 1790–1880" *Scandinavian Economic History Review*, vol. 60 no. 3 (Nov. 2012), pp. 254-89.

Herlihy, David and Christiane Klapisch-Zuber. *Tuscans and Their Families: A Study of the Florentine Catasto of 1427*. New Haven, CT: Yale University Press, 1985 (orig. publ. 1978).

Inglehart, Ronald and Christian Welzel. *Modernisation, Cultural Change and Democracy. The Human Development Sequence.* Cambridge: Cambridge University Press, 2005.

Italy. Ministero della Pubblica Istruzione. *Relazione annuale del regio ispettore scolastico, circondario di Lagonegro*. Archivio centrale dello Stato, Roma. MPI, Direzione generale dell'istruzione primaria e popolare 1860-96, busta 299.

… . Ministero dell'Agricoltura e dell'Industria. Direzione generale della Statistica. *Censimento della popolazione del Regno d'Italia al 31 dicembre 1881. Relazione generale e confronti internazionali*. Roma: Tipografia Eredi Botta, 1885a.

… . Ministero dell'Agricoltura e dell'Industria. Direzione generale della Statistica. *Sulla classificazione per età della popolazione del regno secondo il censimento 31 dicembre 1881.* (*Estratto dagli Annali di statistica, serie 3a, vol. 14o*). Rome: Tipografia dei Fratelli Bencini, 1885b.

Jowett, John and Yuan-Qing Li, "Age-Heaping: Contrasting Patterns from China," *GeoJournal*, vol. 28 no. 4 (1992), pp. 427-42.

Juif, Dácil and Jörg Baten, "On the Human Capital of the "Inca" Indios Before and After the Spanish Conquest. Was there a "Precolonial Legacy"?', *Explorations in Economic History,* vol. 50 (2013), pp. 227-41.

Kaiser, David and Peyton Engel, "Time- and Age-Awareness in Early Modern Russia," *Comparative Studies in Society and History*, vol. 35 no. 4 (1993), pp. 824-39.

La Sorsa, Saverio, "Storia e folclore della mena delle pecore in Puglia," *Archivio Storico Pugliese*, vol. 6 (1953), pp. 487-99.

Lee, Melissa M. and Nan Zhang, "Legibility and the Informational Foundations of State Capacity", *Journal of Politics*, Vol. 79, no. 1 (2016), forthcoming.

Luria, Alexander. *Cognitive Development. Its Cultural and Social Foundations*. Cambridge, Mass: Harvard University Press, 1976.

Macry, Paolo. *Giocare la vita. Storia del lotto a Napoli tra Sette e Ottocento.* Roma: Donzelli, 1997.

Manzel, Kerstin, Jörg Baten, and Yvonne Stolz, "Convergence and divergence in numeracy: the development of age-heaping in Latin America from the seventeenth to the twentieth century", *Economic History Review,* vol. 65 no. 3 (2012), pp. 932-60.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao, "Poverty Impedes Cognitive Function," *Science*, vol. 341 (2013), pp. 976-80.

Mastroluca, Simona and Mariangela Verrascina, "L'evoluzione dei contenuti informativi dei censimenti della popolazione", *Annali di Statistica*, Series XII, Vol. 2 (2012), pp. 77-119.

Menninger, Karl. *Number Words and Number Symbols: A Cultural History of Numbers*. Cambridge, Mass.: MIT Press, 1977 (first German edition 1934).

Mokyr, Joel. *Why Ireland Starved: A Quantitative and Analytical History of the Irish Economy, 1800-1850,* London: Allen & Unwin, 1983.

Patriarca, Silvana. *Numbers and Nationhood. Writing Statistics in Nineteenth Century Italy*. Cambridge: Cambridge University Press, 1996.

Sorokin, Pitirim and Robert Merton, "Social Time: A Methodological and Functional Analysis," *American Journal of Sociology*, vol. 42 no. 5 (1937), pp. 615-29.

Patriarca, Silvana. *Numbers and Nationhood. Writing Statistics in Nineteenth-Century Italy*. Cambridge: Cambridge University Press, 1996.

Price, Glanville, "Romance," Ch. 13 in *Indo-European Numerals*, Jadranka Gvozdanovic, ed. (Berlin: Mouton De Gruyter, 1991), pp. 447-96.

Randeraad, Nico. *Authority in Search of Liberty. The Prefects in Liberal Italy*. Amsterdam: Thesis Publishers, 1993.

Randeraad, Nico. *States and Statistics in the Nineteenth Century. Europe by Numbers*. Manchester: Manchester University Press, 2010.

Rocco, Emanuele. "Il gioco della mora" in Francesco de Bourcard (ed.), *Usi e Costumi di Napoli e contorni,* vol 1. Napoli, 1857.

Roggero, Marina, "Conti sulle dita, calcoli a penna. L'aritmetica elementare a fine Settecento," *Studi Storici*, vol. 35 no. 4 (Oct.-Dec. 1994), pp. 1039-60.

Romani, Mario. *Storia economica dell'Italia nel secolo XIX, 1815-1882*, Sergio Zaninelli, ed. Bologna: Il Mulino, 1982.

Saxe, Geoffrey, "The Mathematics of Child Street Vendors," *Child Development*, vol. 59 (1988), pp. 1415-25.

Schmitt, Jean-Claude. *L'invenzione del compleanno*. Bari: Laterza, 2012.

… , "L'invention de l'anniversaire," *Annales. Histoire, Science Sociales*, vol. 62 no. 4 (2007), pp. 793-835.

Serao, Matilde. *Il ventre di Napoli*. Milano: Fratelli Treves, 1884.

Spitilli, Gianfranco. "Di passata in passata. Il gioco della morra nell'Abruzzo settentrionale," in *Giochi, sport tradizionali e società. Viaggio tra la Valle d'Aosta, l'Italia e l'Unione Europea*, Anna Maria Pioletti, ed. Aosta: Musumeci Editori, 2012.

Thomas, Keith, "Numeracy in Early Modern England: The Prothero Lecture," *Transactions of the Royal Historical Society*, series 5, vol. 37 (1987), pp. 103-32.

Thompson, Edward P., "Time, Work-Discipline, and Industrial Capitalism," *Past and Present*, vol. 38 no. 1 (1967), pp. 56-97.

Tollnek, Franziska and Jörg Baten, "Farmer Families at the Heart of the Educational Revolution: Which Occupational Group Inherited Human Capital in the Early Modern Era?" *EHES Working Paper* no. 33, 2012.

… , "Age-Heaping based human capital estimates," in Diebolt, Claude and Michael Haupert (eds.), *Handbook of Cliometrics* (Heidelberg: Springer-Verlag, 2016).

Vecchi, Giovanni (ed.). *In Ricchezza e in Povertà. Il benessere degli Italiani dall'Unità ad oggi.* Bologna: Il Mulino, 2011.

Whipple, George Chandler. *Vital Statistics. An Introduction to the Science of Demography.* London: John Wiley, 1919.

Zamagni, Vera. *Dalla periferia al centro. La seconda rinascita economica dell'Italia (1861-1990)*, 2e. Bologna: Il Mulino, 1993.
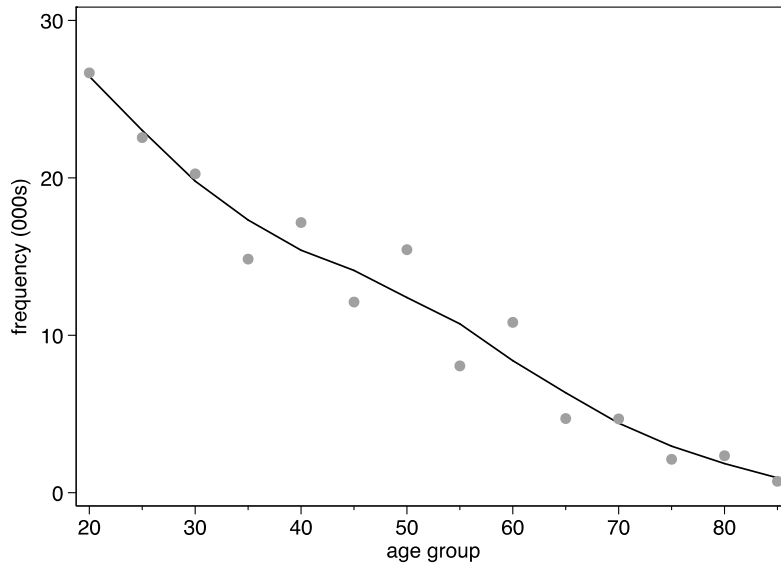
Zanden, Jan Luiten van, ''The skill premium and the 'Great Divergence','' *European Review of Economic History*, vol. 13 (2009), pp. 121-53.

Appendix 1. *Computing the R index.*

The *R* index (*R* for "ratio") is designed specifically for use with age frequency data defined on five-year age-groups beginning with multiple of five ages, e.g. 20-24, 25-29 and 30-34. We denote these frequencies as $f_{20\text{-}24}, f_{25\text{-}29}, f_{30\text{-}34}$, and so on. The censuses of 1881, 1901, and 1911 report these (and only these) directly; for 1861 and 1871 we sum the single-year frequencies. *R* is a measure of heaping on multiples of ten. When there is significant heaping on ages with terminal digit 0 (importantly, significantly *more* than on terminal digit 5), then in any successive pair such as ($f_{20\text{-}24}$, $f_{25\text{-}29}$) in which the first group contains a multiple of ten, their ratio will exceed one. As with the Whipple index, we do not pool the data but employ a two-step procedure, first calculating the index for each ten-year interval, then averaging across intervals.

Because ratios like ($f_{20\text{-}24}/f_{25\text{-}29}$) are *always* greater than one when population declines with age, even in the absence of heaping, an adjustment is required. We make the adjustment at the level of the individual province. The first step is to generate predicted frequencies by smoothing the actual distribution using a locally-weighted regression of frequency on age. Figure A1 illustrates using the example of Agrigento in 1871. We call these predictions $\hat{f}$.

Fig. A1. Actual and predicted age-group frequencies, Agrigento

Note: grey markers indicate observed frequencies for each five-year age-roup. The line is the smoothed distribution resulting from locally weighted regression, with bandwidth at 50% of sample.

Next we compute R_a^r ratios (the superscript *r* indicating "raw" and the subscript *a* referring to the "tens" digit of the ten-year age interval) using both the actual and predicted *f* values:

$$R_a^r = f_{a0\text{-}a4}/f_{a5\text{-}a9} \text{ and } \hat{R}_a^r = \hat{f}_{a0\text{-}a4}/\hat{f}_{a5\text{-}a9}.$$

Our "adjusted" index for a given ten-year interval is the ratio of actual to predicted:

$$R_a = R_a^r/\hat{R}_a^r.$$

Finally, we average these R_a values across all age-intervals from 20-29 to 70-79, i.e. for $a = 2\ldots7$ and multiply by 100:

$$R = 100 \cdot \frac{1}{6}\sum_{a=2}^{7} R_a.$$

Table A1 sets out the elements of the calculation for Agrigento province in 1871. $R =$ 140 for this province. The interpretation of this value is that 0-4 ages predominate over 5-9 ages by 40% more than we would expect, given the age structure of the province's population. Note how R heaping rises with age in Agrigento, up to a point, just as Whipple index heaping does.

Table A1. *R* ratios for Agrigento, 1871

age group	observed R_a^r	predicted $\hat{R}_a^r$	Ratio R_a
20-29	1.18	1.15	103
30-39	1.36	1.14	119
40-49	1.42	1.09	130
50-59	1.92	1.16	166
60-69	2.30	1.32	174
70-79	2.21	1.49	148
average	1.73	1.23	**140**

Note: R values calculated as described in the text. The average reported in the final row is the simple mean across all age groups.

Appendix 2. *Construction of the death record dataset*

Data were drawn from twelve neighbourhoods of greater Naples (Avvocata, Chiaia, Fuorigrotta, Miano e Marianella, Montecalvario, Posillipo, S. Carlo all'Arena, S. Ferdinando, S. Giuseppe, S. Lorenzo, Stella, and Vomero) and 8l towns ("*comuni*") in the rural provinces of Bari and Reggio. In the provinces, an informal sampling strategy was adopted based on the alphabetical ordering of towns and aimed at broad coverage. Towns with poor quality digital images were not sampled, but those with poor handwriting were included (at considerable effort!). The most difficult handwriting was typically not sloppy, but rather over-stylized in the service of elegance.

Our sampling procedure is neither truly random nor stratified. Even if it were, we would face the difficulty that witnesses are not a random selection of adult males. For this exploratory study, the aim was to cover, at low cost, the widest possible range of circumstances and individuals, rather than assemble a representative sample. In each town or neighbourhood, data collection proceeded chronologically starting with 1 January 1861 and continued until i) a reasonable number (a few tens) of useful observations was obtained, ii) the end of calendar year 1861 was reached, or iii) strongly diminishing returns set in due to repetition of the same witnesses. Some towns were "oversampled" when they yielded a wide range of occupations, suggesting a degree of socioeconomic differentiation.

Among witnesses, it was common for the same individuals to appear repeatedly in the records, creating a problem of duplication. Identifying perfect matches on surname, first name or initial, age, occupation, and locality is straightforward, but throws up false matches because the range of occupations, names, and surnames was narrow in small towns. We assumed all matches were valid and retained only one of the duplicate records for analysis. The same rule was applied to partial matches in which all variables except age differed. It is likely that many of these cases represent the same individual reporting different ages on different occasions. This would make these cases particularly interesting to study, if we could be certain they were repeat observations on the same individual. Because we cannot, we retain only the first instance of such partially-matched observations. Table A2 gives the number of observations available.

Table A2. Observations in the micro database

province	records	persons	unique	age 22+
Napoli	1,203	3,609	2,392	1,806
Bari	857	2,571	2,236	1,754
Reggio	336	1,008	919	793
Total	2,396	7,188	5,547	4,353

We rely on witness signatures as a measure of literacy. Figure A2 illustrates the signatures on the death record of six-year old Elvira Romano, who died in the Avvocata neighbourhood of Naples on 7 January, 1861. At the bottom is the stylized, elegant signature of Luigi Pagano, the local councillor in charge of vital records. At the top is the name, neatly printed, of Vincenzo Pallacicino, a 41 year old white collar municipal employee. Interestingly, this unusual surname is spelled differently by Pagano (or his clerk) elsewhere in the document, where it is rendered as Pallavicino, a more common family name. Finally, in the middle is the childlike, only partly legible effort of the 60 year old Antonio Castaldo, whose occupation, incredibly, is "*scrivente*", or scribe.

Fig. A2. The death record of Elvira Romano, 7.01.1861, detail