

Martinez, Marco

Working Paper

Dimensions of illiteracy: A quantitative and comparative approach from Italy, circa 1815

LEM Working Paper Series, No. 2022/28

Provided in Cooperation with:

Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies

Suggested Citation: Martinez, Marco (2022) : Dimensions of illiteracy: A quantitative and comparative approach from Italy, circa 1815, LEM Working Paper Series, No. 2022/28, Scuola Superiore Sant'Anna, Laboratory of Economics and Management (LEM), Pisa

This Version is available at:

<https://hdl.handle.net/10419/273630>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

INSTITUTE
OF ECONOMICS



Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

Dimensions of Illiteracy: A Quantitative and Comparative Approach from Italy, circa 1815

Marco Martinez ^a

^a Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy.

2022/28

October 2022

ISSN(OBJECT) 2284-0400

Dimensions of Illiteracy: A Quantitative and Comparative Approach from Italy, circa 1815

Marco Martinez

Institute of Economics, Sant'Anna School of Advanced Studies, Pisa, Italy

Piazza Martiri della Libertà 33 56126 Pisa (PI), email: m.martinez@santannapisa.it.

This paper quantitatively assesses to what extent signatures in marriage certificates can inform about literacy rates in pre-industrial states. The direct estimates are based on a novel and balanced random sample of marriage certificates for pre-unification Italy in 1815. Such figures are compared to all alternative sources available for close years, including direct and indirect approaches who focused on selected areas of Italy. The new empirical methodology is as important as the results found with it. Two main findings emerge. First, marriage certificates can accurately measure literacy rates in pre-industrial Italy, but only when accompanied by rigorous sampling procedures. Indeed, the proposed empirical approach allows to go from local to aggregate estimates that generally are in line with other estimates for the period, and can be scaled up or applied in other contexts. Second, North vs Centre-South divides in 1815 are lower than previously suggested. This supports the hypothesis that French reforms and later restoration governments triggered a process of widening North-South literacy divides.

Keywords: literacy rates, random sampling, marriage certificates, human capital

JEL codes: N01, N33, J24

Acknowledgments

I am grateful to my supervisor Alessandro Nuvolari for the constant encouragement and insightful discussions. This paper benefitted from the careful comments of Carlo Ciccarelli (discussant), Gabriele Cappelli (discussant), Giuliana Freschi, Martina Occelli (discussant), Giovanni Vecchi (discussant), Lorenzo Del Panta, and Leigh Shaw-Taylor, as well as from the helpful suggestions of David Chilosì, Giovanni Federico, Jacob Weisdorf, Alessandro Pietropaoli, Michele Postigliola, Mauro Rota, Michelangelo Vasta and of the participants at the Spring 2020 Sant'Anna Quarantine Seminar, Fall 2020 Sant'Anna Workshop, 2021 Economic History Society Conference, 2021 Lost Highways Seminar (LHSem), and the 3rd Sapienza Economic History Workshop. I am also grateful to the staff of the libraries of Scuola Normale Superiore, Scuola Sant'Anna, to the libraries of Economics and of History and Philosophy of the Università di Pisa and of the state archives of Florence, Naples, and Palermo for their support.

1. Introduction

Literacy rates are the most widely adopted measure of average human capital and primary education outcomes by economists, economic historians, demographers and education historians. Measuring the ability to read and write becomes challenging in historical settings where censuses did not collect information on literacy, such as ancient and medieval societies, eighteenth-century European countries, and less developed countries in the twentieth-century.

For at least the past century, historians of modern continental Europe and its colonies have made use of signatures to infer literacy rates. Some of the earliest attempts to use signatures to infer literacy rates include Sargant (1867) and Bruce (1910), but a notable example is also Cipolla (1969). 50 years ago, Roger Schofield published his study *Dimensions of Illiteracy, 1750-1850* in 1973. In his study, Schofield (1973) derived a random sample of marriage registers from 274 parishes and analyzed the fall of illiteracy that occurred in the entire England over a century. This publication of Schofield marked a turning point in the already established measurement of literacy rates through signatures because he first attempted to scale up the micro-level estimates derived from marriage signatures to the level of England. The article keeps being cited today, often to justify the use of marriage signatures as a proxy for literacy rates and human capital (e.g., Meier zu Zeilhausen, Van Leeuwen, and Weisdorf 2018; Álvarez and Palencia 2018). Indeed, Schofield (1968, 1973) also defended signature literacy as a good proxy for literacy because standard – a single measure “from one person to the next, one group to the next, and from an historical period to the next” and direct – not relying on asking an opinion about the ability to read or write.¹

¹ Schofield was not the only one to argue in favor of signature-literacy at the time. See Rachal (1987) for a discussion and other examples of signature literacy. Pellizzari (2000, 17-26) and Houston (2014, 1–37) offer a complete review of the various generations of studies using signatures to infer literacy rates.

Yet, other authors also raised concerns about the use of signatures as a proxy for literacy for four main reasons. First, the social and political corpora of individuals that were selected to sign for special occasions often does not represent the entire society (think of wills, court depositions, jury inquests and deeds). This is more of a problem for some documents than for others, and marriage certificates are not exempt from this issue as not the entire population decided to marry, possibly in a selective way – think of the clergy and the mendicants, or individuals of different religious creeds. For instance, Astoul (1992, 182) raised concerns belonging for measuring literacy rates with marriage signatures (*le methode “Maggiolo”*) in the French *Midi*. Poussou (1993, 214-215) also admitted that locally, the results can be erroneous and called for a check of the methods by collecting more signatures in specific areas. Second, the question of whether a sample of signatures represents the macroscopic population of regions or countries is hard to assess empirically when no other sources are available to compare the resulting literacy rate with, despite all efforts that can be put in creating a statistically random sample. Third, women are not well represented in signatures of public documents because they were less involved in public affairs than men, although marriage certificates are an exception to this.² Fourth, signature literacy is a crude proxy for the ability to read and write well, because many people who do not know how to write complicated sentences could still know how to sign. This criticism is often just mentioned, but Quiroga (2003) also introduced a measure of signature quality to account for the fact that signatures with poor handwriting may measure lower levels of literacy rates.

But neglecting to measure literacy rates on these grounds would leave historians ill-equipped to study important questions such as the causes and consequences of early-stage development processes and related path-dependency in later educational and social outcomes. The lack of a comparison benchmark for literacy rates in pre-industrial settings, such as

² Main (1991) is an early attempt to deal with gender related issues in measuring literacy rates.

censuses, and the inherent difficulty in collecting often handwritten data from marriage registers scattered over a country, have discouraged the use of marriage certificates to derive direct, country-wide estimates of pre-industrial literacy rates from marriage signatures. Resultingly, signatures in marriage certificates are mostly used to study educational levels in small and selected areas of countries. However, this skepticism with signature-based methods to infer literacy rates has not yet received solid empirical groundings.³

In this paper we ask a relatively simple question: should we trust literacy rate estimates derived from micro-level signatures to infer macro-level literacy rates? This is an interdisciplinary question as it has broad implications for quantitative demographers, historians of education, economic, and social historians.

The Italian context is a unique case study to empirically test the value of marriage certificates to learn about pre-industrial literacy rates for the following reasons. First, in the early nineteenth-century is possible to compare direct estimates of literacy rates from marriage signatures with more indirect but comprehensive sources such as census data from unified Italy, whose back-casted estimates can be traced back at most to 1821 (e.g., Ciccarelli and Weisdorf 2019). Second, marriage certificates are available across almost the entire unified country, and the ease of access allows systematic and rigorous sampling procedures. The certificates are heterogeneous in format as they were collected in different preunitary states.⁴ Third, Italy in 1815 was still a relatively backward country. It thus represents better other pre-industrial contexts where it would be useful to use marriage signatures to infer literacy rates because no alternative data is available than other countries in the same period. Fourth, the Italian case is historically relevant on its own, because learning more about literacy rates across the country

³ This skepticism is generally implicit in the studies. Cremin (1970, 526-540) is an example of explicit criticism.

⁴ See Lemerrier and Trivellato (2022) for a similar context with Italian notarial records.

before the unification informs about the one of the longest-standing debates in the Italian economic and social history, that on the origins of the North-South divides.⁵

I offer a systematic account of literacy rates from all the sources that, to my knowledge, are available to comprehensively study literacy rates in pre-unitary Italy, including novel ones. To do so, I draw from the toolbox of a wide set of disciplines, including demography, statistics, and economic history. The main data source is a novel and representative cross-sectional sample of 1,315 individual-level marriage certificate signatures dating back to around 1815. I randomly select villages with marriage certificates in the Antenati.it online genealogical archives following Schofield (1973), Rossi, Toniolo and Vecchi (2001), and Shaw-Taylor and Wrigley (2014). The sample (henceforth "Antenati" sample) is constructed to have representative rates at the level of Italian provinces, regions and macro-areas (North, Centre, and South). I account for data under-coverage issues by constructing sampling weights and using regressions following Clark (2005) and Federico et al. (2019). This allows to safely divide the literacy rates of each area into three main occupational classes. From marriage certificates I also collect information about the ability to sign of spouses but also about their occupations, and about parents' occupation. Then I construct a novel data-set containing all the available signatures, rather than a sample of it, for a large area of the provinces of Udine and Salerno,

⁵The economic historiography has provided convincing evidence that Italy was already divided at least in two at the time of the Unification, measuring economic performance and living standards through real wages (Federico, Nuvolari, and Vasta 2019, contra: Daniele and Malanima 2017), life expectancies and heights (A'Hearn, Auria, and Vecchi 2011), social capital and culture (Guiso, Sapienza and Zingales 2016; Putnam, Leonardi and Nanetti 1994), GDP (Felice 2014, contra: Malanima and Daniele 2007; Daniele and Malanima 2011), Human Development Index (Felice and Vasta 2015), or income inequality (Felice 2019). The social capital explanation that dates back North-South divides to as far as the Middle Ages has received alternative (Cappelli 2017b) and complementary explanations (A'Hearn 1998; Mariella 2022). These explanations focus on the post-unitary determinants of persistent divides.

and collect all possible information from eighteenth-century sources (*Processetti* and *Documenti Matrimoniali*) for the city of Naples and the province of Florence. I also collect literacy rates from studies who focused on specific parts of Italy in and around 1815 (also referred to as "local studies"). Finally, I consider unitary censuses and related literature.⁶

I find that the after following a stratification and regression approach, a sample of signatures across Italy reconstructs well the literacy patterns that were also documented by local studies and indirect approaches in the period of interest. The rich information of marriage certificates also enables to look at new dimensions of illiteracy. For instance, the urban-rural gap was 11% and the gender gap of 22% for an average male literacy of 31%. The North-South gap in female literacy rates was 2.2, lower than previously estimated (2.7) for 1821. The North-South gap was particularly high among farmers as the literacy rate of Northern Italian farmers was similar to the average literacy in Centre and Southern Italy. The North-South divide in literacy was of 12%, less marked than previously found.

This study has implications for two main streams of research. The first is the historical debate on the origins of the Italian North-South divides. The findings of a smaller North-South gap than previously estimated, although preliminary, allow new speculations on the origins of North-South literacy divides, and to attribute a crucial role to the early nineteenth century

⁶ Local studies include Scirocco (1987), Toscani (1987), Piseri (2002), Toscani (1993), Brambilla (1991), Milanesi (1991), and Ferraresi (1991). I do not compare my study only with studies that either focused on later periods (study only with studies that either focused on later periods (i.e., Griseri (1973) for Piedmont in 1831-1856), earlier periods (i.e., Pellizzari 1987 for mid eighteenth-century Kingdom of Naples) or on areas not collected in this sample as Sardinia (Pruneri 2011). Considered official statistics are the estimates of Ciccarelli and Weisdorf (2019) for up the period 1821-1851, and the 1861, 1871, and 1881 unitary censuses: MAIC (1866a, 388-411), MAIC (1875, 4-239), MAIC (1883, 217-564), alongside other proxies of literacy after the unification derived from Ministero dell'Istruzione (1890).

Napoleonic reforms rather than to the persistence of earlier institutions.⁷ Indeed, the Napoleonic schooling reforms oriented the Italian primary educational system around public and universalistic principles, but in practice decentralized public funding. Consequently, only wealthy municipalities, more frequent in Northern Italy, could fund primary schools. Indeed, "municipalities, besides few bequests and tributes, only disposed of the variable revenue streams of taxation" and "it could happen that the local administrations, often expression of notables whose interests diverged from the ones imagined by the central government, did not have any will to increase taxation in order to obtain the resources to be distributed as a public service" (own translation from Lupo [2005, 70-74]). This, combined with a particularly conservative restoration of monarchic powers in the Bourbonic South after the congress of Vienna of 1815, blocked the development of literacy rates in the South but not in the rest of Italy.

This paper also speaks to scholars who wish to learn about literacy levels and education supply at early-stages of the industrialization process in other countries. This study provides an operative methodology to do so, scaling up the local studies' estimates with a bottom-up approach to safely learn about the many dimensions of literacy rates in wider areas. It also corroborates the empirical validity of such an exercise with a particularly rich comparison with other sources.

The remainder of the paper proceeds as follows. Section 2 places the reader in the context of the history of education of Italy over the period 1770-1861. Section 3 describes the sources. Section 4 describes the research design and discusses its validity. Section 5 presents the main results about literacy rates and places them in context. Section 6 concludes.

⁷ This hypothesis is in line with a 'French influence' argument that underlies some recent work (Dincecco and Federico 2021; Postigliola and Rota 2020; Postigliola and Rota 2021).

2. The Evolution of Literacy, 1770-1861: Historical Background

This section covers a brief history of the educational policies enacted in Italy from the 1770 to 1861 and will be useful for the reader to adequately place the historical results of this paper in context (see Piseri [2017, 21-60] for an excellent introduction). Until the 1770s, the absolutist governments of the many Italian states did not conceive education as a state affair. Most schools were privately founded where the demand was highest. Primary schools were demanded especially by small landowners, manufacturers and the ones in need of skills to emigrate (Toscani 1993). Teaching was individual rather than classroom-based and teachers were ecclesiastical (Genovesi 2010). The so-called enlightened despots gradually became aware of the importance of economic growth for political legitimacy and military power. Even catholic rulers as Maria Theresa of Habsburg-Teschen and her government first reduced the feudal rights (i.e., through taxation), and ecclesiastical privileges (i.e., in education provision). Relatedly, the members of the Company of Jesus were expelled from various European States.

Jesuits had been central providers of higher education in Italy since the foundation of the first Jesuit College in Messina in 1604. In many states, the Jesuits continued to be influential in the intellectual life also after the expulsion, as in the reformist Habsburg Empire. But the Jesuit expulsion spurred a general tendency to find alternatives to religious education also at the level of primary education. Solutions varied from the decision of the Papal State to keep the education entirely in the hands of ecclesiastical teachers, to intermediate forms that handed education to laic orders, such as the *Scolopi* in the Grand Duchy of Tuscany, to the first attempts to have a public education system in the Kingdom of Naples of Ferdinand IV of Bourbon and in the Lombardy ruled by emperor Maria Theresa of Habsburg-Teschen. In the Kingdom of Naples, the diffusion of new collective teaching methods (normal schools) and new public primary schools was effective in metropolitan Naples and neighboring areas (Terzi 2001; Zazo 1927), but less so in rest of the kingdom due to insurmountable obstacles posed by

local noblemen and clerics (Lupo 2005). This explains the exceptionally high levels of literacy rates in Campania, where Naples is. Also, the reforms by Maria Theresa were successful in urban centers, but less so in rural centers (Toscani 1993). Importantly, Maria Theresa also suppressed the religious orders to fund public and - ideally - mandatory primary schools. The political government of her descendant Joseph II of Habsburg-Lorena specifically targeted areas with fewer private schools. Toscani (1993) observes how in the agrarian region of *Bassa Padana*, the schools were now many, but the schooling demand remained low and so did literacy.

The first phase of French domination (1796-1799) was too short-lived for the invaded states to adopt a consistent set of reforms in line with the revolutionary ideologies, although the French imposed very high tributes to all conquered regions (Federico and Dincecco [2021, 7]). The second phase of French domination (1801-1815) saw a gradual incorporation of all the countries of pre-unitary Italy to what became the French empire in 1804, excluding Sicily and Sardinia.⁸ This time the domination was also political, and was exercised directly by Napoleon Bonaparte, Bonaparte family ties, or close allies. The Napoleonic government introduced radical reforms to various aspects of the civil and administrative system in all the Empire, creating what is sometimes referred to as “administrative monarchy” (Benigno 2014). The reforms came at the cost of higher taxation relative to *ancien régimes*, also to fund the French and Italian military forces (Federico and Dincecco, [2021, 7-8]). This mentality shift was also reflected in the design of public schooling policies: from 1802 on, various laws and decrees aimed at centralizing the education around the state, who decided the syllabi, imposed free, mandatory attendance to primary schools, prepared and funded teachers, and directly controlled private, often ecclesiastical, schools. The administrative uniformity of the Napoleonic period implied that the public schooling reforms of the time was eventually imposed all over continental Italy. And yet Vigo (1971) stressed how only the states that were not under the direct control of Napoleon – the Kingdom of Naples and especially the Cisalpine Republic which included Lombardy – could have the necessary public autonomy to create a coherent set

⁸ Postigliola and Rota (2020) provides a precise timeline of the incorporations to the French domination.

of legislation that was tailored to the local issues. Yet, in Lombardy public expenditure in schooling varied considerably by department (Piseri, 2017, 105-119, 139–142, 222–236). However, in the Kingdom of Naples even the larger autonomy seemed to be insufficient: the first set of reforms of 1806 proved to be ineffective until central authorities were given more reach over the provinces with the *Decreto Organico della Pubblica istruzione*.⁹ Consistent with this view, Toscani (1993) finds a doubling in the number of primary schools in Lombardy after the Napoleonic domination.

The most evident effect of these reforms was to centralize and boost public education provision, although with important regional variation. This centralization did not mean that funding was also completely centralized. The central government only had the duty to inspect the schools, probably as a consequence of the crisis of the public finances that followed the collapse of the absolute monarchies. Moreover, while the number of schools increased, this did not necessarily go hand in hand with a shift in the public perception of primary schools as having the different features from “ancient” schools, influenced by religious beliefs, local élites, and parental choices.¹⁰

After the Restoration (1815) every Italian state at least formally reverted to the administrative systems of *ancien régimes*, with a notable exception: the newly formed Kingdom of Two Sicilies, composed of the Kingdom of Naples and the island of Sicily. Arguably to reduce popular support for revolutionary ideas, taxation was kept low (Federico & Dincecco, 2021). As in many other fields such as taxation of land (Felice, 2013), the Borbonic king remained idle in terms of creating new primary schools: Gargano (2012) documents a fall in

⁹ For a detailed description of every legislative measure enacted in the Kingdom of Naples in the Napoleonic decade (1806-1815), see Lupo, Gargano, and Marra (2014).

¹⁰ Banti (1996, 55-97) and Macry (2002) show how noblemen, landowners, small business owners, and professionals in the North and in the South found ways to keep the family wealth concentrated that resembled the ones prohibited by Napoleonic Code on succession (in particular, the *majorat*) up to the late nineteenth century. It is possible to imagine that families retained a similar unchanged attitude towards education despite the institutional reforms.

the number of *public* male schools (from 2,498 to 822) and a fall in public female schools (from 1,556 to 1,142). Also the number of *private* schools decreased. Alternatives to public schools (i.e., evening schools, adult schools) concretely diffused only after 1861 (Vigo, 2017). Meanwhile the monarchic response to the revolts of 1820-1821 was particularly repressive and gave back a prominent role to the clergy in the public education, in line with the conservative attitude that other Italian States had in 1815 (Lupo, 2005).

In Northern and Central Italy, except in the Papal State, the role of the State in the public education system was gradually re-introduced. Indeed, the new absolutist governments were not able to restore the set of the privileges that their aristocratic title or ecclesiastical status entitled them to before the Napoleonic period. As a result, already in 1818 the Austrian Empire extended its public schooling system to *Lombardo-Veneto*, for boys and girls alike. In *Savoia* (today's Piedmont), the schools remained in the hands of the clergy until the Boncompagni law 1848. The Boncompagni Law re-affirmed the central role of the state in the education system by controlling also private and ecclesiastical schools, and was the steppingstone of the 1859 Casati Law, later imposed to unified Italy. Instead, the legislative efforts of the Bourbons in the South were increasingly dedicated to secondary and higher education, and the response to the 1848 insurgency was again repressive (Lupo, Gargano, and Marra 2014).

3. Sources

3.1 The Antenati marriage certificates

In Italy, historical civil records are scattered across many different archives, making it difficult to conduct comprehensive studies using these sources. Here I transcribe a random sample of scanned certificates from the Antenati genealogical platform. This initiative is an ongoing work of civil registration records scanning from various Italian state archives. Out of the 137 Italian State archives, 85 contain marriage certificates for the period of interest. Of these, 31 are still awaiting digitisation, so I considered the certificates from the available 54 State archives. The fact that the digitisation effort is still ongoing could raise concerns about whether it makes sense to already use this source. Yet, the majority archives with pending digitisation are minor ones (i.e., Assisi, Chieti, Pistoia, Massa, Frosinone), while there are large

areas where the data is structurally unavailable, either because the certificates do not have signatures (i.e., in Genova) or because marriage certificates are located elsewhere, such as parishes and civic courts (i.e., Milan, Nuoro). These instances will not be solved even when the full digitisation of State archives will be complete: even having the full sample from Antenati would not give a fully representative picture of Italy, at least because the entire regions of today's Latium except the southernmost provinces, Sardinia, Umbria, and the city of Milan are not covered. For this reason, after having stratified the sample to account for population imbalances, we can at most claim that the sample is representative of the regions covered by the digitized archives, which however include most of Italy. The research can later be extended with the remaining archives, and the sampling approach presented here serves this purpose, although data preservation issues make it clear that for some areas there will still be no marriage certificates with signatures in the future (see next section, figure 3, figure 7 in appendix "1" and supplemental material for the spatial coverage of the Antenati). I will refer to the sample of marriage certificates drawn from this source as the Antenati sample. The paper also makes use of datasets derived from the published census reports of 1861, 1871, and 1881, and of other official documents.¹¹

3.2 The Processetti and Documenti Matrimoniali

Before marrying, spouses had to submit documents to the priest in preparation for the upcoming marriage. Documents included declarations of not being already married, or about being no legal opposition to the marriage. Such documents were legally binding, so spouses were often required to sign them, even when the marriage certificates were only signed by the

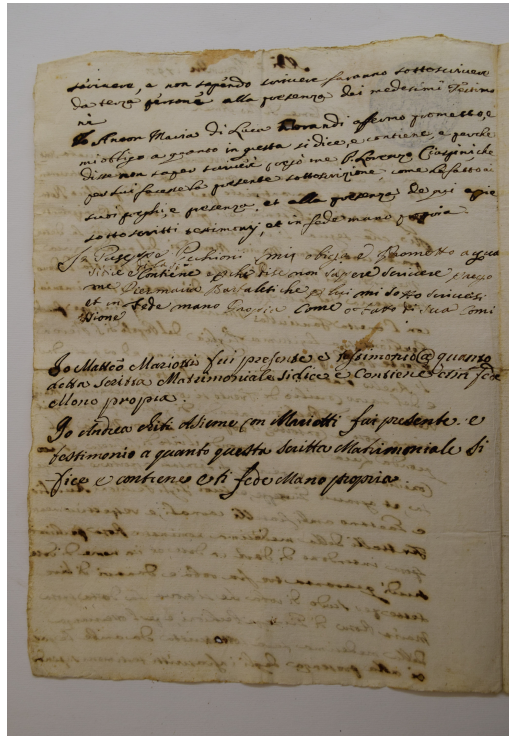
¹¹ Occupations were coded according to the HISCO and HISCLASS classifications of Van Leeuwen (2020) and Van Leeuwen and Maas (2011). See supplemental material for further details. The official sources I use are MAIC (1866a, b); MAIC (1875); MAIC (1883); MAIC (1990), and Ministero dell'Istruzione (1890). I also refer to the elaborations of census data of Ciccarelli and Weisdorf (2019).

priests, as before the Napoleonic reforms. For this reason, the *processetti matrimoniali* are an important alternative source to marriage certificates in Italy for the pre-Napoleonic period. However, this is not the case for the entire Italy: virtually every parish in Northern and Central Italian states did not require the signatures of the spouses, so the *processetti* are a useful source to study pre-Napoleonic literacy rates only in the Southern Kingdom of Naples. And even there the evidence is hard to come by: every parish has its own archive, and the preservation quality varies considerably. The Diocesan State Archive of Naples is an exception as it contains the *processetti* relating to marriages of the entire city of Naples and is accessible up to the year 1800 not included. I thus collect the signatures of 224 marriage *processetti* of brides and grooms whose surname started with the letter 'B' in 1750 and with letters 'B and L' in 1775 in Naples.¹²

For the rest of Italy, the evidence is even more scant as only few parishes decided to ask the spouses to sign certificates of equivalent to *processetti*, but we do not know in advance which and each parish keeps its own archive: for the province of Florence, out of more than 200 parishes gathered in the *Archivio Arcivescovile*, only 6 had *processetti* for the eighteenth century, and only one parish (Santa Maria La Romola) had signatures of 26 spouses and they covered the period 1767-1797 (Figure 1).

Figure 1. Second page of a marriage document from Santa Maria La Romola, Tuscany, 1767.

¹² I checked whether Northern and Central Italian archives contained signatures of spouses before Napoleon by contacting or visiting the Diocesan archives of Milan, Como, Brescia, Bologna, Florence, Pisa, and Lucca. Northern and Central Italian parish archives sometimes contain signatures for legal processes regarding divorces. However, there are just few such processes, and they are very selected because at the time it was only possible to ask for divorce due to very severe situations of violence, and the interested spouses were often wealthy.



3.3 Measuring literacy rates in the Antenati sample

The ability to sign in marriage certificates is the proxy for literacy rates used. The ability to sign can be indicative of functional literacy adjusted to the standards of the early nineteenth century Italy.¹³ Despite that, comparing literacy rates derived from signatures with stated literacies in censuses or conscript literacies may not be legitimate if they measure different dimensions of literacy. In Italy, literacy inferred from marriage records in 1872 matches very closely census self-reported literacy for the 1871 census. This partially suggests that it can be legitimate to compare marriage certificates with other sources of literacy rates, although Italy in 1815 was still pre-industrial, and thus being able to read and write was probably less endorsed

¹³ If a person can sign a certificate, this does not mean that she can access book production or the popular press (see Marchesini 1992). There are sources that enable measurement of more complex literacy skills, such as receipts of payments but these sources are scattered and hardly systematic (see Petrucci 1987). I classified trembling signatures as “poor handwriting”.

by the population than when needed to access better occupational positions. But even in 1815 the ability to sign predicts well the socio-economic status of spouses.¹⁴

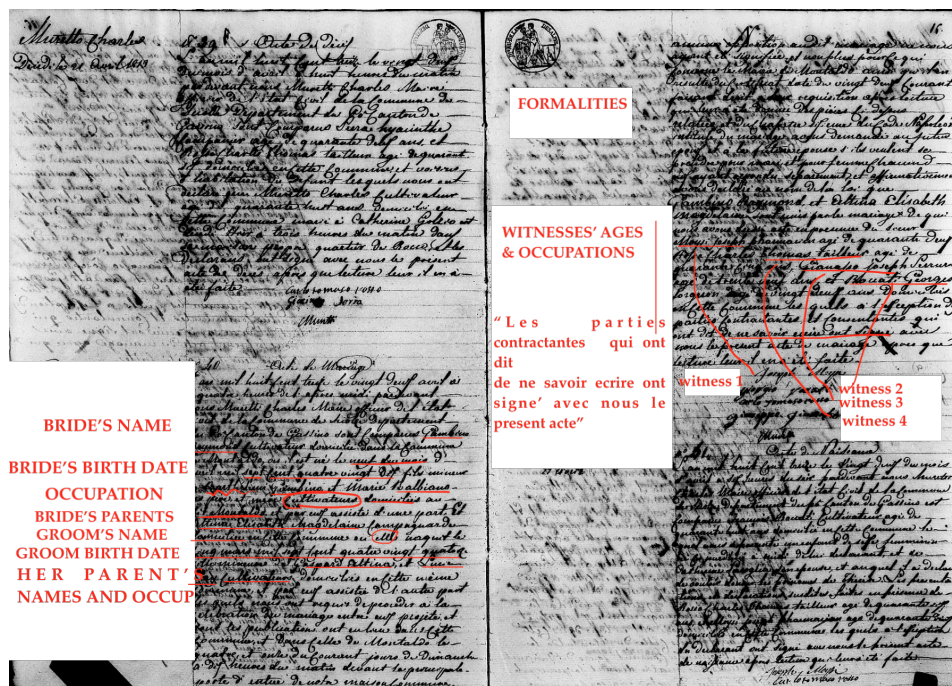
Another concern might be that as spouses are on average younger than the average population who can become literate (above 6 or 12 years old), they might be more literate than the average if the literacy rates are growing in 1815. Marriage certificates in Italian civil records also provide information about occupation, age, and place of residence not only of spouses but also of their parents, and of the witnesses. Moreover, signatures of the spouses in marriage certificates suffer less selection biases than other sources such as the signatures of witnesses in death records or the signatures of sailors' boarding certificates (Rediker 1989) because every strata of the population married, and spouses did not have particular reasons to learn specifically how to sign when compared to the rest of the population.

I focus on the year 1815 also because it is the only out of which we can have comprehensive geographical picture of literacy based on signatures. Indeed, civil registries were abolished in most Restored States after 1815. Parish registers regained the responsibility of recording births, deaths, and marriages, as with the *status animarum*, but now recording information according to the Napoleonic format. Civil registrations were only maintained in the Kingdom of the Two Sicilies and in the Duchy of Modena and Reggio; for the latter, there are only remnants for Southern Liguria digitized before 1860. The Civil Registration was also introduced in Sicily in 1820. Figures 2 show an example of Napoleonic marriage certificate. Only the format of Restoration civil records record is standardized and based on a pre-printed

¹⁴ See figures 10 and 11 in “appendix 1” and Brian A’Hearn, Alexia Delfino, and Alessandro Nuvolari, “Rethinking age heaping: a cautionary tale from nineteenth-century Italy”, *The Economic History Review* 75, no. 1 (2022): 111–137. In the first three unitary Italian censuses 1861, 1871, and 1881 the enumerators asked if individuals were able to at least read, to read and write, or were illiterate. Each family head received *schede del censimento* from enumerators (Mastroluca and Verrascina 2012). The illiterate family head was had to find a literate individual among neighbors to fill out the form.

format. I collect Restoration-period records for Southern regions for years as close as possible to 1815 (usually 1816) after the end of the Napoleonic period. For Northern Italy, with only certificates dating back to the Napoleonic period, I collect information that is as close as possible to 1815 (usually 1814) to exclude years where the number of certificates or the type of spouses could have been affected by the novelty of the civil registration institution, and to have a sample with comparable years. Excluding second marriages, the mean age of spouses in the sample is 26.76 years old for grooms and 23.25 years old for brides, in line with Rosella Rettaroli (1990).

Figure 2. Marriage certificate from Sciolze, Piedmont (1814).



4. Research Design

4.1 Sampling Strategy

I adopt a clustered random sampling scheme, randomly selecting villages within the village list of each archival unit. Then I weight the total sample size by the regional population weight and the number of certificates to have in each archive over the total of the region, and start collecting the required marriage certificates from the randomly selected villages within

each archival unit. Each of the 49 State archives with full marriage certificates, or archival unit, has a list of ‘villages’. These are often municipalities, but in some cases correspond to smaller villages or even parishes.

Obtaining a list of the villages from the online platform is straightforward, but it would be very difficult to obtain an ordered list of marriage certificates within each archival unit, because in some cases there are more than one certificate per scanned page, and in other cases there is a certificate every two or even four pages. A clustered random sampling scheme solves these difficulties because it samples units (villages) that are more aggregated than marriage certificates.¹⁵

I then use two more weights to post-stratify the obtained sample in an iterative fashion to have enough observations in each gender, region, and urban-rural cell. The iterative approach is based on three weights (see figure 12 in “appendix 3” and supplemental material). The first weight corrects for between-region variability in archive coverage with regional and provincial population estimates for 1815 drawn from Mariella, Postigliola, and Rota (2020), whose data reconstruction can be considered the most updated and complete. This weight stratifies by the population by unitary region when the Antenati archive does not include all the provinces within a region, and by the smaller level of pre-unitary province when the data collection is more complete. The second corrects for the share of urban areas in the region, derived the

¹⁵ Schofield (1973) cannot observe beforehand how many marriage certificates are contained in each set of parish records, and therefore how to construct a weighted sample of the population, so inference is valid only at the parish-level. Here I can make a step forward because observing the total number of marriages within an archive allows to obtain a balanced sample of marriage certificates at the archive level. Linking this to the number of certificates needed by region observed by post-stratifying by population guarantees that that the Antenati sample is representative at the archive level and, most importantly, at the provincial/regional level. Also Cappelli and Vasta (2020b) use a stratified sample of socio-economic and institutional features of municipalities to study the impact of the Daneo-Credaro schooling reform of 1911 on municipal-level literacy.

population from the 1800 estimations of Malanima (2016). I consider 9,000 in 1800 as the minimum population threshold for larger towns in 1815, and 4,000 as minimum threshold for smaller towns. I also assume that when town size in 1800 is smaller than 5,000, it is 4,000. I introduce an indicator to distinguish cities from agricultural agglomerations ('agro-towns') and only include cities in the urban population (see supplemental material).¹⁶ Sensitivity analyses suggest that the results and coverage in the Northern Italy would be dramatically underestimated if not using weights, while they are quite similar when considering the 15-30 marriage cohort only and the 30-40 marriage cohort separately instead of the entire one (see figures 8, 9, and 10 in "appendix 1"). Alternatives such as simple random sampling would imply to open all pictures first and create a one-marriage, one-picture set of variables to then neatly randomize the selection procedure.¹⁷

As a result, the weights that I use refer to the same objects during the sampling process and there is no need to worry about updating. Another advantage of this strategy is that it allows to conduct analyses that represent the specific village better than randomly picking certificates from different villages. It also allows to integrate the research with qualitative information by

¹⁶ Chilosi and Ciccarelli (2021) show that the agro-town issue is present much more in unitary censuses than in pre-unitary ones: the agricultural labour share in the South is roughly on par with the one of the Centre-North in unitary censuses while it is much higher in the South in pre-unitary censuses. Given that the provincial population data I use is not based on pre-unitary censuses but rather on a set of other official publications collected by Mariella, Postigliola and Rota (2010). I adjust the urban indicator for agro-town population.

¹⁷ One might be concerned that the sample is not fixed, because the online uploading of new certificates is ongoing, and this might invalidate the weights because they would not represent the different parts of the sample as new data is introduced. However, the clustered sampling scheme solves this difficulty because it considers villages as sampling units, and if a village is available on the Antenati platform it means that the collection is complete for that village. This means that even though the online source is updated over time, my sampling frame does not change as it is based on the list of 5,874 villages with full marriage certificates available in the platform at a specific point in time (September 2022). See also supplemental material table A1.

village. Also, the process of random selection can be documented clearly because it is based on the list of potential villages to select created before starting to sample. The collection covers quite uniformly the areas of Italy that are contained in digitized archives. Each considered region or province, and each urban and rural area has a balanced amount of certificates (see table 4 in “appendix 1”). Although three regions are missing altogether, the universe of interest for which this sample is representative at the level of region and province is a vast portion of continental Italy (Figure 3 and 4 and figure 7, 8 in “appendix 1”).

Figure 3. Spatial distribution of randomly sampled clusters, or ‘villages’. Larger dots indicate more marriage certificates collected. Provinces are the contemporary ones.

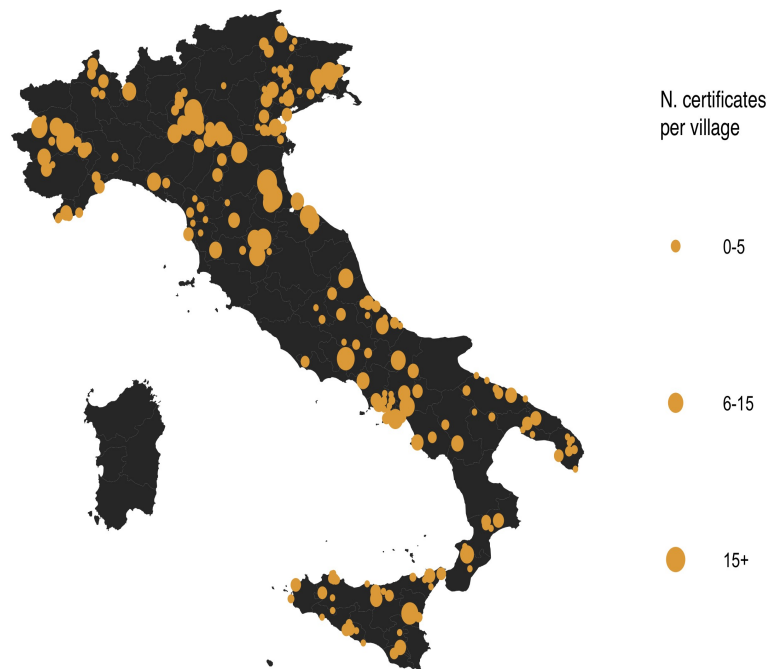
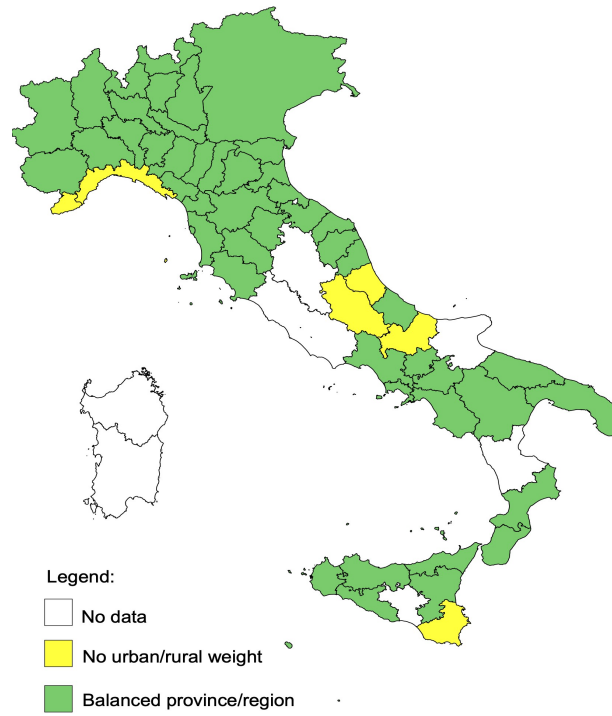


Figure 4. Sample coverage by region in the Centre-North and pre-unitary province in the South.



4.2 Regression-based adjustments

I then use a 'regression' approach to further correct for sampling imbalances. This approach was introduced by with similar weight to the second weight of this sample: the share of each region's population over the national population times the number of observations per region. I adapt the approach of Clark (2005) and of Federico *et al.* (2020) to a context where the variability is not over time but across regions and provinces within a single cross-section and where the dependent variable is binary. I estimate literacy rates in a (logit) regression framework controlling for skill, gender, urban/rural and regional dummies as follows:

$$Lit_i = \alpha + \sum_{i=1}^N \beta_i Skill_i + \gamma Gender_i + \sum_{j=1}^J \delta_j URBRUR_j + \sum_{j=1}^J \theta_j LOC_j + \varepsilon_i \quad (1)$$

Where Lit_i is the ability to sign of the i -th observation; $Skill_i$ is a categorical variable with three values three values: primary, secondary and tertiary derived from HISCLASS 12; $Gender_i$

is a gender dummy; $URBRUR_j$ is a categorical variable with the share of urban and rural population within area j ; and LOC_j is a categorical variable with a value for each province, region or macro-area considered. As it was not possible to stratify the sample by skill levels without sampling all marriage certificates, in order to consider also the skill dimension in the regression I aggregate the results by macro-area (North, Centre and South) to have enough observations for each skill level. To obtain shares of literate individuals out of the logit regression odds I compute the predicted probabilities of being literate conditional on each reference level of the other covariates (Table 5 in “appendix 1”).

To obtain aggregate literacy rates across macro-areas, gender, and urban-rural dimensions I considered two options. The first consists in using shorter forms of the regression to compute the values across the entire distribution of the other regressors. For instance, to obtain literacy rates by gender and area across urban-rural and skill levels, I run a regression with only $Gender_i$ and LOC_j as regressors, and then observe the literacy rates by gender in the North, Centre, and South respectively. This allows not to specify the reference category of the remaining regressors (i.e., rural and of high status), but the estimates may suffer from omitted variable biases.

Table 1. Italian literacy rates by gender, location, and occupation category (%) predicted with a weighted logit regression approach.

Macro Area	Summary		Gender		Urban-Rural		Occupational Sector			
	N	Average Literacy	Male	Female	Urban	Rural	Services	Industry	Agriculture	GPI (%)
North	1,295	25.92	38.70	13.15	36.17	22.42	55.61	36.17	14.52	39.97
Centre	274	16.06	24.16	7.97	17.83	14.59	38.85	21.37	7.11	32.18
South	1,061	13.86	21.74	6.00	15.64	13.13	38.56	21.15	7.02	27.60
Italy	2,630	19.32	29.27	9.38	21.85	18.34	45.92	27.40	10.14	34.09
N	2,630	2630	1,310	1,320	574	2,056	252	692	1,385	

*The occupation categories correspond to Services: HISCLASS 1-5, equivalent to 'Higher Status';

Industry: HISCLASS 6, 7, and 9, and 11; Agriculture: HISCLASS 8, 10, and 12.

** The weights used to aggregate are the new urban-rural weight with the population weights derived from Mariella, Postigliola, and Rota, and the skill weights in 1815 derived from Chilosi and Ciccarelli.

*** GPI stands for Gender Parity Index: the ratio of female over male literacy rates.

The second and preferred option uses the fully specified regression, disaggregated across the various dimensions (see Table 1) and aggregates the different cells using population, urban-rural, and skill weights obtained by North, Centre, and South. The population and urban-rural weights correspond to the weights 2 and 3 used to stratify the sample and aggregated by macro-area, respectively: as some provinces and regions were not covered in the sample, I prefer to use weights that only include covered areas and that refer to a date as close as possible to 1815. The three-categories skill weight was obtained from Chilosi and Ciccarelli (2021).¹⁸

The regression should yield more precise estimates in cells that for some reason were not accurate with the sampling method because of sampling imbalances and in the skill cells that were not stratified for in the sample. The estimates are similar across all dimensions (compare table 1 and “appendix 1” Table 6).

5. *New Estimates of Literacy Rates and Comparison with previous studies*

5.1 *The Antenati Estimates*

In this section I describe and interpret the results weighted using the regression approach. I find significant North-South and gender gaps in literacy. On average, spouses of Northern Italy signed 12% fewer certificates than Southern spouses (Table 1). The gap appears lower when directly using sample averages (9%), but also there, considering the lower bound of Southern literacy rates of 14% and the upper bound of Northern literacy rates of 26%, the gap

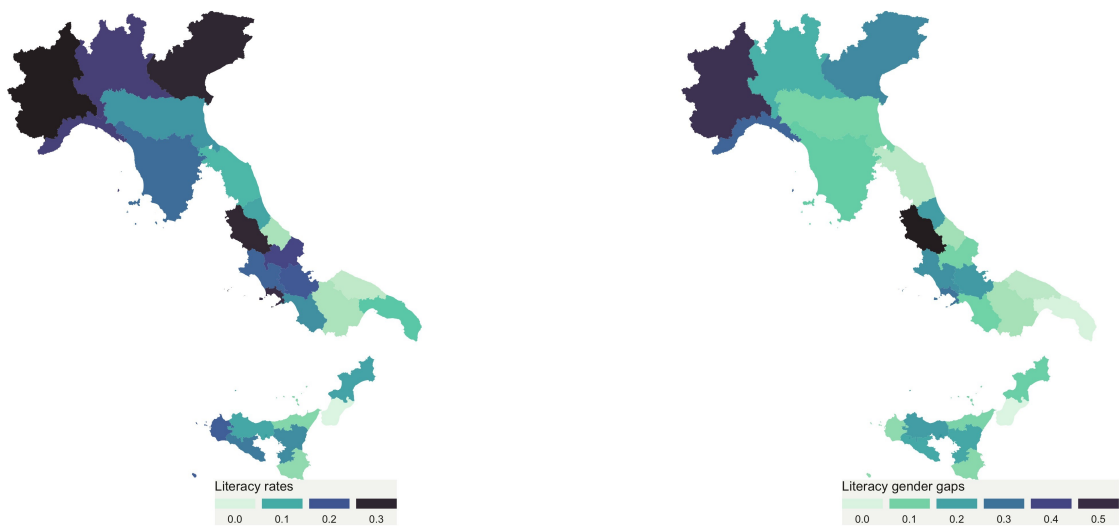
¹⁸ Suppose that, for instance, we are interested in observing the literacy rates of women in Northern Italy across urban-rural and skill dimensions. Calling a , b , c the literacy rates of women in urban areas and d , e , and f the literacy rates of women in rural areas across the three occupational sectors respectively, I aggregate as follows: $w_{women, North} = w_{rur} * (w_{serv, North} * a + w_{ind, North} * b + w_{agr, North} * c) + w_{urb} * (w_{serv, North} * d + w_{ind, North} * e + w_{agr, North} * f)$.

becomes approximately 12%. Importantly, the North-Centre gap is comparable in size to the North-South gap. Even though Northern Italian spouses could sign marriage certificates only 26% of the time, their Southern and Central Italian counterparts did so 14% and 16% of the time on average, respectively. The gap is mostly explained by the high literacy rates of Piedmont in the North and the low literacy rates of most Southern and Central Italy except for Tuscany, Basilicata and Campania (Figure 5a). The provinces in Calabria, Apulia, and Basilicata display the lowest literacy rates in the South; the central regions of Tuscany and Marche closely match the literacy levels of southern regions; The literacy rate of the Emilia-Romagna region in the Centre-North is half the literacy rate of the neighboring region of Lombardy, and similar to that of the province of Principato Citeriore (Salerno) in the South.

Figure 5. The literacy rates for the Antenati sample are plotted over contemporary provinces in southern Italy and regions in Central and Northern Italy.

(a) Literacy rates

(b) Literacy gender gaps (Male-Female)



Literacy gender gaps are consistently high throughout Italy, and lower only in the regions of Italy with lower average literacy rates (Figure 5). 29.27% of grooms signed the marriage certificates, while only 9.38% of brides did. The resulting literacy Gender Parity Index

(GPI), the ratio of female to male literacy is equal to 34.09% across Italy, and with a clear North-South pattern increasing as moving North (see also figure 5b). The North-South ratio for female literacy rates is 2.2. This is lower than 2.7, the value provided by Cappelli and Vasta (2020a) and based on elaborations from Ciccarelli and Weisdorf (2019) data. This ratio reached levels as high as 3.5 in 1860, suggesting that it widened markedly during the early nineteenth-century. Southern and Central Italy are surprisingly similar in terms of aggregate literacy rates also across gender, urban rural and occupational dimensions. In the Centre-North the urban-rural literacy gap is very marked, while in the South is almost non-existent (Table 1): the opportunities for becoming literate were of lower quality but more diffused in the South, while they were concentrated in urban areas in central Italy.

This might explain why over time, literacy rates in central Italy increased while they did not in the South: the new model of school-based education was centered around urban agglomerations before diffusing in rural areas. The fact that urban areas in Central Italy had higher literacy rates than urban areas in the South suggests that in Central Italy the process of schooling diffusion had already started and would have radically changed the scale of literacy diffusion over time. The wider spatial diffusion but lower overall level of literacy rates in the South instead suggests that besides large cities such as Naples, the process still had to be initiated there. The literacy rates of farmers (HISCLASS codes 8, 10, and 12) were of 10.14%, while workers in industry (HISCLASS codes 6,7, 9, and 11) could sign the marriage certificates 27.40% of the time on average. Among spouses of higher status (HISCLASS codes 1 to 5), mostly working in services still just 45.92% of them could sign a marriage certificate. The literacy rate of Northern Italian farmers was double to the average literacy rate of central and southern Italy. This can be explained with the effort of school diffusion in rural areas in particularly in the Lombardy ruled by Joseph II of Habsburg-Lorena. The literacy rates in the North are especially high among farmers (14.52%) and among higher status workers (44.61%) comparing to the rest of Italy.

5.2 Comparison with unitary censuses

Here I compare the Antenati results with unitary census-based estimations. Ciccarelli and Weisdorf (2019) uses the cohort of individuals aged 90 to 100 years old to infer the literacy rates of the group aged 30-40 in 1821. These figures will be the main comparison of our figures, which date back to around 1815 for spouses aged 23 years old on average. Figures 5 and 6 show the evolution over time of literacy rates by macro area and gender, respectively. The estimates of Ciccarelli and Weisdorf are higher for Northern and Central Italy than the 1815 Antenati sample estimates, while they are similar for Southern Italy: in the North, literacy rates were 26% in the Antenati sample and 37% in Ciccarelli and Weisdorf.¹⁹

An underestimation driven by using older cohorts in more recent censuses can explain the discrepancies between the results using the 1871 census and the estimates by Ciccarelli and Weisdorf, who used the 1881 census, but cannot entirely explain the difference that we observe between our sample and either the 1881 or 1871 census retropolations. Other explanations can relate to issues of census data retropolations: selective mortality, cohort effects, differences between stated and proved literacy abilities, and age heaping (see supplemental material for a technical discussion). Taking the uncertainty that all these biases – which cannot be circumvented -- imply for an even comparison, marriage certificates' and census' figures are comparable.

¹⁹We also use the 1861, 1871, 1881 censuses (see note 10 for more details).

Figure 5. Literacy rates by source and area (North, Centre, South).

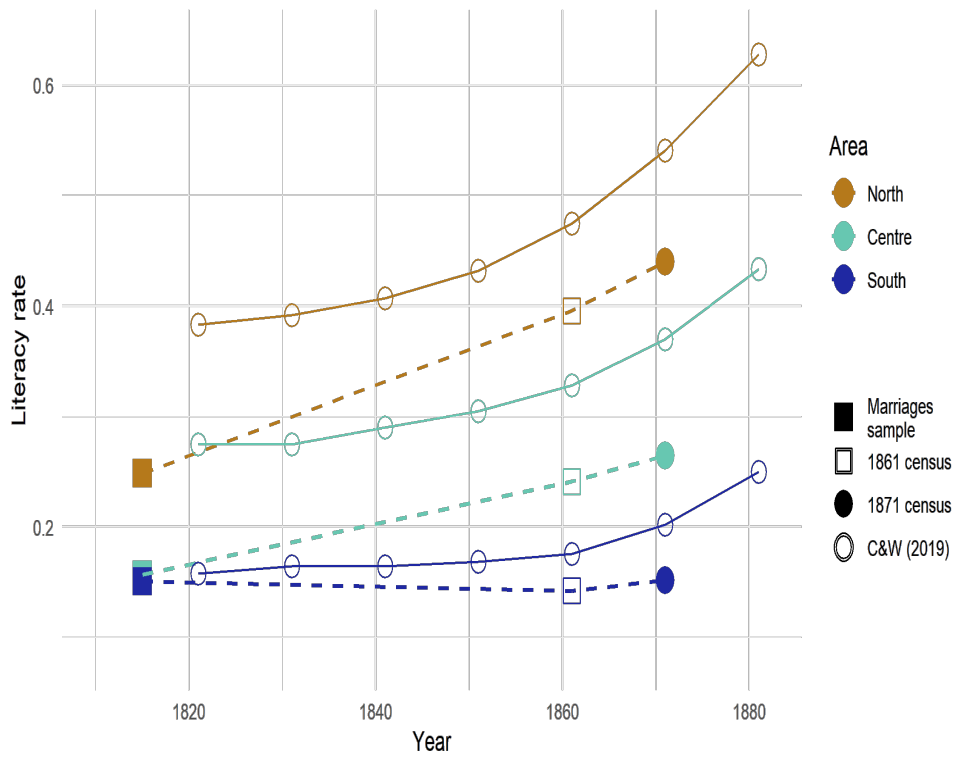
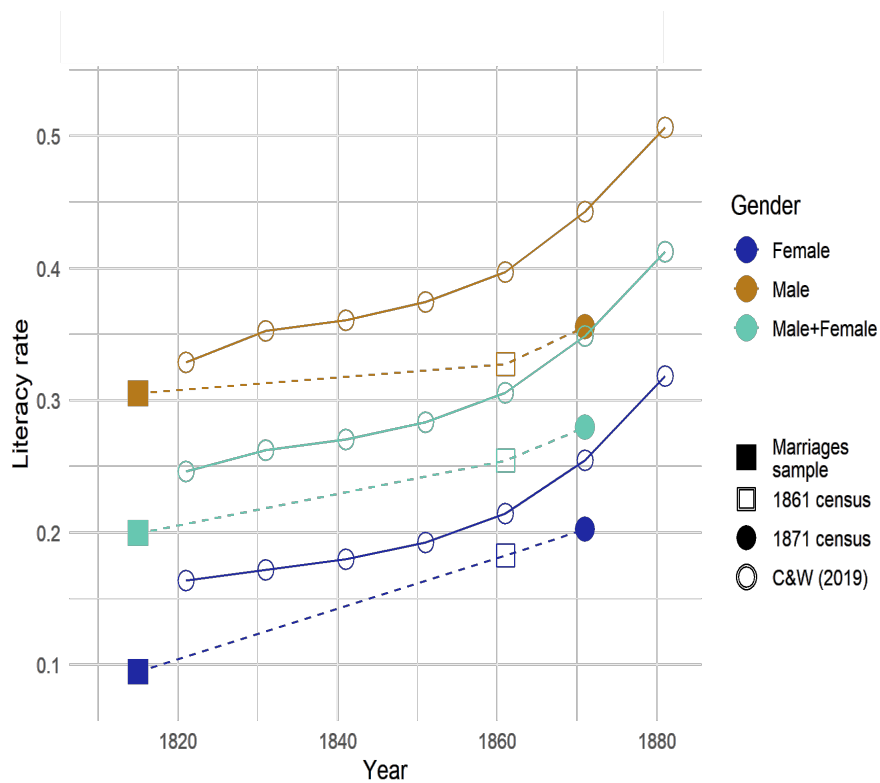


Figure 6. Literacy rates by census source and gender.



5.3 Comparison with Local Studies

I proceed by comparing the estimates of literacy rates with a set of local estimates partially collected also for this study. In this paper, a "local study" is a study that contains quantitative estimates of literacy rates based on signatures in local areas such as villages and that can sum up to provinces or regions (see Table 2). We need to be very careful when comparing the estimates, as they often refer to selected areas of the selected regions or cities.²⁰

In the supplemental material I compare more in detail the cases of Lombardy, Udine, and Salerno and find that after considering the skill composition of each sample, the estimates are comparable, although the Antenati sample seems to underestimate literacy rates, but uniformly across Italy (see also tables 7, 8, and 9 in "appendix 1). A more historically informative set of observations can be derived for the case study of Naples. We can see on literacy rates were already high for grooms in central Naples in 1750 (28.76%). They increased in 1775 for men (34.56%) but not so much for women. In 1810 and 1815, just before the Napoleonic reforms, but after the effect of Absolutist rulers' reforms could be evident, the literacy rates reached levels as high as 57% for grooms according to the full set of certificates collected by Alfonso Scirocco (1987), and 42% according to Antenati, which considers also more peripheral areas of the city such as Miano and Marianella and Stabilimento dell'Annunziata. This indicates that the city of Naples witnessed important literacy rate increases

²⁰ "Local studies" include Toscani (1987); Piseri (2002); Toscani (1993); Brambilla (1991); Milanese (1991); Ferraresi (1991). Other studies estimated literacy rates in areas of Italy in the period, but to the best of my knowledge these are the only ones that cover the areas from 1800 to 1861 and for areas also covered in the Antenati sample. For instance, the local studies for Lombardy oversample the areas in the mountains, where literacy rates are higher and the number of landowners is larger; The local estimates for Piedmont and Marche are not very reliable as they only include a very selected part of the region (4 villages for Piedmont and 30 for Marche, concentrated around the regional capital). See the supplemental material for the maps with covered villages. The estimates for Naples are very comprehensive but only focus on the central neighborhoods of the city.

which placed it at a comparable level to the literacy rates of Lombardy in the North as measured by data-rich studies of Xenio Toscani. This coincided with the primary school reforms of the late nineteenth century, rather than those of the Napoleonic period. In support of this explanation, Tuscany did not experience relevant schooling reforms before Napoleon and indeed the literacy rates we observe in the remote Tuscan village of La Romola in the late eighteenth century are very close to the average Grand Duchy of Tuscany literacy rates just before the Napoleonic period.

Table 2. Comparison of considered studies and new area estimates.

Reference	Region	Source	Period	N	N	Literacy	Litera
					Villages	certificates	male
							cy
							female
Antenati sample	City of	<i>Processetti</i>	1750	1	81	0.287	0.1034
	Naples	<i>Matrimoniali</i>				6	
Antenati sample	City of	<i>Processetti</i>	1775	1	148	0.345	0.1234
	Naples	<i>Matrimoniali</i>				6	
Scirocco (1987)	City of	Napoleonic	1810	12	1,922	0.57	0.25
	Naples	Records					
Antenati sample	City of	Restoration	1815	14	50	0.42	0.12
	Naples	Civil Records					
Processetti	La Romola	<i>Documenti</i>	1767-1790	1	25	0.214	0.00
sample	(Province of	<i>Matrimoniali</i>				2	
	Florence)						
Antenati sample	Tuscany	Napoleonic	1810-1815	36	118	0.24	0.06
		Civil Records					

Toscani (1987)	Lombardy	Napoleonic	1806-1810	176	14,072	0.43	0.08
		Civil Records					
Piseri (2002)	Lombardy	Napoleonic	1800-1830	29	302	0.38	–
		Civil Records					
Toscani (1993)	Lombardy	Napoleonic	1800-1830	341	13,625	0.49	–
		Civil Records					
Antenati sample	Lombardy	Napoleonic	1810-1814	26	203	0.32	0.15
		Civil Records					
Milanesi (1991)	Lombardy	Archivio di	1806-1815	4	683	0.22	0.12
		Stato di					
		Mantova					
Brambilla	Marche	Archivio di	1808-1814	30	--	0.19	0.06
(1991)		Stato di					
		Macerata					
Antenati sample	Marche	Archivio di	1814	5	57	0.11	0.07
		Stato di					
		Macerata					
Ferraresi	Piedmont	Napoleonic	1806-1810	4	656	0.25	0.13
(1991)		Civil Records					
Antenati	Piedmont	Napoleonic	1806-1810	28	170	0.60	0.09
sample		Civil Records					
Antenati	Udine	Napoleonic	1810-1814	8	60	0.18	
sample	province	Civil Records					
Antenati full	Udine	Napoleonic	1810-1814	4,67	88	0.14	
count	province	Civil Records		4			

Ciccarelli and Weisdorf (2019)	Udine province	1881 Census	1821	--	--	0.29
Antenati sample	Salerno province	Restoration Civil Records	1816-1818	5	76	0.20
Antenati full count	Salerno province	Restoration Civil Records	1816-1818	49	1,950	0.12
Ciccarelli and Weisdorf (2019)	Salerno province	1881 Census	1821	--	--	0.18

* The term local studies encompass the studies on pre-unitary Italian literacy and alphabetization of X. Toscani, M. Piseri, E. Brambilla, A. Milanesi, A. Ferraresi.

† For areas covered by more than one study (Lombardy) I considered the literacy rates only once per municipality. The ‘Antenati full count’ uses the full set marriage certificates available around 1815 in the area between the rivers Isonzo and Livenza for Udine, and north of Eboli (included) for Salerno.

6. Conclusion

50 years ago, Roger Schofield first attempted to scale up the conclusions from a sample of marriage certificate to country-wide estimates. Can we trust literacy rates collected from marriage signatures? This article collects evidence from an interdisciplinary set of approaches to measure literacy rates in Italy over the period 1750-1815, with a focus on 1815. The random sample of marriage certificates transcribed from the Antenati platform, together with the consequent weighting and regression approach gives a positive answer, as the estimates are generally in line with other micro and macro-level approaches. Methodologically, it shows that the key to make a random sample of marriage certificates valid to scale conclusions up to macro areas is to balance the sample characteristics by

region, province, occupational class, and urban-rural dimensions. This article proposes approaches to mitigate imbalances, but doing so requires a wealth of data that is not always easy to find pre-industrial periods. This can be a limitation of the approach in contexts with limited data to balance the sample with.

Relating to the debate on the origins on the North-South divides, this article also finds that North-South gaps in 1815 were smaller than previously thought, because literacy rates were lower in Northern and Central Italy. The evidence that literacy rates were particularly high in areas exposed to the public education reforms – Austrian-ruled Northern region of Lombardy and Borbonic Southern region of Campania, especially Naples – shows that despite differing socio-economic backgrounds, centralized State-led economic reforms are at least associated with increasing literacy rates. The new estimates give credit to new speculations: North-South gaps as we know them from later historical accounts may be originating in the Napoleonic period rather than being inherited from earlier ages. Disentangling the role of the Napoleonic domination to that of later administration of pre-unitary states to understand the respective role in the literacy rates and divides would be a valuable direction of future research.

Data Availability Statement

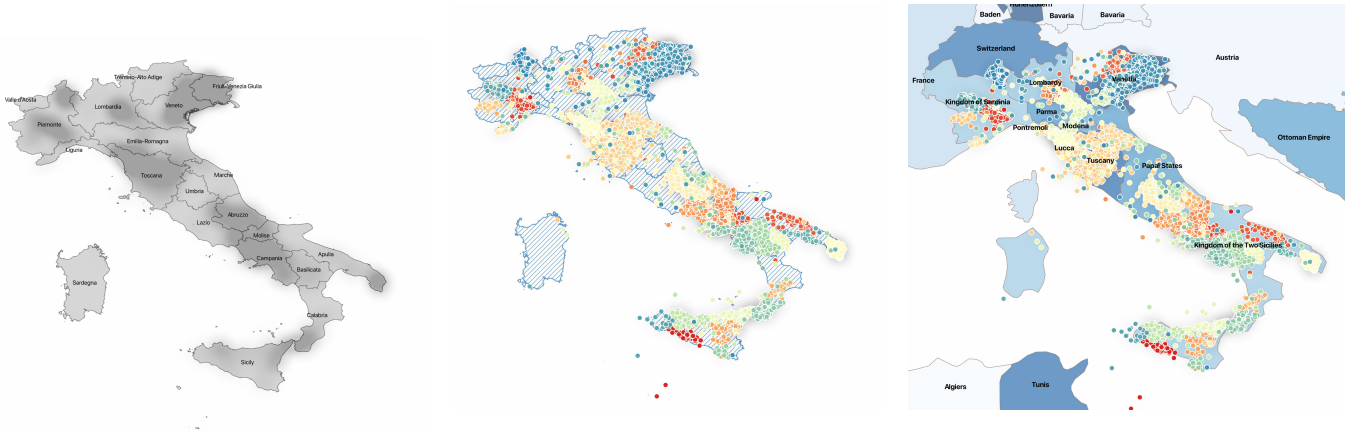
The data, methods, and replication codes underlying this article are available for further studies and replication on a anonymized repository that can be accessed via the following link:

<https://deposit.icpsr.umich.edu/deposit/claimResource?tenant=openicpsr&claimId=45579>

Appendix 1

See Figs. 7, 8, 9, 10, and 11 and Tables 4, 5, 6, 7, 8, 9, and 10.

Figure 7. Concentration of villages in the Antenati platform.

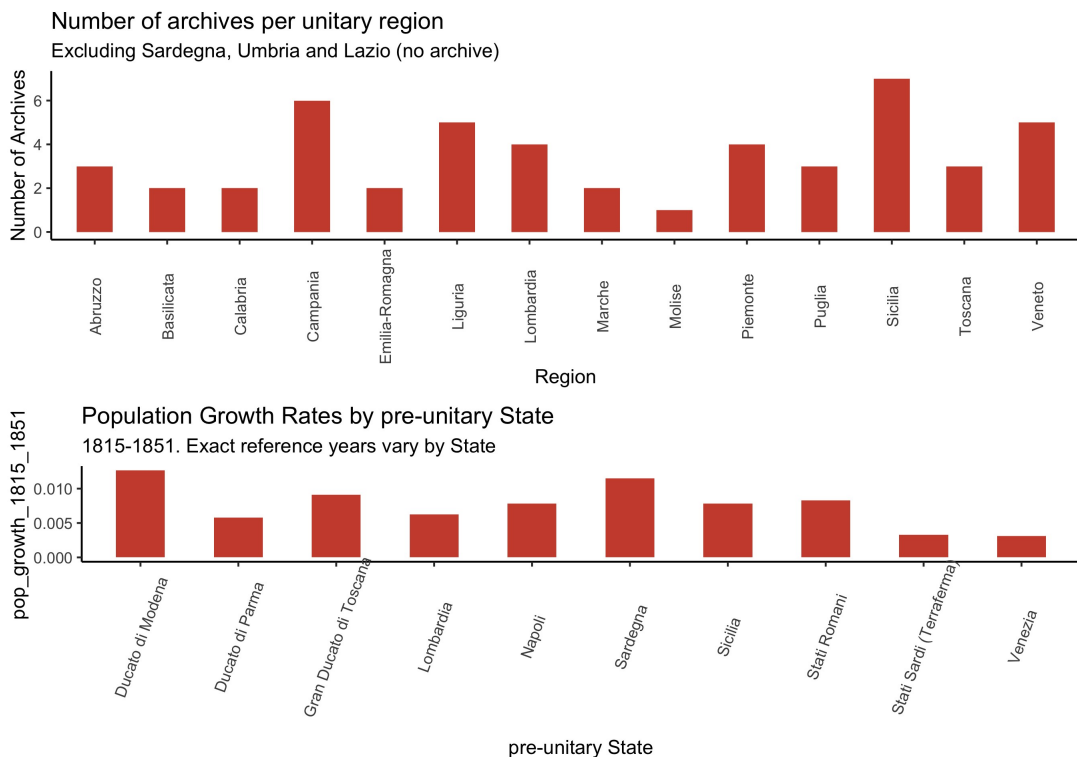


(a) Concentration of villages over according to today's regional configuration of Italy.

(b) Every point represents a village, and every color an archival unit. The boundaries are today's regional boundaries.

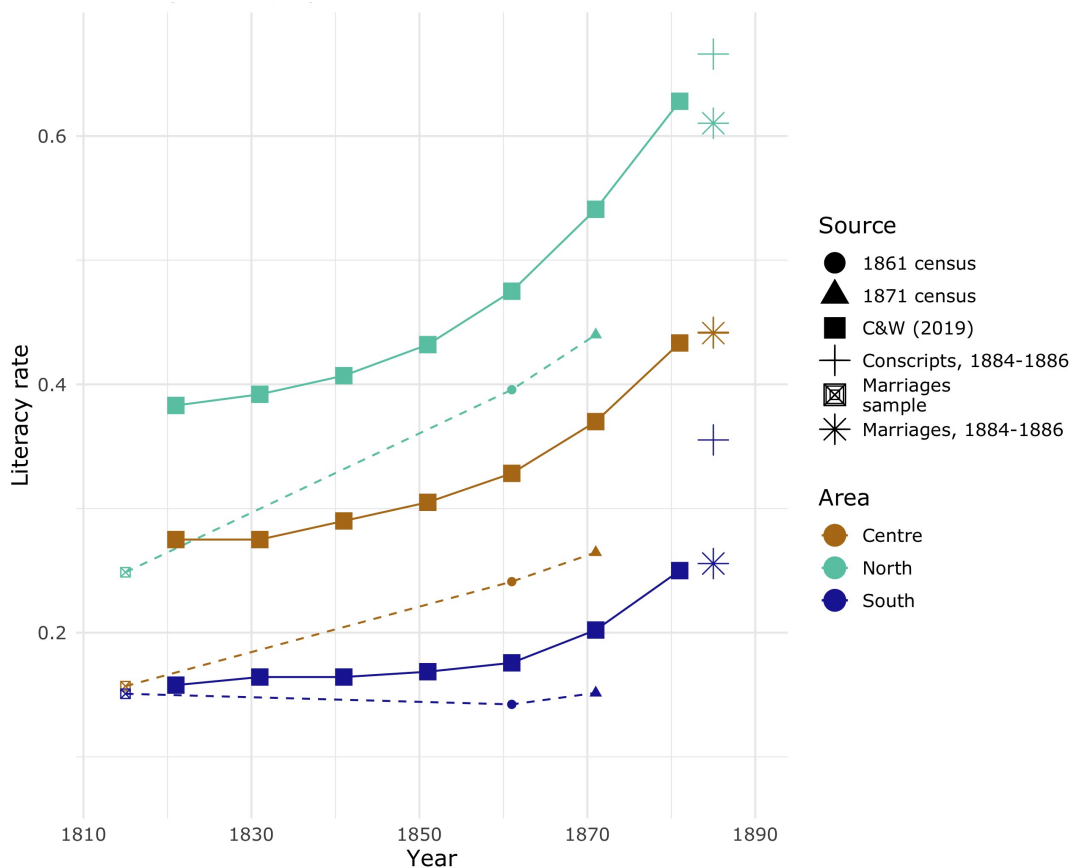
(c) Villages are plotted over 1815 boundaries.

Figure 8. Number of State archives by region and population growth in pre-unitary States (1815-1851).



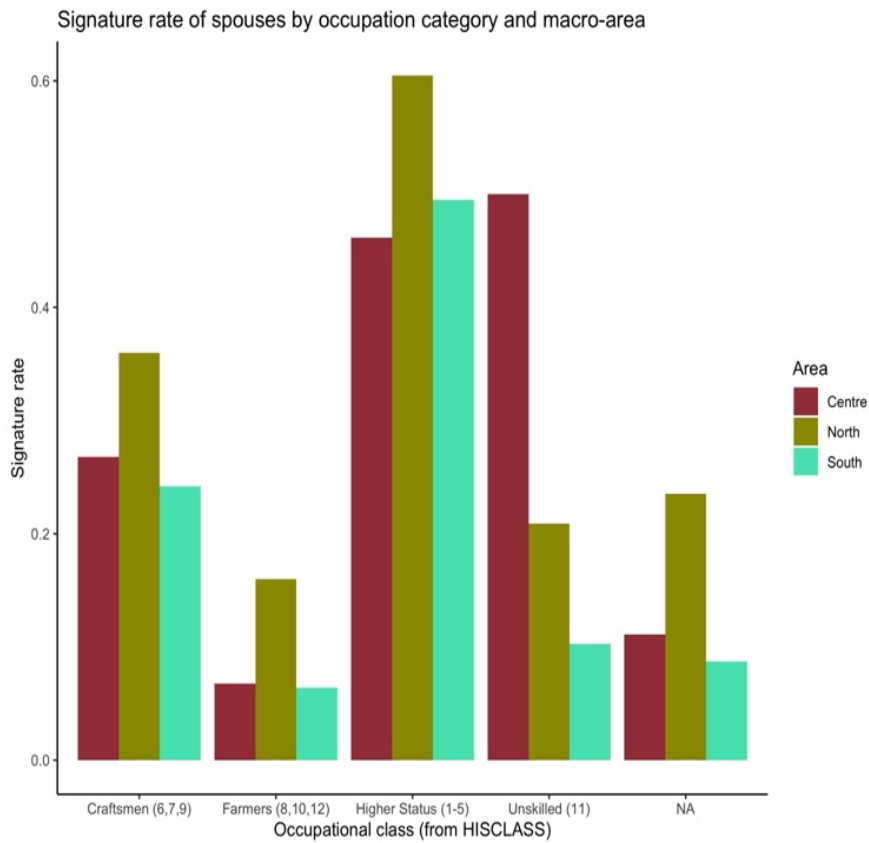
Sources: ASI (1853), ASI (1964), Castiglioni (1962)

Figure 9. Literacy rates by source and area, including additional information from Ministero dell'Istruzione (1890).



The dashed lines indicate the new series linking this sample's estimates for 1815 with the ones of the 1861 and 1871 unitary censuses, using the 19+ and 30-40 age cohorts, respectively.

Figure 10. Histogram showing the percentage of spouses able to sign by broad socio-occupational class, by macro-area.



36

Figure 1.: Histogram showing the percentage of spouses able to sign by broad socio-occupational class, by gender.

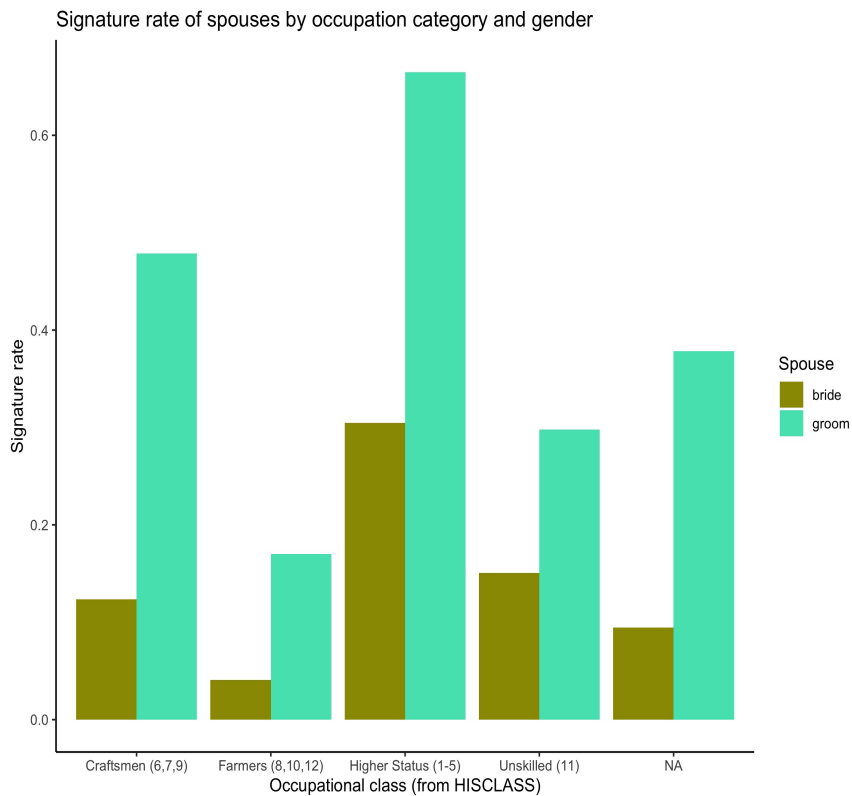


Table 4 Post-stratification of Marriage certificates, by Population of the Area and by City Size.

Region-Province	Collected certificates			Minimum Cell Size		
	Small Urban	Large	Rural	Small Urban	Large	Rural
Abruzzo Citra	0	2	16	1.6	2.1	12.9
Abruzzo Ultra I	0	0	14	0.7	0.0	10.4
Abruzzo Ultra II	0	0	20	0.9	1.2	14.5
Basilicata	0	2	13	0.9	1.0	11.3
Calabria Ultra I	0	1	18	0.7	1.0	16.8
Calabria Ultra II	0	1	16	0.0	0.8	15.2
Catania	1	3	18	1.4	6.1	13.9
Emilia-Romagna	3	23	95	4.1	22.6	91.3
Girgenti	0	8	8	0.0	4.8	9.5
Liguria	16	0	34	3.0	6.6	31.7
Lombardia	7	37	153	7.1	32.4	151.7
Marche	0	19	35	6.8	7.2	39.5
Messina	2	7	12	1.6	7.2	9.8
Molise	0	0	25	3.0	0.0	15.9
Napoli	7	43	21	7.1	28.4	9.9
Noto	0	0	12	0.8	2.0	11.2
Palermo	0	15	18	0.4	16.0	15.0
Piemonte	10	13	139	8.9	13.6	135.4
Principato Citeriore	15	1	18	0.6	0.7	28.0
Principato Ulteriore	0	0	35	0.0	1.9	33.5
Terra d'Otranto	0	7	22	1.3	4.6	15.7
Terra di Bari	0	9	18	0.0	7.1	17.1
Terra di Lavoro	2	2	25	3.5	1.9	27.3
Toscana	19	17	55	6.3	16.0	63.6
Trapani	0	8	5	0.4	5.4	4.9
Veneto	2	34	114	3.2	30.4	100.3
Total	84	252	959	64.5	221.1	906.2

* Each cell contains a number of certificates. Small urban areas have a population of less than 9,000 in 1800, and large urban areas have a population of more or as much as 9,000 in 1800.

† Group 1 contains the certificates collected in the post-stratified sample. If the number of collected certificates is zero in a cell where it should be positive, it means that there is no such a category in the list of villages.

‡ Group 2 contains the minimum number of certificate required in each to have a representative sample.

Table 5 Literacy rates obtained through logit regressions (in %)

Area	Sector	Men		Women	
		Urban	Rural	Urban	Rural
North	Industry	60.37 (15.29)	53.02 (13.26)	21.46 (15.88)	16.83 (13.85)
	Agriculture	28.23 (16.29)	22.56 (10.08)	6.59 (19.30)	4.97 (14.35)
	Services	80.16 (19.14)	74.96 (16.38)	42.03 (20.22)	34.94 (17.56)
Centre	Industry	39.36 (22.59)	32.47 (23.41)	10.43 (23.69)	7.94 (24.42)
	Agriculture	14.36 (23.01)	11.04 (21.48)	2.92 (25.86)	2.18 (24.46)
	Services	63.26 (24.80)	56.05 (24.76)	23.60 (26.27)	18.62 (26.18)
South	Industry	39.01 (15.12)	32.15 (13.91)	10.29 (17.41)	7.83 (16.30)
	Agriculture	14.17 (17.44)	10.90 (12.78)	2.88 (21.62)	2.15 (18.00)
	Services	62.92 (18.58)	55.69 (16.44)	23.33 (21.07)	18.40 (19.14)

* Each cell represents the predicted value of literacy rates at the margin of each level of each model term (gender, urban-rural, skill, area).

† Standard errors (in %) are in parentheses.

Table 6 Unadjusted literacy rates by gender, location, and occupation category (in %) from the post-stratified sample.

Macro Area	Summary		Gender		Urban-Rural		Occupation (HISCLASS)				GPI (%)
	N	Average Literacy	Male	Female	Urban	Rural	High Status	Craftsmen	Farmers	Unskilled	
North	1295	24.86 (1.20)	38.27 (1.91)	11.44 (1.25)	41.18 (3.20)	21.19 (1.26)	60.47 (4.32)	35.96 (3.19)	15.99 (1.35)	20.87 (3.81)	29.89
Centre	274	15.69 (2.20)	21.90 (3.55)	9.49 (2.51)	27.17 (4.66)	9.89 (2.22)	46.15 (9.97)	26.79 (5.97)	6.78 (1.89)	50.00 (22.36)	43.33
South	1061	15.08 (1.10)	23.24 (1.85)	7.09 (1.11)	17.62 (2.44)	14.32 (1.23)	49.48 (5.10)	24.19 (2.72)	6.38 (1.13)	10.26 (4.92)	30.50
Italy	2630	19.96 (0.78)	30.53 (1.27)	9.47 (0.81)	28.92 (1.89)	17.46 (0.84)	54.76 (3.14)	29.51 (1.98)	11.55 (0.86)	19.38 (3.13)	31.01
N	2,630	2,630	1,310	1,320	574	2,056	553	833	1,686	461	

* The occupation categories correspond to the following HISCLASS codes: Higher Status for HISCLASS 1-5; Craftsmen for HISCLASS 6, 7, and 9; Farmers for HISCLASS 8, 10, and 12; Unskilled for HISCLASS 11.

** GPI stands for Gender Parity Index: the ratio of female over male literacy rates.

*** Standard errors are in parentheses.

Table 7 Literacy rates from marriage certificates and local studies, by occupation category.

Region-Province	Landowners		Primary		Secondary		Tertiary	
	Local	Sample	Local	Sample	Local	Sample	Local	Sample
Lombardia	0.65	0.64	0.19	0.16	0.52	0.55	0.60	0.47
Marche	0.78	0.50	0.07	0.05	0.37	0.20	0.48	0.50

* Comparison of literacy rates among occupational structures estimated in various local studies and in this sample, at the regional level. I converted the HISCO and HISCLASS classification into primary, secondary, tertiary, and landowners sectors to follow the classification of most local studies.

Table 8 Share of male spouses in the different occupational groups, in local studies and in this sample.

Region-Province	N		% Landowners		% Primary		% Secondary		% Tertiary	
	Local	Sample	Local	Sample	Local	Sample	Local	Sample	Local	Sample
Lombardia	1,162	193	18.59	5.70	51.64	59.07	19.10	27.46	10.67	7.77
Marche	786	56	30.41	3.57	25.70	75.00	27.61	17.86	16.28	3.57

* I converted the HISCO and HISCLASS classification into primary, secondary, tertiary, and landowners sectors to follow the classification of most local studies.

Table 9. Literacy rates by gender and location, Antenati sample, 15-30 cohort

Macro Area	Summary		Gender		Urban-Rural	
	Average Literacy	N	Male	Female	Urban	Rural
Centre	12.32	211	20.43	5.93	25.71	5.67
North	24.23	986	39.91	11.30	42.51	20.51
South	14.80	824	22.56	7.53	16.30	14.37
Italy	19.14	2021	30.60	9.23	28.27	16.75
N	2021	2021	938	1083	421	1600

* This balanced sample only considers spouses aged 15 to 30 [15-29 included] in around 1815.

Table 10 Literacy rates by gender and location, Antenati sample, 30-40 cohort

Macro Area	Summary		Gender		Urban-Rural	
	Average Literacy	N	Male	Female	Urban	Rural
Centre	22.92	48	25.00	16.66	29.41	19.35
North	19.78	182	25.41	8.33	23.68421	18.75
South	17.56	148	26.83	6.06	28.20513	13.76
Italy	19.31	378	25.83	7.97	26.59574	16.90
N	378	378	240	138	94	284

* This balanced sample only considers spouses aged 30 to 40 [30-39 included] in around 1815, excluding second marriages.

Appendix 2. List of sources for literacy rates

Primary sources

ASI – Annuario Statistico Italiano (1853) Annuario Economico-Statistico dell'Italia (Vol. 1027).
Turin, Ferrero e Franco.

Antenati – Ministero per i beni e le attività culturali e per il turismo (2022) In: Buzzi AM, Vitali S, Corbellini R, Corbo A, Ferrari D, Gagliardoni L, Garis E, Maestrini F, Salemme F, Vedovotto N, and Nicoletti R (Eds.), Antenati <http://dl.antenati.san.beniculturali.it/gallery>
Accessed 24 Set 2022

MAIC – Ministero dell'Agricoltura, dell'Industria e del Commercio (1866a) Censimento generale della popolazione del 31 dicembre 1861 Vol. II. Tipografia Letteraria, Florence

MAIC – Ministero dell'Agricoltura, dell'Industria e del Commercio (1866b) Censimento generale della popolazione del 31 dicembre 1861 Vol. III. Tipografia Letteraria Florence

- MAIC – Ministero dell’Agricoltura, dell’Industria e del Commercio (1875) Censimento generale della popolazione del 31 dicembre 1871 Vol. II. Tipografia Cenniniana, Rome.
- MAIC – Ministero dell’Agricoltura, dell’Industria e del Commercio (1883) Censimento della popolazione del Regno d’Italia al 31 dicembre 1881. Vol. III. Tipografia Bodoniana, Rome
- MAIC – Ministero dell’Agricoltura, dell’Industria e del Commercio, Direzione Generale di Statistica (1900) Studi e proposte per l’esecuzione del IV censimento generale della popolazione del Regno. Tipografia Nazionale di G. Bertero, Rome
- Ministero dell’Istruzione (1890) Delle condizioni della istruzione elementare in Italia e del suo progresso dal 1861 in poi. Stabilimento Tipografico Sinimberghi, Rome

Secondary sources

- Correnti C, Maestri G (1864) Annuario Statistico Italiano. Ministero dell’Industria, Agricoltura e Commercio, Rome
- Chilosi D, Ciccarelli C (2021) Southern and Northern Italy in the Great Divergence: New Perspectives from the Occupational Structure. Bank of Italy Economic History Working Paper 47
- Curtis D (2013) Is there an “agro-town” model for Southern Italy? Exploring the diverse roots and development of the agro-town structure through a comparative case study in Apulia. *Continuity and Change* 28:377. <https://doi.org/10.1017/S0268416013000362>
- Griseri G (1973) L’istruzione primaria in Piemonte: 1831-1856. Deputazione subalpina di storia patria, Turin
- King R, Strachan A (1978) Sicilian Agro-Towns (Sizilianische Agro-Städte). *Erdkunde*, 110–123.
- Brambilla E (1991) Alfabetismo e società nelle Marche in Età Napoleonica. In: Bartoli Langeli A, Toscani X. (eds.) Istruzione, alfabetismo, scrittura. Saggi di storia dell’alfabetizzazione in Italia (sec. XV-XIX). Franco Angeli, Rome
- Castiglioni P (1862) Introduzione storica sopra i censimenti della popolazione italiana. Ministero dell’Industria, Agricoltura e Commercio, Rome
- Ciccarelli C, Weisdorf J (2019) Pioneering into the past: Regional literacy developments in Italy before Italy. *European Review of Economic History* 23:329–364. <https://doi.org/10.1093/ereh/hey014>
- Ferraresi A (1991) Alfabeti e analfabeti nell’Oltrepò pavese alla fine dell’Antico regime: Voghera, Stradella, Casteggio, Casatisma. In: Bartoli A, Toscani X (eds.) Istruzione, alfabetismo, scrittura. Saggi di storia dell’alfabetizzazione in Italia (sec. XV-XIX). Franco Angeli, Rome, 245–331

- Malanima P (2016). Italian Urban Population 1300-1861 (The Database).
- Mariella V, Postigliola M, and Rota M (2020) A Reconstruction of the Italian Provincial Population. *Rivista di Storia Economica* 36:1–39. <https://doi.org/10.1410/98355>
- Milanesi A (1991) La città e il borgo: un caso padano. In: Bartoli Langeli A, Toscani X. (eds.) Istruzione, alfabetismo, scrittura. Saggi di storia dell'alfabetizzazione in Italia (sec. XV-XIX). Franco Angeli, Rome, 218–244
- Petrucchi A (1987) Prospettive di Ricerca e Problemi di Metodo per una Storia Quantitativa dell'Alfabetismo. In: Pellizzari MR, Sulle vie della scrittura: alfabetizzazione, cultura scritta e istituzioni in età moderna. Atti del convegno di Studi del centro studi “Antonio Genovesi” per la storia economica e sociale. Edizioni Scientifiche Italiane, Salerno, 21–27
- Piseri M (2002) L'alfabeto delle riforme: Scuola e alfabetismo nel basso Cremonese da Maria Teresa all'Unità. Vita e Pensiero, Milan
- Salvo C, Zichichi L (2003) La Sicilia dei signori: il potere nelle città demaniali. Sellerio, Palermo
- Scirocco A (1987) Note sull'alfabetizzazione a Napoli nell'Ottocento attraverso i Registri Matrimoniali. In Sulle vie della scrittura: alfabetizzazione, cultura scritta e istituzioni in età moderna. Atti del convegno di Studi del centro studi “Antonio Genovesi” per la storia economica e sociale, ed. M.R. Pellizzari, Edizioni Scientifiche Italiane, Salerno, 259–291
- Toscani X (1987). L'alfabetismo nelle campagne dei Dipartimenti del Mincio e del Mella e nelle Alte Valli del Serio e dell'Adda. In: Pellizzari MR, Sulle vie della scrittura: alfabetizzazione, cultura scritta e istituzioni in età moderna. Atti del convegno di Studi del centro studi “Antonio Genovesi” per la storia economica e sociale. Edizioni Scientifiche Italiane, Salerno, 549–609
- Toscani X (1993) Scuole e alfabetismo nello Stato di Milano da Carlo Borromeo alla Rivoluzione. La Scuola, Brescia

Appendix 3. Bottom-up illustration of the sampling strategy

This section is intended to clarify the precise steps taken to obtain a sample that is representative at the regional level in the Centre-North and at the provincial level in the South, starting from the village level and going up to the regional level (see also figure 12). This is just intended as a conceptual illustration. For technical details please see the supplemental material. Here I consider the provinces of Naples and the region of Lombardia to illustrate the steps. the pre-unitary province of Naples there are 81 villages in the province, and 18 neighborhoods of the city. The process goes as follows.

First, I count the average number of marriage certificates that there are in 1815 in each neighborhood or village: 140.8 marriages per neighborhood in Naples, and 57.85 marriages in the province.²¹ Second, I look at the number of archives in the province of interest: is there more than one archive? In the case of Naples, the city and provincial data are collected by the same state archive of Naples, but the number of marriage certificates varies considerably in the two areas, so I treat them as separate archival units.²² In this step, I calculate the weight one, the relative weight assigning the number of certificates to collect from each archive within a region: $0.71 * \frac{140.8}{(140.8+57.85)}$ for the city of Naples, and 0.29 for the province.²³ Third, I look at the population sources (primarily mariella2020reconstruction) to know how much population there overall in the province of Naples in 1815 (677,389), and the overall population in the parts of Italy covered in this sample 16,916,109. The ratio is the population weight, $(\frac{677,389}{16,916,109})=0.04$. According to available source ASI (1853), the number of marriages per 100 inhabitants in the kingdom of Naples was 0.72, so the marriage certificates equivalent of the population size in Naples is 4877,20. The minimum sample size in terms of marriage certificates is 1,216, so we need a total of $1,216*(0.04)=48.64$ certificates in Naples: according to weight 1, 34.53 in the city, and 14.11 in the province. In terms of population, weight 1 also tells us that the population in the city of Naples was $677,389*0.71=480,946$ and 199,442 in the province.

Fourth, to account for the share of urban and rural population within both the city of Naples and its province, I look at the 1800 figures from Malanima (2016). The city of Naples is entirely urban. In the province, according to Malanima (2016), in 1800 the total urban population size was 120,000, of which *Afragola*, *Torre Annunziata*, and *Pozzuoli* counting as large cities with more or as much as 9,000 inhabitants and totalling 35,000 inhabitants. I assign weights to each of the large urban, small urban, and rural areas in the province of Naples accordingly, and sample from each the required number of certificates.

The process is very similar in the region of Lombardy, and for this reason it is applied recursively over the remaining provinces and regions. The only difference is that Lombardy, as other regions in the North, are represented by fewer archives than their population would suggest: the archives of Como, Brescia, and Mantova have provided the scanned certificates, but the archives of Milano and Bergamo did not. The population weight ensures that the region is represented well in terms of total number of certificates, but sampling biases may arise if the available archives are systematically

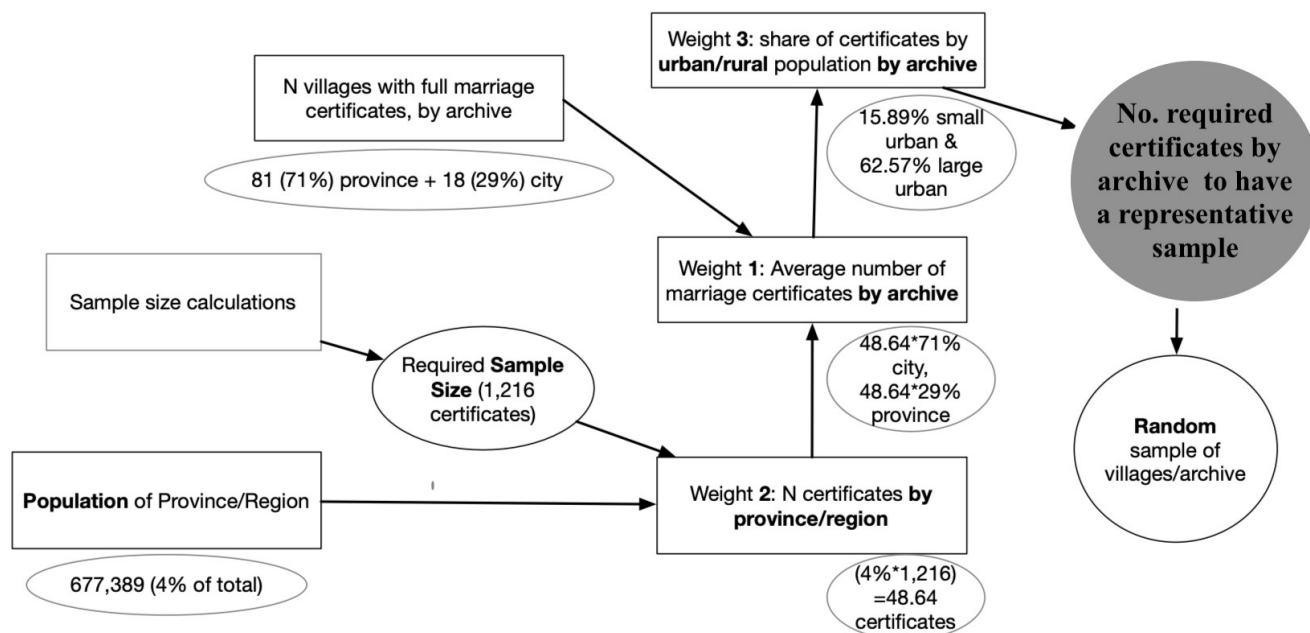
²¹ When the number of villages was very large, as in the case of Udine, I took a random sample of them, covering at least 25% of the considered villages.

²² I also treated other large urban areas separately from their province (Palermo and Turin).

²³ In many cases, there was only one archive serving a province, so this weight did not matter.

different than the unsampled ones. For Lombardy, this is partially the case because excluding Milan means to exclude the most dynamic, and probably literate, area in Lombardy. This bias is partly accounted for by including the large city weight that samples more from other larger urban areas and takes these additional observations as substitutes for the ones that are missing for Milan. In other regions, the selected archives include a fair share of large cities as well as more remote areas.

Figure 12. Infographic illustrating the sampling and post-stratification approach.



Round boxes provide example figures for the province of Naples. Unlike most other cases, even though the State archive is just one, the province of Naples was treated as composed by two separate entities: city and province. The city of Naples is entirely urban. In the province, according to Malanima (1016), in 1800 the total urban population size was 120,000, of which Afragola, Torre Annunziata, and Pozzuoli counting as large cities with more or as much as 9,000 inhabitants, and totalling 35,000 inhabitants.

References

A'Hearn B, Auria C, Vecchi G (2011) Istruzione. In: Vecchi G (ed.) *In Ricchezza e in Povertà.*

Il Benessere degli Italiani dall'Unità a Oggi. Il Mulino, Bologna

A'Hearn B (1998) Institutions, externalities, and economic growth in southern Italy: evidence from the cotton textile industry, 1861-1914. *The Economic History Review* 51:734–762.

<https://doi.org/10.1111/1468-0289.00112>

A'Hearn, B, Delfino A, Nuvolari A (2020) Cognition, Culture, and State Capacity: Age-Heaping in XIX Century Italy, mimeo.

A'Hearn B, Delfino A, Nuvolari A (2022) Rethinking age heaping: a cautionary tale from nineteenth-century Italy. *The Economic History Review* 75:111–137.

<https://doi.org/10.1111/ehr.13087>

Álvarez B, Palencia FR (2018) Human capital and earnings in eighteenth-century Castile.

Explorations in Economic History 67:105–133. <https://doi.org/10.1016/j.eeh.2017.10.005>

Astoul G (1992) L'alphabétisation en Haut-Languedoc au XVIIIe siècle: les aléas de la

méthode Maggiolo dans le diocèse de Montauban. *Annales du Midi* 104.198: 175–194

Banti AM (1996) *Storia della borghesia italiana: l'età liberale* (Vol. 1). Donzelli Editore, Rome

Benigno F (2014) *L'età moderna: dalla scoperta dell'America alla Restaurazione*. Laterza, Rome-Bari

Berengo M (1974) La città di antico regime. *Quaderni storici* 661–692

Bianchi A (2019a) L'istruzione in Italia tra Sette e Ottocento. Dal Regno di Sardegna alla Sicilia

Borbonica: istituzioni scolastiche e prospettive educative. *La Scuola*, Brescia

Bianchi A (2019b) La scuola primaria nell'alto Piemonte-Valle d'Aosta Napoleonici. I dipartimenti della Dora e della Sesia. In: Angelo B (Ed.), *L'istruzione in Italia tra Sette e Ottocento. Dal Regno di Sardegna alla Sicilia Borbonica: istituzioni scolastiche e prospettive educative*. *La Scuola*, Brescia

- Blok A (1969) South Italian agro-towns. *Comparative Studies in Society and History* 11:121–135.
<https://doi.org/10.1017/S0010417500005235>
- Bruce PA (1910) *Institutional History of Virginia in the Seventeenth Century (Vol. 1)*. GP Putnam's sons, New York-London
- Cappelli G (2017a) The missing link? Trust, cooperative norms, and industrial growth in Italy. *Journal of Interdisciplinary History* 47:333–358,
https://doi.org/10.1162/JINH_a_01014
- Cappelli G (2017b) Was Putnam Wrong? The Determinants of Social Capital in Italy around 1900. *Rivista di Storia Economica* 33:277–314. <https://doi.org/10.1410/89359>
- Cappelli G, Vasta M (2020a) A “Silent Revolution”: school reforms and Italy’s educational gender gap in the Liberal Age (1861–1921). *Cliometrica* 15:1–27, <https://doi.org/10.1007/s11698-020-00201-6>
- Cappelli G, Vasta M (2020b) Can school centralization foster human capital accumulation? A quasi-experiment from early twentieth-century Italy. *The Economic History Review* 73:159–184.
<https://doi.org/10.1111/ehr.12877>
- Chilosi D (2007) Old wine in new bottles: civic nation-building and ethnic nationalism in schooling in Piedmont, ca. 1700–1861. *Nations and Nationalism* 13:417–436.
<https://doi.org/10.1111/j.1469-8129.2007.00294.x>
- Cipolla CM (1969) *Literacy and Development in the West (Vol. 1027)*. Penguin, London
- Clark G (2005) The condition of the working class in England, 1209–2004. *Journal of Political Economy* 113:1307–1340. <https://doi.org/10.1086/498123>
- Costa DL (1993) Height, weight, wartime stress, and older age mortality: evidence from the Union Army records. *Explorations in Economic History* 30:424–449.
<https://doi.org/10.1006/exeh.1993.1018>

- Costa DL (2015) Health and the Economy in the United States from 1750 to the Present. *Journal of Economic Literature* 53:503–70. <https://doi.org/10.1257/jel.53.3.503>
- Crayen D, Baten J (2010) Global trends in numeracy 1820–1949 and its implications for long-term growth. *Explorations in Economic History* 47:82–99.
<https://doi.org/10.1016/j.eeh.2009.05.004>
- Cvrček T (2020) *Schooling under control: The origins of public education in Imperial Austria 1769–1869 (Vol. 1)*. Mohr Siebeck, Tübingen
- Daniele V, Malanima P (2011) *Il Divario Nord-sud in Italia, 1861–2011 (Vol. 273)*. Rubbettino Editore, Catanzaro
- Dickinson RE (1954) Land reform in southern Italy. *Economic Geography* 30:157–176.
doi: 10.2307/142101
- Engelsing R (1973) *Analphabetentum und Lektüre: Zur Sozialgeschichte des Lesens in Deutschland zwischen feudaler und industrieller Gesellschaft*. J. B. Metzler, Stuttgart
- Federico G, Dincecco M (2021) Napoleon in Italy: a legacy of institutional reform? In: O'Brien P (ed.), *The Crucible of Revolutionary and Napoleonic Warfare and European Transitions to Modern Economic Growth*. Brill, Leiden, 141–170
- Federico G, Nuvolari A, Ridolfi L, Vasta M (2019) The race between the snail and the tortoise: skill premium and early industrialization in Italy (1861–1913). *Cliometrica* 15:1–42.
<https://doi.org/10.1007/s11698-019-00200-2>
- Federico G, Nuvolari A, Vasta M (2019) The origins of the Italian regional divide: Evidence from real wages, 1861–1913. *The Journal of Economic History* 79:63–98.
<https://doi.org/10.1017/S0022050718000712>
- Felice E (2013) *Perché il Sud è rimasto indietro*. Il Mulino, Bologna
- Felice E (2014) *Il Mezzogiorno fra storia e pubblicistica. Una replica a Daniele e Malanima*. *Rivista di Storia Economica* 30:197–242

- Felice E (2019) Regional income inequality in Italy in the long run (1871-2001). Patterns and determinants. In: Roses J, Wolf N (eds.), *The Economic Development of Europe's Regions. A Quantitative History Since 1900*. Routledge, London, 177–203
- Felice E, Vasta M (2015) Passive modernization? The new human development index and its components in Italy's regions (1871–2007). *European Review of Economic History* 19: 44–66. <https://doi.org/10.1093/ereh/heu018>
- Fleury M, Valmary P (1957) Les progrès de l'instruction élémentaire de Louis XIV à Napoléon III, d'après l'enquête de Louis Maggiolo (1877-1879). *Population* 71–92
- Gargano A (2012) Numeri in dubbio. Scuola pubblica e scuola privata nell'Italia meridionale attraverso l'inedita inchiesta del 1861. In: Bianchi A (ed.), *L'istruzione in Italia. La Scuola* Brescia, 761–802
- Genovesi G (2010) *Storia della scuola in Italia dal Settecento a oggi (Vol. 1)* Laterza, Rome-Bari
- Guiso L, Sapienza P, Zingales L (2016) Long-term persistence. *Journal of the European Economic Association* 14:1401–1436. <https://doi.org/10.1111/jeea.12177>
- Houston RA (2014) *Literacy in early modern Europe*. Routledge, London
- Hoyler M (1998) Small town development and urban illiteracy: comparative evidence from Leicestershire marriage registers 1754-1890. *Historical Social Research/Historische Sozialforschung*, 202–230
- Lemercier C, Trivellato F (2022) 1751 and Thereabout: A quantitative and comparative approach to notarial records. *Social Science History* 46:555–583. <https://doi.org/10.1017/ssh.2022.8>
- Lupo M (2005) *Tra le provvide cure di sua maestà: stato e scuola nel Mezzogiorno tra Settecento e Ottocento*. Il Mulino, Bologna
- Lupo M, Gargano A, Marra A (2014) *Collezione delle Leggi de' Decreti e di altri atti riguardante la Pubblica Istruzione promulgati nel già Reame di Napoli dall'anno 1806 in poi*. ISMED - Consiglio Nazionale delle Ricerche (CNR), Naples
- Macry P (2002) *Ottocento. Famiglia, élites e patrimoni a Napoli*. Il Mulino, Bologna

- Main GL (1991) An inquiry into when and why women learned to write in colonial England. *Journal of Social History* 24:579–589. <https://doi.org/10.1353/jsh/24.3.579>
- Malanima P, Daniele V (2007) Il prodotto delle regioni e il Divario Nord-Sud in Italia (1861-2004). *Rivista di Politica Economica* 97:1–49
- Marchesini D (1992) Il bisogno di scrivere: usi della scrittura nell'Italia moderna. Laterza, Rome-Bari
- Mariella V (2022) The agrarian origins of social capital. *Journal of Economic Behavior & Organization* 193(1): 543-568. <https://doi.org/10.1016/j.jebo.2021.11.029>
- Marzeddu R (2019) Fonti per la storia dell'istruzione in piemonte tra sette e ottocento. Spunti e itinerari di ricerca. In: Angelo B (ed.) *L'istruzione in Italia tra Sette e Ottocento. Dal Regno di Sardegna alla Sicilia Borbonica: istituzioni scolastiche e prospettive educative. I-Studi; II-Carte Storiche*. La Scuola, Brescia, 57–75
- Mastrangelo G (2008). *Le "Scuole Reggimentali", 1848-1913. Cronaca di una forma di istruzione degli adulti nell'Italia liberale*. Ediesse, Rome
- Mastroluca S, Verrascina M (2012) L'evoluzione dei contenuti informativi del censimento della popolazione. *Annali di statistica* 12
- Meier zu Selhausen F, Van Leeuwen MH, Weisdorf JL (2018). Social mobility among Christian Africans: evidence from Anglican marriage registers in Uganda, 1895–2011. *The Economic History Review* 71:1291–1321. <https://doi.org/10.1111/ehr.12616>
- Pellizzari MR (1987) Alfabeto e fisco. Tra cultura scritta e oralità nel Regno di Napoli a metà Settecento. In: Pellizzari MR, *Sulle vie della scrittura: alfabetizzazione, cultura scritta e istituzioni in età moderna. Atti del convegno di Studi del centro studi "Antonio Genovesi" per la storia economica e sociale*. Edizioni Scientifiche Italiane, Salerno, 99–152
- Pellizzari M (2000) *La penna e la zappa. Alfabetizzazione, cultura e generi di vita nel Mezzogiorno moderno*. Laveglia e Carlone editore, Salerno

- Piseri M (2012) *L'alfabeto in montagna. Scuola e alfabetismo nell'area alpina tra età moderna e XIX secolo: Scuola e alfabetismo nell'area alpina tra età moderna e XIX secolo.* FrancoAngeli, Rome
- Piseri M (2017) *La scuola primaria nel Regno italico: 1796-1814.* FrancoAngeli, Rome
- Postigliola M, Rota M (2020). Institutions and literacy rates: the legacy of Napoleonic reforms in Italy. *European Review of Economic History*, 25:757-779.
<https://doi.org/10.1093/ereh/hea021>
- Postigliola M, Rota M (2021) Institutional shocks and social capital: evidence from the French domination in Italy. *Journal of Modern Italian Studies* 1:1–24,
<https://doi.org/10.1080/1354571X.2021.1912251>
- Poussou JP (1993) La méthode Maggiolo et la mesure de l'alphabétisation du Sud-Ouest de la France à la fin du XVIIIème siècle. *Annales du Midi* 105:209–223
- Pruneri F (2011) *L'istruzione in Sardegna 1720-1848.* Il Mulino, Bologna
- Putnam RD, Leonardi R, Nanetti RY (1994) *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton University Press, Princeton
- Quiroga G (2003) Literacy, education and welfare in Spain (1893–1954). *Paedagogica Historica* 39:599–619. <https://doi.org/10.1080/0030923032000128818>
- Rachal JR (1987) Measuring English and American historical literacy: A review of methodological approaches. *International Journal of Lifelong Education* 6:185–198.
<https://doi.org/10.1080/0260137870060302>
- Rediker M (1989) *Between the Devil and the Deep Blue Sea: merchant seamen, pirates and the Anglo-American maritime world, 1700-1750.* Cambridge University Press, Cambridge
- Rettaroli R (1990) Age at marriage in nineteenth-century Italy. *Journal of Family History* 15:409–425. <https://doi.org/10.1177/036319909001500123>

- Rossi N, Toniolo G, Vecchi G (2001) Is the Kuznets curve still alive? Evidence from Italian household budgets, 1881–1961. *The Journal of Economic History* 61:904–925.
doi:10.1017/S0022050701042024
- Sargant WL (1867) On the progress of elementary education. *Journal of the Statistical Society of London* 30:80–137
- Shaw-Taylor L, Wrigley EA (2014) Occupational structure and population change. In: Floud R, Johnson R (eds.) *The Cambridge Economic History of Modern Britain (Vol. 1)* Cambridge University Press, Cambridge 53-88
- Schofield RS (1973) Dimensions of illiteracy, 1750-1850. *Explorations in Economic History* 10:437–454. [https://doi.org/10.1016/0014-4983\(73\)90026-0](https://doi.org/10.1016/0014-4983(73)90026-0)
- Schofield RS (1968) The measurement of literacy in pre-industrial England. *Literacy in traditional societies* 10:310–26
- Terzi L (2001) *Le scuole Normali a Napoli tra Sette e Ottocento. Documenti e ricerche sulla “pubblica uniforme educazione” in antico regime*, L’Orientale Editrice, Naples
- Van Leeuwen MH, Maas I, Miles A (2002) *HISCO: Historical International Standard Classification of Occupations*. Leuven University Press, Leuven
- Van Leeuwen MH, Maas I (2011) *HISCLASS: A Historical International Social Class Scheme*. Leuven University Press, Leuven
- Vasta M (1999) Capitale umano e ricerca scientifica e tecnologica. In: Amatori F, Bigazzi D, Giannetti R, Segreto L (eds.), *Storia d’Italia, Annali*. Einaudi, Turin, 1041–124
- Vigo G (1971) *Istruzione e sviluppo economico in Italia nel secolo XIX*. ILTE, Turin
- Vigo G (2017) *Il vero sovrano dell’Italia: l’istruzione degli adulti nell’Italia dell’Ottocento*. Il Mulino, Bologna
- Zamagni V (1993) *The Economic History of Italy 1860-1990*. Clarendon Press, Cambridge
- Zazo A (1927) *L’istruzione pubblica e privata nel Napoletano (1767-1860)*. Il Solco, Città di Castello.

