

STATE CAPACITY AND THE UNEVEN  
COST OF NATION BUILDING:  
LANGUAGE MISMATCH AND  
LITERACY LEVELS IN VALENCIA

Francisco J. Beltrán Tapia

Alfonso Díez-Minguela

Alicia Gómez-Tello

Julio Martinez-Galarraga

Daniel A. Tirado-Fabregat

---

*UB Economics Working Paper No. 414*

**Title:** State capacity and the uneven cost of nation building: Language mismatch and literacy levels in Valencia

**Abstract:** This paper studies the impact generated by the existence of a mismatch between the language of instruction and the language of use of the population in the context of the construction of the liberal state in Spain. In particular, the work analyzes the effects of the presence of this linguistic distance on the unequal diffusion of literacy among the municipalities that made up the former Kingdom of Valencia from 1860 to 1930. For the development of the analysis, a novel data set has been constructed with information that includes the literacy rates of the 524 municipalities that make up the region of Valencia (Valencian Community) in three points in time (1860, 1900 and 1930), the linguistic domain to which each municipality belongs, as well as the institutional, geographic and economic characteristics of each municipality at the end of the Ancien Régime (1787). Based on the available information, the analysis uses Propensity Score Matching techniques to verify the existence of an effect on the literacy levels recorded in Spanish-speaking municipalities with respect to Catalan-speaking ones. Two main results are obtained. The first is to identify the existence of differences in educational outcomes derived from the presence of a mismatch effect. Secondly, it is also shown that this effect only appears when the Spanish state enjoyed the capacity to force compliance with language regulations in public schools, in parallel with the advance of its financial and administrative capacity and the incipient advance of a democratic regime.

**JEL Codes:** I28, O17, N33, N43, J24

**Keywords:** Language mismatch, literacy, regional inequality, economic history, Valencia

---

UB Economics *Working Paper No. 414*

**Authors:**

|  |                         |                         |
|--|-------------------------|-------------------------|
| Francisco J. Beltrán Tapia                               | Alfonso Díez-Minguela   | Alicia Gómez-Tello      |
| Norwegian University of Science<br>and Technology (NTNU) | Universitat de València | Universitat de València |

**Email:**

francisco.beltran.tapia@ntnu.no

**Email:**

alfonso.diez@uv.es

**Email:**

alicia.gomez-tello@uv.es

Julio Martinez-Galarraga  
Universitat de Barcelona

Daniel A. Tirado-Fabregat  
Universitat de València

**Email:**

julio.martinez@ub.edu

**Email:**

daniel.tirado@uv.es

**Date:** November 2021

**Acknowledgements:** The authors would like to thank the Bank of Spain for its financial support through the “Programa de Ayudas a la Investigación en Macroeconomía, Economía Monetaria, Financiera y Bancaria, e Historia Económica”. Financial support from the Spanish Ministry of Science and Innovation, Projects ECO2015-65049-C12-1-P, ECO2017-90848-REDT, HAR2015-64076-P, PGC2018-095821-B-I00, PGC2018-095529-B-I00 (MCIU/AEI/FEDER, UE), the Generalitat Valenciana (AICO/2018/1930 and PROMETEO/2020/083), FBBVA (Ayudas a equipos de investigación en Humanidades Digitales) and the Norwegian Research Council, Project 249723, is also gratefully acknowledged.

## **State capacity and the uneven cost of nation building: language mismatch and literacy levels in Valencia**

Francisco J. Beltrán Tapia (NTNU)

Alfonso Díez-Minguela (Universitat de València)

Alicia Gómez-Tello (Universitat de València and IVIE)

Julio Martínez-Galarraga (Universitat de Barcelona)

Daniel A. Tirado-Fabregat (Universitat de València)

### **Abstract**

This paper studies the impact generated by the existence of a mismatch between the language of instruction and the language of use of the population in the context of the construction of the liberal state in Spain. In particular, the work analyzes the effects of the presence of this linguistic distance on the unequal diffusion of literacy among the municipalities that made up the former Kingdom of Valencia from 1860 to 1930. For the development of the analysis, a novel data set has been constructed with information that includes the literacy rates of the 524 municipalities that make up the region of Valencia (Valencian Community) in three points in time (1860, 1900 and 1930), the linguistic domain to which each municipality belongs, as well as the institutional, geographic and economic characteristics of each municipality at the end of the Ancien Régime (1787). Based on the available information, the analysis uses Propensity Score Matching techniques to verify the existence of an effect on the literacy levels recorded in Spanish-speaking municipalities with respect to Catalan-speaking ones. Two main results are obtained. The first is to identify the existence of differences in educational outcomes derived from the presence of a mismatch effect. Secondly, it is also shown that this effect only appears when the Spanish state enjoyed the capacity to force compliance with language regulations in public schools, in parallel with the advance of its financial and administrative capacity and the incipient advance of a democratic regime.

**Keywords:** Language mismatch, literacy, regional inequality, economic history, Valencia

**JEL Codes:** I28, O17, N33, N43, J24

## 1. Introduction

In this article we analyse the impact generated by the existence of a mismatch between the formative language and the language used by the population in the context of the construction of the liberal state in Spain. Particularly, the work analyses the effects of this linguistic distance in the uneven spread of literacy in the municipalities that made up the former kingdom of Valencia from 1860 to 1930.

By doing that, the study intends to contribute to two strongly connected strands of research. On the one hand, the so-called “nation building” literature. In this line of research, different works have pointed out that population heterogeneity in terms of ethnic fractionalization, language, religion or cultural values could harm economic development of societies by reducing the provision of public goods (Alesina and La Ferrara, 1999; Miguel 2004) or increasing the probability of political unrest (Alesina, Reich and Riboni, 2020). In these circumstances, states would have had an incentive to implement policies intended to homogenize the culture of population around national values. In this framework, Alesina and Reich (2014) and Alesina, Giuliano and Reich (2019) maintain that “homogeneity amongst people can be built with education, teaching a common language”, in a kind of benevolent homogenizing policy.<sup>1</sup> In order to illustrate this idea, these works offer qualitative evidence on the growing use by European liberal estates of compulsory primary instruction in a national language as a sort of national identity building policy, during the second half of the XIXth and the first third of the XXth centuries.<sup>2</sup> Additionally, they also point out to the fact that, during these years, the intensity of use of homogenization policies around a national identity increased with the threat that the extension of the franchise put in risk the political stability of States.

On the other hand, it has been also pointed out that the effectivity of linguistic homogenization policies grew in parallel to the financial and administrative capacity of the liberal States (Alesina, Giuliano and Reich, 2019). So, nation building and state capacity literatures appear strongly related. State capacity, understood as the capacity to implement official goals, especially over the actual or potential opposition of powerful social groups (Skocpol, 1985), would act as contingent

---

<sup>1</sup> The use of compulsory primary education, and in particular its implementation in a single language as a policy of homogenization of the population around national cultural values, is a widespread idea in the literature on the history of education, political science and sociology.

<sup>2</sup> In particular, the work of Alesina and Reich (2014) provides descriptive evidence about training in a national language in the cases of Italy and France.

of national homogenization policies. It is needed a State with high financial and administrative capacities in order to secure the success of the aforementioned policies.

However, the empirical analysis of the causes and consequences of the deployment of national homogenization policies, and in particular of educational policies, has faced the difficulty of identifying this type of relationship. This is why, as Alesina, Giulano and Reich (2019) point out, it constitutes an excellent line of research. In this sense, as regards the causes that drive the use of nation-building policies, it has been analyzed whether the generalization in the use of this type of policies is related to the level of heterogeneity of the individuals that make up a society. In this line, Bandiera et al. (2018) study the existence of a link between the temporal sequence in which compulsory schooling was implemented in North American states along the second half of the nineteenth century and the heterogeneity of their population, measured through the relative importance and particular characteristics of the immigrant population.

Regarding its effectiveness, empirical work has focused on the study of the results derived from the imposition of policies intended to the homogenization of the political preferences of citizens. Thus, Cinnirella and Schueler (2018) identify the impact of the use of schooling as a nation-building policy from the study of the existence of a relationship between public educational spending and voting for nationalist parties in nineteenth-century Prussia. Kersting and Wolf (2019), for their part, have studied the impact derived from this type of policy in the context of the construction of a German national identity, analyzing its incidence on the choice of a name of Germanic origin for children.<sup>3</sup>

In relation to these aspects, this paper proposes to identify the effectiveness of the use of a cultural homogenization policy, such as compulsory education in a national language, by means of the study of the educational cost generated by this policy in those territories where, historically, the language used by the population was not the national language. In what follows, it is argued that this approach allows us to identify the real effectiveness of this policy throughout the processes of national construction of the European liberal states of the second half of the nineteenth century and the first third of the twentieth century.

---

<sup>3</sup> From a more general perspective, another line of work has focused on the study of the effectiveness of this type of policies by analyzing their effects on the economic development of societies, measured through the quantification of existing differences in the provision of educational or sanitation infrastructure between countries that have made more or less intensive use of nation-building policies in the context of the decolonization processes in Africa (Miguel, 2004).

The existence of a training cost associated with linguistic homogenization has been well documented in the literature. From the supply-side perspective, its presence finds an economic foundation in works such as those by Alesina et al. (1999), that designed a theoretical framework with which to analyse the provision of certain public services in the presence of heterogeneous agents. This literature has pointed out that the rules governing the education system can affect the results achieved if they give rise to a mismatch or linguistic distance between the language of schooling and the population's language of use.<sup>4</sup> Particularly, it has been posed that this effect could be especially relevant in cases where education funding was a municipal-level responsibility and the local elites believed that the compulsory use of a language of schooling that was not their own posed a threat to their position of privilege. Cinnirella and Schueler (2016) show that the provision of primary education in Prussia was lower in those territories where German-speaking citizens lived alongside speakers of other Slavic languages (mainly Polish) in a context in which a Germanization policy had imposed German as the only language of schooling throughout the Empire after its foundation in 1871. Cvrcek and Zajicek (2013) find evidence that the literacy process followed the opposite dynamic in the Austro-Hungarian Empire. In this case, in territories where the language of schooling was the same as that of the local elites, education received greater financial support than in those where there was a distance between the two.

Other line of research, however, has stressed that the existence of a negative relationship between linguistic distance and education is a result of the effect the former can have on demand for the latter. Jain (2017) analyses the impact of the mismatch between the official language and the local language in districts in India and concludes that those in which a higher linguistic distance was present had significantly lower literacy rates than those where the language of use and the language of schooling were the same. The economic mechanism behind this relationship would be that being schooled in a language not their own requires individuals to make more of an investment. As shown in investigations like that by Williams and Cooke (2012), aimed at identifying the best strategies for bringing literacy to the population in recent development projects, education results are lower when these projects involve teaching in a language other than that of the potential beneficiaries. Similar conclusions are found when analysing the results

---

<sup>4</sup> Particularly, Alesina and La Ferrara (2005) carry out an in-depth survey of empirical works that analyse the existence of a relationship between the ethnolinguistic diversity of a territory and the provision of public goods.

achieved by immigrant groups in host countries. Bleakley and Chin (2004), for example, show that the education results for children who immigrate to the US from English-speaking territories are significantly better than those for children with different mother tongues. This type of evidence leads one to conclude that schooling in a language that is not the language of use calls for a greater investment than that needed when learning in one's own language.

With the aim of contributing to these debates, in this article we analyse the determinants of the uneven literacy levels in the municipalities of the region of Valencia between 1860 and 1930, using this territory as a case study in which to identify the impact caused by the State's establishment of an education system that created a linguistic distance between the language of schooling and the language of use in any municipality. Particularly, the choice of Valencia makes it easier to identify the impact caused by this homogenization policy for different reasons. Firstly, in Valencia two languages coexisted in the past and still do. A third of its municipalities are located in areas where *castellano* (henceforth Spanish) predominates, while two thirds are in the Catalan (locally known as Valenciano) language area. Secondly, linguistic diversity in this region was totally established during the Ancien Regime as far as it stems mainly from the Reconquista and the origins of the settlers who came to populate the different territories.<sup>5</sup> Thirdly, in this region the linguistic lines of demarcation are clear-cut (because apart from in certain cities, in the mid-nineteenth century the two languages did not coexist). Finally, although the introduction of regulations that laid down that the language to be used in Spanish primary schools came along other Bourbon reforms in the second half of the eighteenth century, it was the liberal state who increasingly enforced the accomplishment of this regulation by means, firstly, of a new Public Instruction Act passed in 1857 (Moyano Act) and, secondly, by the creation of the Ministry of Public Instruction in 1900. All in all, the study of the case of Valencia makes it possible to quantify the language mismatch effect and the changes it could experience along the different educational regimes implemented during the second half of the nineteenth century and the first years of the twentieth century.<sup>6</sup>

---

<sup>5</sup> In the case of lordships that were formerly inhabited by Moriscos, however, the language of use depends not on the origins of the settlers or their lords after the Reconquista, but on the origins of the new colonists who settled after 1609 (Casanova, 2001)

<sup>6</sup> The compulsory use of Spanish for teaching throughout the entire kingdom of Spain dates back to 1768, as established by Royal Decree by Carlos III.



To carry out the analysis we constructed a data set with information including the literacy rates recorded in the 524 municipalities that made up the region of Valencia (or *Comunidad Valenciana*) in three points in time (1860, 1900 and 1930), the language area in which each entity was located, the institutional and economic characteristics of the various municipalities at the end of the Ancien Regime (1787). Besides, also included is varied information on those elements that from a first- or second-nature geographical point of view could affect literacy levels and need to be taken into account in the analysis. Particularly, the dataset even provides information on settlement patterns in each local entity, i.e. the volume of population that lived in each of the local entities that made up the municipalities (towns, villages, hamlets, homesteads, rural dwellings, etc.), as well as the occupational structure of the municipalities.<sup>7</sup>

Considering the information available, the analysis brings two main empirical results. Firstly, it is shown that the presence of a linguistic distance between the language of schooling (Castilian) and the language of use (Valencian) in a significant group of municipalities acted as one of the determinants of the huge differences in literacy rates existing across the territory of the region of Valencia. So, in the case of Valencia is possible to identify the existence of a language mismatch effect on education outcomes. Secondly, the study also shows that this effect only appeared when the Spanish state had the capacity to enforce the accomplishment of the linguistic normative in the public schools, in parallel with the growth of its financial and administrative capacity and the ascend of an incipient democratic regime. As a corollary it can also be concluded that compulsory education in Spanish either in Valencian or in Castilian speaking municipalities caused a territorial uneven distribution of the costs of Spanish nation building.

The article is developed as follows. In section 2, the historical background is presented. First, information is provided on the process of formation of the linguistic areas in the region of Valencia. Secondly, a brief review is made of the historical evolution of the regulations that governed primary education in Spain, of the regulations that established the obligatory use of Spanish in primary classrooms and of the potential capacity of the State to impose its

---

<sup>7</sup> In the original Spanish these entities were known as *villas, lugares, aldeas, arrabales, caseríos*, etc. Constructing the data set meant digitalizing the information and homogenizing it at a level of territorial disaggregation corresponding to 524 municipalities. Today the *Comunitat Valenciana* comprises 542 municipalities, and so because of the changes that have taken place over the years, some municipalities have had to be grouped together in order to create a sample of 524. In the Appendix we give details of this municipal homogenization with reference to the main sources used: the Census of Floridablanca, 1787; the Population Censuses of Spain, 1860, 1900 and 1930; and the Nomenclator of Spain, 1887.

compliance. Section three is devoted to presenting the variables used in the analysis and to offering qualitative evidence about the existence of a cost in educational terms as a result of the mismatch between the language of instruction and the language of use in Valencian municipalities. Section four presents the strategy followed in the empirical analysis and discusses the results obtained. The paper closes with a section devoted to a summary of the main conclusions.

## **2. Historical background**

### *2.1. The creation of Spanish and Catalan linguistic areas in the Valencia region*

It is fortunate that an analysis of the determinants of literacy levels in the municipalities of Valencia in the period 1860-1930 provides us with a good opportunity to identify the existence of a mismatch effect, and for at least two reasons. The first is that, now and then, two main languages coexist in the region, these being Spanish and Catalan. According to the law of 1983 governing the use and teaching of Valenciano (as it is referred to locally), Catalan is the main language of use in two thirds of municipalities, while Spanish predominates in the other third. The situation in the mid-nineteenth century was not very different, since the territorial spread of one language or the other depended mainly on what happened at two points in history – both before the development of the liberal state – and the situation has seen no great changes up to the present day.<sup>8</sup>

Particularly, according to Casanova (2001), the establishment of one language or the other is related with two main historical events. Firstly, it can be traced back to the granting of lordships by the Crown of Aragon to the nobles who took part in the conquest of Valencia (Aragonese or Catalans) and who in exchange were given rights over the various territories and guided the

---

<sup>8</sup> Possibly the most significant change that occurred during the contemporary era was the incorporation into the province of Valencia of the municipalities located in the *comarca* or district of Utiel-Requena. This territory, which until 1851 was part of the kingdom of Castile, was Spanish-speaking. The same thing happened with the municipalities of Sax, which was transferred to the province of Alicante in 1836 from the province of Murcia (the provinces having only been created in 1833), and Villena, which that same year became part of the province of Alicante having previously belonged to Albacete. Another change occurred in Paterna, a Catalan-speaking municipality that was joined by a local council area called San Antonio de Benagéber in 1957. This small settlement was home to a group of people who came from a Spanish-speaking municipality, Benagéber, which disappeared after the construction of a reservoir (of the same name) in 1952. In 1989 San Antonio de Benagéber and Paterna separated and the new municipality, San Antonio de Benagéber, became a Spanish-speaking island surrounded by Catalan-speaking municipalities.

settling of new inhabitants. As far as the geographical distribution is concerned, Catalan lords tended to be located in territories nearest the coast, while the Aragonese were in the interior (see Figure 1).<sup>9</sup>

Besides, the map was partially redefined by the shock caused to the Valencia area by the expulsion of the Moriscos in 1609 and the subsequent repopulation process. Historiographers have analysed the origins of the settlers and recorded that most of them came from within the kingdom of Valencia, generally from neighbouring districts, although some groups have also been identified as coming from coastal towns and even from outside the region (from Mallorca, La Mancha, France and Italy).<sup>10</sup> The language that ultimately became the language of use in each territory was therefore not necessarily that spoken by the local lord, but that of the new settlers in each of these municipalities. Until 1609, for example, the linguistic frontier in the *comarcas* of the Sierra de Espadán and La Plana Baixa was marked by the frontier between the Diocese of Tortosa (Catalan) and the Diocese of Segorbe (Spanish). However, after the expulsion of the Moriscos, repopulation took the Catalan language to the municipalities of Aín, Artana, Tales, Suera, Betxí and Veo. In the Camp de Morvedre, despite the general predominance of Catalan, municipalities such as Gátova and Marines became Spanish-speaking areas after 1609 (Casanova, 2001). According to Sanchis Guarner (1973), Catalan used to be spoken in the Vega Baja del Segura due to the origins of its settlers in the fourteenth and fifteenth centuries, but from the sixteenth century it gradually became a Spanish-speaking area because of repopulation by settlers from Murcia, which increased after the expulsion of the Moriscos.

The second reason, as noted by linguists, is that in Valencia there are no transitional dialects separating Spanish-speaking and Catalan-speaking areas.<sup>11</sup> Naturally the two languages have had an influence on each other in frontier areas, basically Catalan to Spanish until the modern era,

---

<sup>9</sup> This aspect, as it is posed in section 4, needs to be taken into account in the subsequent analysis given that it could generate an element of endogeneity in the distribution of the languages over the territory, thus introducing a bias into the study of its impact on literacy levels.

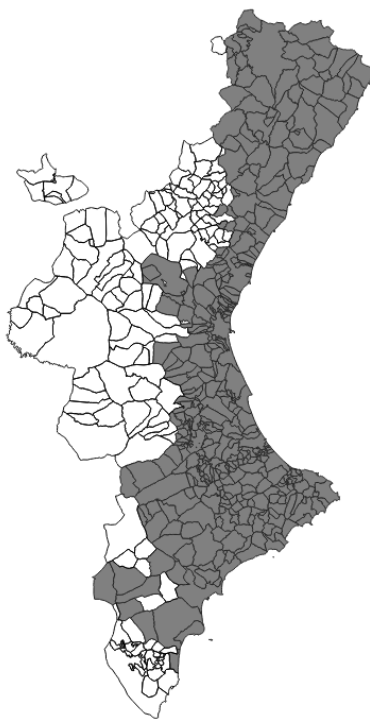
<sup>10</sup> The repopulation of the municipality of Tàrbena with settlers from Mallorca is particularly well-known, resulting in the village belonging to the Catalan linguistic domain. Even today it still retains vocabulary and certain cultural aspects typical of the Balearic Islands.

<sup>11</sup> Aragonese, a dialect spoken by some of those who settled in the kingdom of Valencia after the conquest, left a linguistic trace in some of these territories, but it gradually disappeared as a language of use. In territories that today are Catalan-speaking it was absorbed into Catalan for a number of reasons, including the demographic hegemony of the settlers from Catalonia and the prestige of a language that became the official language of the kingdom of Valencia due to its use by the political authorities in administering the municipalities and the incipient State. In what are today's Spanish-speaking areas it was absorbed in the same way by the Spanish language over the centuries following the conquest, especially during the modern era (Guinot, 1999, p.262).

and Spanish to Catalan since then. But it can safely be assumed that, apart from among certain local elites (justice, administration), the languages used by the vast majority of people in territories identified as Catalan-speaking or Spanish-speaking were Catalan and Spanish respectively.<sup>12</sup>

All in all, it can be considered that the map that draws the use of Catalan or Spanish in the region of Valencia was completely defined before the period under analysis and that the use of one or the other language was related to historical processes that occurred independently of the level of economic development of the territories occupied during the Reconquest. This element limits the relevance of the potential problem of endogeneity that the subsequent empirical analysis must face, although, as can be seen in Figure 1, it points to the existence of a relationship between linguistic dominance and geographical aspects.

Figure 1. Valencian municipalities by language



Source: see text.

---

<sup>12</sup> This means that when studying the case of Valencia, we do not need to estimate the percentage of the population that had one or the other language as their language of use, as was indeed necessary in other research aimed at analysing the impact of the linguistic mismatch in Prussia (Cinnirella and Schueler, 2016) and India (Jain, 2017).

## 2.2. *The enforcement of the compulsory use of Spanish language in primary education*

Anyway, an analysis of the existence of a linguistic effect that may have affected educational demand in the case of Catalan speaking municipalities in Valencia would make it necessary to take the preliminary step of proving that during the period analysed teaching the three Rs in schools was actually done in Spanish. From a formal perspective, the obligation for the language of schooling to be Spanish dated back almost a century, to 1769. This was done in a royal decree of Charles III as part of the centralization of the State imposed by the Bourbon dynasty after their victory in the Spanish War of Succession (1714).

It is important to point out that the Royal Decree establishing the obligatory use of Castilian as a formative language in primary schools also sanctioned two other measures aimed at building a more homogeneous society. On the one hand, in the same royal decree, the tariffs levied on judicial processes for the whole territory were unified and, on the other hand, the Castilian *real de vellón* was imposed as the unit of account for their payment. These are three measures that jointly promote monetary, fiscal and language unification, with the aim of give an impulse to the integration of the Spanish market and to increase the national mindset among citizens.

This commitment to the cultural and economic unification of the country was in the philosophy of the enlightened liberals of the eighteenth century, and was given shape and form in the Constitution of 1812, and was continued during the transition from the Ancien Regime to the nineteenth-century liberal state. With regard to education, the compulsory use of Spanish as the common language appeared in all the documents involving projects and plans put forward from that time on (the Quintana Report of 1814, the Plan Rivas of 1836, the Primary Instruction Plan of 1838). Nevertheless, it was not until 1857, a time marked by great socioeconomic change in which various countries were taking their first steps towards mass schooling and the use of education as a nation-building policy, that in Spain the Public Instruction Act (PIA), commonly known as the Moyano Law, was passed. This regulated the Spanish education system from 1857 to 1970, when it was replaced by the General Education Act (GEA).<sup>13</sup> Indeed, the Moyano Law was one of the great reforms introduced in Spain in the nineteenth century.<sup>14</sup> Education was split

---

<sup>13</sup> Although the PIA continued until 1970, successive changes were introduced during the 113 years that it remained in force. Compulsory education, for example, initially from ages 6 to 9, was extended to age 12 in 1909 and age 14 in 1964.

<sup>14</sup> This was not the only area in which the deployment of the liberal state gradually promoted administrative, territorial or economic homogenization, as well as the economic integration of Spanish regions. In a non-exhaustive manner, we could highlight the creation of a homogeneous territorial administrative structure with the creation of

into primary education and higher education (Art. 1), with a curriculum being established for both. At the same time, primary education would become “*compulsory for all Spaniards*” (Art. 7) between the ages of 6 and 9, but was free only in cases where the “parents, guardians or providers are unable to pay for it” (Art. 9).<sup>15</sup> Besides, according to Art. 2 of the PIA, primary education comprised 6 subjects: Christian doctrine and basic scripture, reading, writing, principles of Spanish grammar, principles of arithmetic, and basic knowledge of agriculture, industry and commerce, this latter subject depending on location, reaffirming the compulsory nature of the study of the first letters in Spanish.

To summarize, the regulatory framework for primary education in force in Spain in 1860 served to organize and homogenize the education system, set out its stages and establish the content of each subject. It also listed the obligations that fell to families (the schooling of their children) and municipalities (the opening of schools). However, from the point of view of funding, the reality that was characteristic of the Spanish education system under the Ancien Regime remained unchanged. The provision of educational infrastructures in 1860 was the responsibility of the local authorities and funding was still expected to come from local entities and families and, to a lesser extent, from secular or religious foundations. In these circumstances, the State's capacity to promote primary education and force the compliance with the regulations that established Spanish as the compulsory learning language was very limited.

In fact, the existence of local languages in many of the regions that made up the kingdom of Spain, their widespread use and even their survival in many documents of the time leads us to think that perhaps the vernacular languages were used for teaching in municipal schools, especially by parish priests or charitable bodies. Although how widespread this practice is impossible to quantify, it has been confirmed by various investigations into the history of education in Spain (Escolano, 1997). Even so, it seems to be a reasonable hypothesis that the construction of the liberal State and the growing interest in creating a kind of national mindset would have favoured the establishment of Spanish as the language of teaching in municipal

---

the provinces (1836) and the organization of the territory into judicial districts and municipalities, the reform of the Treasury (1845), the deployment of the postal service, the law for the organization of the telegraph network (1855), the law to promote the construction of railway infrastructures (1855) or the monetary unification around the peseta (1869). All of them, as a whole, would make up a broad program aimed at the creation of a homogeneous national conscience and the economic impulse through the integration of the Spanish market.

<sup>15</sup> The compulsory nature of education was not absolute, since pupils could ask to be excused when they were “*sufficiently provided with this type of education in their homes or in a private establishment*” (Art. 7). And to obtain free primary education, a “*certificate issued by the relevant parish priest and endorsed by the town mayor*” had to be provided (Art. 9).

schools and that the use of regional languages would have gradually diminished.<sup>16</sup> Historians of education have defended the growing use of Spanish in schools on the basis of evidence relating to teacher training. De Gabriel (1994), for example, has shown that, since the Normal Schools for teacher-training were created in 1849, their curricula only ever considered Spanish as the language in which to teach grammar or the history and geography of Spain. Another stream of the literature has tackled the problem by analysing the materials and textbooks used in primary and secondary teaching, confirming that all of them were published in Spanish.<sup>17</sup>

The compulsory use of Spanish in primary education gained momentum with the creation of the Ministry of Public Instruction and Fine Arts in 1900.<sup>18</sup> From the mid-nineteenth century and during the early twentieth century, the Spanish economy and society underwent a number of important changes. Between them, the extension of the franchise, a rapid population growth and accelerated economic structural change.<sup>19</sup> It was in this context of far-reaching socioeconomic change that the Ministry of Public Instruction and Fine Arts was created, followed two years later by the central government taking over the funding of primary education. In particular, since 1902, the state took over the salaries of public school teachers, who were then considered civil servants of the administration. Under these conditions, the construction of a centralized and publicly financed educational system provided the liberal state with the necessary legitimacy to force greater compliance with linguistic regulations. Proof of this are the words of the Minister of the Interior, Eduardo Dato, in his speech to the Senate in 1900:

---

<sup>16</sup> This process of linguistic unification would come about contemporaneously in other European states that were in the process of consolidation over the same period. Fouret and Ouzof (1980) describe the process, seemingly very similar to that of Spanish in Spain, whereby French was introduced in schools in territories that had their own languages such as Corsican and Breton. Cinnirella and Schueler (2016) describe the introduction and spread of German in Polish-speaking territories during the time of the Empire.

<sup>17</sup> It is significant that the widely-used *El instructor de la Juventud* (Young People's Instructor) was published by Esteban Paluzie in 1845 in Barcelona but was of course in Spanish. Textbooks from this publisher, such as the *Manual de Escritura y Lenguaje* (Manual of Writing and Language), were used in the Normal School of Barcelona.

<sup>18</sup> With the Finance Act of 31 March 1900 came the reorganization of the Ministry of Public Works, and after the Royal Decree of 18 April 1900 came the creation of the Ministry of Agriculture, Industry, Trade and Public Works and the Ministry of Public Instruction and Fine Arts (1900-1936). The Ministry of Agriculture, Industry, Trade and Public Works would revert to being called the Ministry of Public Works from 1905 onwards.

<sup>19</sup> General male suffrage replaced census suffrage in 1890. On the social and economic changes recorded in this period see Pérez Moreda et al. (2015) and Prados de la Escosura (2017).

*Es sensible que no todos los españoles conozcan el idioma nacional; pero es un hecho innegable respecto al cual no cabe hacer más que una constante propaganda encaminada a extender el conocimiento de la lengua oficial, evitando que en las Escuelas del Estado se enseñe otro idioma que el español, no permitiendo textos escritos en ningún dialecto, como libros de enseñanza, y adoptándose por las autoridades las medidas que se han adoptado al efecto, y que aplicará este Gobierno como las han aplicado todos los anteriores Gobiernos de España.*

With the legitimacy granted by the financing of primary education in public schools from the general State budget, the Government decreed a new regulation that imposed sanctions on teachers, now State officials, who failed to comply with the linguistic regulations. Thus, the Royal Decree of November 21, 1902, signed by King Alfonso XIII, established in its article 2:

*Los Maestros y Maestras de instrucción primaria que enseñasen á sus discípulos la doctrina cristiana u otra cualquiera materia en un idioma o dialecto que no sea la lengua castellana, serán castigados por primera vez con amonestación por parte del Inspector provincial de primera enseñanza, quien dará cuenta del hecho al Ministerio del ramo; y si reincidiesen, después de haber sufrido una amonestación, serán separados del Magisterio oficial, perdiendo cuantos derechos les reconoce la ley.*

The approval of this decree raised a wave of protests from regionalist political circles or from the Catholic Church itself, responsible for a relevant fraction of the primary schools. In the parliamentary debates opened after its promulgation, the Minister of Public Instruction, *Conde de Romanones*, vehemently exposed the reasons why he considered the pertinence of the Decree:

*Para ser maestro no se necesita ser, ni hablar catalán, sino hablar y poseer el castellano. [...] ¿Se puede consentir en buenos principios, sobre todo ahora que los maestros públicos son funcionarios del Estado, por haber pasado al Estado el paso de esta atención; se puede consentir que en las escuelas de España se enseñe un idioma o dialecto que no sea el español? Esta es la pregunta que hago yo a todos los Sres. Diputados. [...] En el seno de las familias, en el hogar doméstico pueden hablarse todos los idiomas o dialectos que se quiera; en la escuela del Estado, sostenida y vigilada por la acción del Estado, no se puede enseñar más que en idioma nacional, y como idioma nacional en España es el castellano...*



Therefore, the existing qualitative evidence points to the strengthening of state pressure to enforce the obligatory use of Spanish in primary education. Moreover, the relevance of the tightening of regulations was argued on two fundamental bases: that primary education was now financed by the State and that the State only recognized one national language, Spanish.<sup>20</sup> As a result of this process, historians of education have argued that Spanish gradually became the only language of schooling in primary education across Spain (Escolano, 2012) and that the use of regional languages practically disappeared during the first years of the twentieth century.

Based on this evidence, it seems reasonable to argue that the increasing imposition of the compulsory use of Spanish in public schools, in a context in which great advances are made in mass schooling (Beltran-Tapia et al., 2019), would have meant the growth of a real differential educational cost between Valencian children who were taught in Castilian living in municipalities belonging to the Castilian linguistic domain and those who were forced to do so in this language, even though in their territories the language of use was Catalan. In these circumstances, it could be hypothesized the potential existence of a relationship between linguistic mismatch and literacy levels in Valencian municipalities and its growing relevance once the liberal state could manage to enforce the compulsory education in Spanish in Catalan speaking territories.

### **3. Sources, data description and descriptive analysis**

To carry out the empirical analysis of this hypothesis an original dataset has been created after manipulating and merging several data sources. One of the complexities to build the dataset appears in the selection of the number of municipalities because it has been variable due to the segregation and fusion phenomena that have taken place along the modern Spanish history. Our sample considers 524 municipalities that appear in the three censuses needed to build the dataset.<sup>21</sup>

Once the number of municipalities has been established, the main variable of the analysis, the literacy rate, is obtained from the censuses of 1860, 1900, and 1930 as the total men that could read and write over the total men population. Only information about men is considered to

---

<sup>20</sup> In this same period, the liberal State limited the use of regional languages in communications with the public administration by postal mail or telegraph. Royal Decree of June 4, 1904, signed by King Alfonso XIII at the proposal of the Minister of the Interior, José Sánchez Guerra.

<sup>21</sup> Beltrán Tapia et al. (2019) provides a detailed explanation about the way in which the changes of the number of municipalities among the different censuses has been managed.

compute literacy rates because at that time there was significant literacy gender gap, phenomenon that could complicate the interpretation of our results.

Table 1 reports basic descriptive statistics of men literacy rate by municipality for the three censuses. The average of male literacy rate increased from 14.3 percent in 1860, to 26.4 in 1900, to 53.6 percent in 1930. Indeed, one of the most striking characteristics of Spanish literacy is that it varied greatly across the territory as a whole. This is true not only when comparing the regions that made up the kingdom of Spain (Núñez, 1992), but also when comparing municipalities within the same region or province (Beltrán Tapia et al., 2019). A good illustration of this variability across municipalities is the region of Valencia, as shown in Table 1 and Figure 2.

The other key variable in this analysis is the one that classifies municipalities according to their linguistic domain, namely between Catalan or Spanish. Following the *Ley de uso y enseñanza del valenciano* published by the Generalitat Valenciana in 1983 (and assuming that this situation is not far from the one at the end of the Old Regime) 381 municipalities are classified as Catalan linguistic domain and as Spanish linguistic domain (Figure 1).

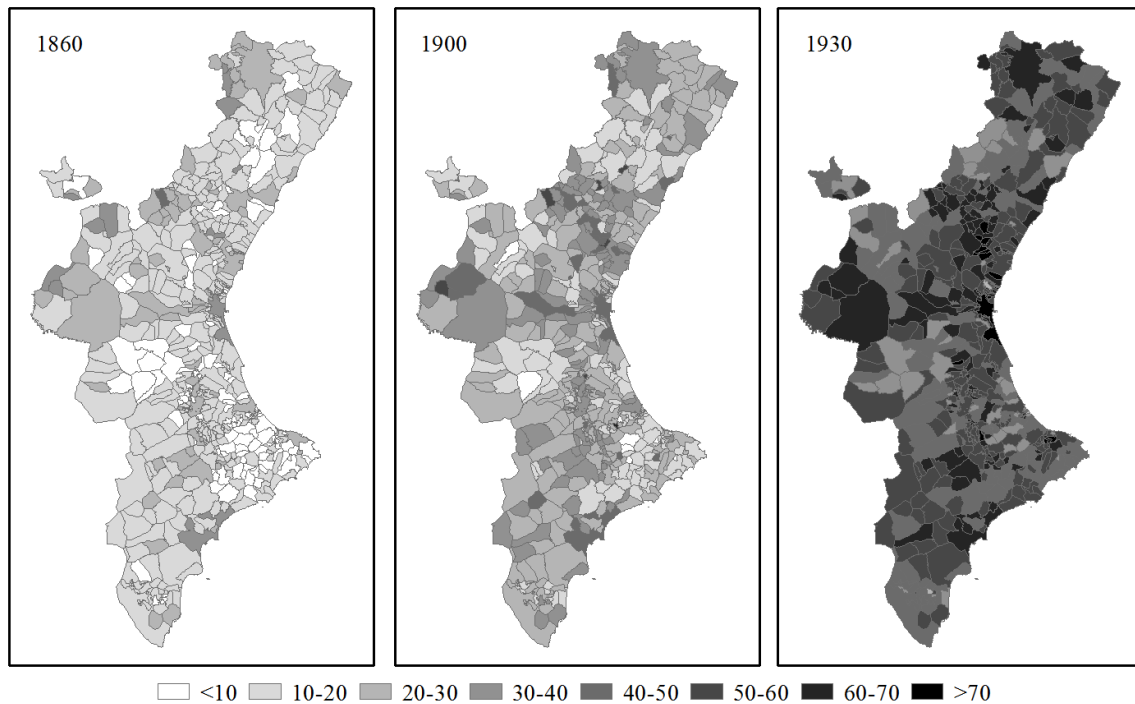
Table 1: Main descriptive statistics of male literacy rate (percentages)

| Sample             | N   | mean | Sd   | min  | p25  | p50  | p75  | Max  |
|--------------------|-----|------|------|------|------|------|------|------|
| Census 1860        |     |      |      |      |      |      |      |      |
| All municipalities | 524 | 14.3 | 6.5  | 0.9  | 10.1 | 13.5 | 17.7 | 44.0 |
| Catalan-speaking   | 381 | 13.4 | 6.1  | 0.9  | 9.4  | 12.9 | 16.5 | 39.7 |
| Spanish-speaking   | 143 | 16.6 | 7.2  | 3.5  | 11.9 | 15.2 | 20.6 | 44.0 |
| Census of 1900     |     |      |      |      |      |      |      |      |
| All municipalities | 524 | 26.4 | 9.2  | 7.4  | 19.8 | 25.5 | 32.2 | 68.9 |
| Catalan-speaking   | 381 | 26.0 | 8.9  | 7.4  | 19.8 | 25.4 | 31.6 | 68.9 |
| Spanish -speaking  | 143 | 27.6 | 9.7  | 9.7  | 20.6 | 26.4 | 34.1 | 55.7 |
| Census of 1930     |     |      |      |      |      |      |      |      |
| All municipalities | 524 | 53.6 | 9.6  | 25.2 | 47.6 | 54.2 | 60.2 | 82.5 |
| Catalan-speaking   | 381 | 54.0 | 9.4  | 25.2 | 48.1 | 54.6 | 60.2 | 82.5 |
| Spanish -speaking  | 143 | 52.4 | 10.3 | 29.3 | 45.0 | 52.1 | 60.8 | 77.1 |

Notes: The percentage of male literacy in any municipalities is computed as the percentage of men that know read and write respect to the total male population of each municipality. This mean is an unweighted average.

Sources: Censuses of 1860, 1900, and 1930, *Ley de uso y enseñanza del valenciano* (Generalitat Valenciana, 1983), and own elaboration.

Figure 2.- Male literacy rates in 1860, 1900 and 1930 by municipality (%)



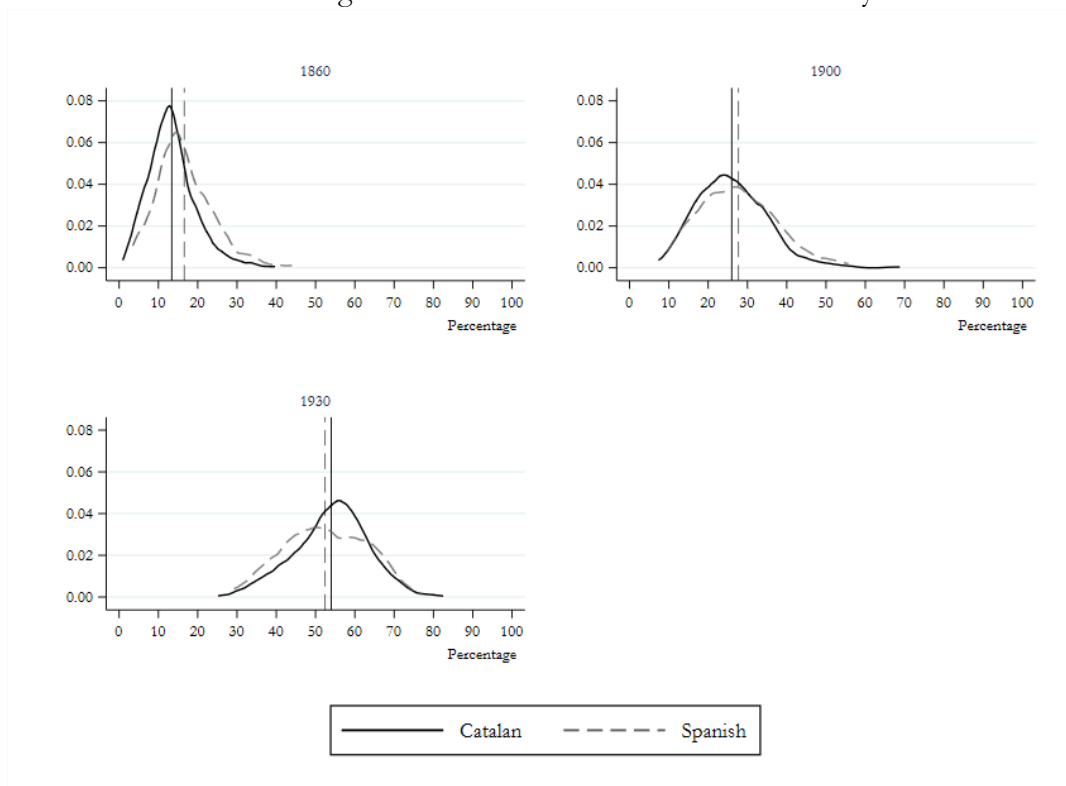
Source: own elaboration based on the 1860 Census of Population.

As shown in Table 1, when we discriminate municipalities by language a new point of interest emerges. In 1860 the average literacy rate in Catalan municipalities was lower than in Spanish municipalities (13.4 versus 16.6). This difference decreased in 1900, where the difference between Catalan and Spanish municipalities was only 1.6 percentage points (26 versus 27.6). Nevertheless, this scenario changed in 1930, where the Catalan municipalities had, on average, a higher male literacy rate (54 versus 52.4).

The basic descriptive statistics for male literacy are complemented with the kernel densities (Figure 3). There are some interesting issues. On the one hand, both distributions move to the right and, on the other hand, distributions were left-skewed in 1860 and 1900, but not in 1930. These two facts indicate that there was a generalized increase in male literacy rates of the municipalities of the former Kingdom of Valencia along the period considered. Concerning the effect of the language, the distribution of Spanish-speaking municipalities was located to the right of Catalan-speaking municipalities in 1860, indicating that the latter had lower literacy level. The situation is not so clear in 1900, where the right tail of the Catalan-speaking municipalities is longer than the one for the Spanish-speaking municipalities; namely, municipalities with highest male literacy rates were Catalan-speaking. It seems that the advantage of the Spanish-speaking

municipalities decreases still more in 1930, where a considerable part of the distribution of Spanish-speaking is located to the left.

Figure 3: Kernel distribution for male literacy



Sources: Censuses of 1860, 1900, and 1930 and own elaboration.

In any case, in order to carry out a precise empirical analysis of the existence of a relationship between language mismatch and literacy rates by municipality, it also needs to include additional elements that generate variability in the costs and benefits of education – and thus in the literacy rates recorded –. In this respect, in the analysis that follows, it has been considered the territorial variability to the institutional characteristics of Valencian municipalities during the Ancien Regime, obtained from the census of 1787, i.e. the Census of Floridablanca. This source provides information on the size, structure and occupation of the population by settlement, and these settlements are classified by category (city, town, village, hamlet...) and jurisdiction. The census classifies the type of jurisdiction of local entities as *royal*, *ecclesiastical lordship*, *secular lordship* and *military orders*. Using the information on the kingdom of Valencia, we have been able to identify the jurisdictional regime of all 524 municipalities. With this information we create the dummy variable *lordship*, which is equal to one when the municipality is classified as lordship and equal to zero when classified as royal. Besides, we have also identified each municipality as being

inhabited by Moriscos until they were expelled in 1609. With this information we have created the dummy variable *morisco*, which is equal to one when the municipality was inhabited by Moorish until 1609.<sup>22</sup>

We have also taken into account the total population as well as an approximation of the settlement pattern in each municipality. This means we can analyse whether the inhabitants were concentrated in the core population entity (towns, villages) or scattered across minor entities (hamlets, homesteads, mills, rural dwellings and isolated buildings). To do this we have created a variable that measures the percentage of population that lived in the main population centre. In this case the information comes from the *Nomenclator* of Spain for 1887. Finally, since geographical characteristics could also affect literacy levels, the analysis includes controls for both first-nature (temperature, rainfall, altitude and ruggedness) and second-nature geography (distance to roads).<sup>23</sup> Finally, the economic structure of the municipalities could have an effect on skilled work demand, so the share of manufacturers and artisans on total employment in 1787 has also been computed.

Table 2. Main descriptive statistics of the control variables

| <b>Variable</b>                     | <b>N</b> | <b>Mean</b> | <b>sd</b> | <b>Min</b> | <b>p25</b> | <b>p50</b> | <b>p75</b> | <b>Max</b> |
|-------------------------------------|----------|-------------|-----------|------------|------------|------------|------------|------------|
| Population 1860                     | 524      | 2,434       | 6,795     | 90         | 709        | 1,229      | 2,386      | 140,61     |
| % population in the core entity     | 524      | 86.03       | 18.14     | 4.64       | 80.68      | 94.21      | 98.15      | 100.00     |
| Lordship                            | 431      | 13.5        | 6.0       | 0.9        | 9.5        | 13.1       | 16.6       | 44.0       |
| Royal                               | 93       | 17.8        | 7.8       | 5.1        | 12.0       | 17.2       | 21.5       | 39.7       |
| Morisco                             | 216      | 11.9        | 5.4       | 0.9        | 7.8        | 11.6       | 15.2       | 32.0       |
| Non-Morisco                         | 308      | 16.0        | 6.7       | 3.4        | 11.4       | 14.6       | 19.9       | 44.0       |
| Distance to main road (km)          | 524      | 17.8        | 15.8      | 0.06       | 4.5        | 13.6       | 27.7       | 69.1       |
| Temperature (degrees)               | 524      | 15.7        | 2         | 9.7        | 14.2       | 16.2       | 17.3       | 18.3       |
| Rainfall (mm)                       | 524      | 470.2       | 62.9      | 280.9      | 439.9      | 464.4      | 507.7      | 627.3      |
| Altitude (m)                        | 524      | 372.5       | 316.4     | 1.2        | 90         | 298        | 624.1      | 1330.6     |
| Ruggedness                          | 524      | 91.7        | 62.5      | 1.5        | 34.8       | 90.6       | 136.6      | 289.8      |
| Share of manufacturers and artisans | 474      | 8.5         | 9.4       | 0          | 2.5        | 6.4        | 10.5       | 66.3       |

Sources: Own elaboration based on the 1787 Census of Population; the 1860 Census of Population; Lapeyre (1959, 1986), *Nomenclator de España*, 1887 and GIS.

<sup>22</sup> A specific analysis of the effects of differences in institutional setting on literacy levels across Valencian municipalities in Beltran-Tapia et al. (2020).

<sup>23</sup> We have also constructed additional variables representative of the second-nature geographic characteristics of the municipalities (distance to the coast, distance to the capital city). Since taking these alternative measures into account does not alter the main results of the empirical exercise, they are not described in the text.

#### 4. Empirical strategy and results

The descriptive analysis confirms the existence of a language effect but, a priori, it seems in favour of the Catalan-speaking municipalities, at least from 1900 on. Nevertheless, other things beyond the language mismatch could be explaining this preliminary result. Then, we propose an econometric analysis that allows us to control for the additional elements that affect literacy rate. The econometric specification for a specific year  $t$  is the following:

$$male\_literacy_{it} = \alpha_0 + \alpha_1 Catalan_i + \beta X_{it} + \delta Z_i + \lambda_j + u_{it} \quad (1)$$

where the subscript  $i=1, \dots, 524$  refers to municipalities,  $j=1, \dots, 17$  refers to districts, and  $t=1860, 1900, 1930$  refers to the year of the census. The dependent or endogenous variable, the male literacy rate, is explained by the language of the municipality ( $Catalan_i$ ), a group of control variables—both time-variant ( $X_{it}$ ) and time-invariant ( $Z_i$ ) controls variables—, districts fixed effects ( $\lambda_j$ ), and an error term ( $u_{it}$ ).

Our main objective is to analyse the impact generated by the existence of a mismatch between the formative language and the language used by the population; namely, we want to know the male literacy rate of a specific Catalan-speaking municipality if in that municipality the official language would have been Spanish. This hypothetical situation, known as counterfactual, is unknown. In any case, if the distribution of the language were exogenous, the estimated coefficient  $\alpha_1$  would inform about the language effect: A negative (positive) coefficient means that Catalan-speaking municipalities has, on average, lower (higher) male literacy rates than Spanish-speaking municipalities.

Nevertheless, as it has been shown in Section 2.1 and Figure 1, the distribution of language was not random since Catalan-speaking municipalities are located in territories nearest to the coast. In this regard it could be argued that municipalities closer the coast, given their characteristics, enabled greater specialization in commercial activities. Then, it could be thought that the Catalan-speaking territories were also characterized by the greater specialization in activities that made more intensive use of human capital, an aspect that would encourage greater educational investment in them. Or on the contrary, that they specialized in agricultural production sectors that were more intensive in unskilled work, which means that families would have to deal with a greater opportunity cost when it came to sending their children to school. Whatever the direction of this effect, it is possible for those Catalan-speaking municipalities to obtain education results significantly different to Spanish-speaking municipalities, without this having to be due to a mismatch between the language of use and the language of schooling but instead

due to other elements which are related to the prevalence of one language or the other and have a direct effect on education results.

Therefore, as the official language of a municipality was not a random phenomenon, using the ordinary least square estimation method will produce a biased estimated coefficient  $\alpha_1$ . In order to overcome this issue, an alternative estimation technique is proposed, the propensity score matching. This technique, very common to assess the impact of specific program or public policies, consists of comparing the literacy rate of a treated unit (in this case a Catalan-speaking municipality) with a very similar one untreated unit (in this case, a Spanish-speaking municipality). Averaging all these comparisons provides the average impact, known as the average treated effect of the treated (ATET).

The technique consists of several steps. In the first step it is computed the propensity score or the probability of being treated, namely, to be a Catalan-speaking municipality. The variables choice this propensity score is one of the more controversial aspects of this technique because the inability to capture well the characteristics that determine whether a unit is treated generates biased estimators (Heinrich et al, 2010). We assume that all the characteristics in which the treated and control groups differ are observed, which is known as conditional independence assumption or unconfoundedness assumption (see the Appendix for a brief explanation). As this assumption cannot be tested, the validity of the model relies on the economic theory and previous empirical findings (Cappelli and Vasta, 2020). In our case the historical evidence shows that the official language of a specific municipality was related to the geographical characteristics of that municipality. Then, we use the geographical variables (both first and second nature)<sup>24</sup> in order to compute the probability of being a Catalan-speaking municipality (Table A1 in the Appendix provides the estimated results). The other assumption that cannot be violated is the common support or overlap condition, which implies that there must be enough municipalities (Catalan or Spanish speaking) with similar propensity scores. The distribution of propensity scores confirms that there is wide common support (see Figure A1 in the Appendix), which ensures that each treated municipality has at least one untreated municipality that behaves as counterfactual.

Finally, we match the similar units in the treatment and control group using a matching technique. Firstly, we rely on the Kernel matching, a non-parametric matching estimator that, in

---

<sup>24</sup> The geographical variables are the following: temperature, rainfall, altitude, ruggedness, distance to the coast, distance to the capital city, and distance to main roads.

order to construct the counterfactual outcome, uses weighted average of all individuals in the control group. As it is used all the available information, the estimator has a low variance, although the bias could be high. In order to compute the ATET we take into account, apart from the geographical variables, the other control variables presented in section 3 and considered in equation 1.<sup>25</sup>

Table 3 presents the magnitude and significance of the average treatment effect of the treated for three different years (1860, 1900, and 1930). The results show that, in terms of male literacy rates, the language did not have any significant effect neither 1860 nor 1900. In 1930, nevertheless, the language mismatch was statistically significant at 10 percent. If Catalan-speaking municipalities were Spanish-speaking municipalities, the male literacy rate would have been 5.9 percentages points higher.

Table 3. Average treatment effect of the treated  
 Matching method: Kernel  
 Number of treated units: 118, Number of untreated units: 112

| Year | Difference (ATET) | Std. Error (Bootstrap) | t-stat        |
|------|-------------------|------------------------|---------------|
| 1860 | -0.776            | 3.109                  | -0.249        |
| 1900 | -3.638            | 3.196                  | -1.139        |
| 1930 | -5.859            | 3.862                  | <u>-1.517</u> |

Notes: Common support: [0.00804, 0.9787].

As far as there is not a best matching procedure (it depends on the data structure), we have also made use of two additional matching techniques (nearest neighbour and stratification) as robustness checks.<sup>26</sup> On the one hand, the nearest neighbour compares each treated municipality with the municipality in the control group that has the more similar propensity score. As not many observations are used (not all the units in the control groups are used), this method computes estimators with low bias, but high variance. On the other hand, the stratification matching consists of dividing the common support into a set of intervals or groups. Then, the program compares treated and control units within each group and compute an average impact. The optimal number of intervals depend on the data, and it is very important to test that, within

<sup>25</sup> The (non-geographical) control variables are: population, settlement pattern, lordship, morisco, the percentage of manufacturers and artisans, latitude and longitude, and districts fixed effects.

<sup>26</sup> See Caliendo and Kopeining (2005) for a description of the matching techniques.



each group, the mean propensity score is not different for treated and controls units. In our case, the program has generated eight different groups.<sup>27</sup>

Tables 4 shows the results of these two techniques, being the panel a for the neighbour matching and panel b for the stratification matching. The results are consistent with the previous one: the language mismatch was not relevant neither 1860 nor 1900 but was statistically significant (at 5 percent) is 1930.

Table 4: Average treatment effect of the treated  
 Panel a: Matching method: Neighbour  
 Number of treated units: 118, Number of untreated units: 30

| Year | Difference (ATET) | Std. Error (Bootstrap) | t-stat        |
|------|-------------------|------------------------|---------------|
| 1860 | 1.260             | 3.243                  | 0.389         |
| 1900 | -4.323            | 3.826                  | -1.130        |
| 1930 | -7.599            | 4.584                  | <u>-1.658</u> |

Notes: Common support: [0.00804, 0.9787].

Panel b: Matching method: Stratification  
 Number of treated units: 118, Number of untreated units: 112

| Year | Difference (ATET) | Std. Error (Bootstrap) | t-stat        |
|------|-------------------|------------------------|---------------|
| 1860 | 1.452             | 1.649                  | 0.881         |
| 1900 | -3.806            | 2.975                  | -1.279        |
| 1930 | -7.615            | 4.179                  | <u>-1.822</u> |

Notes: Common support: [0.00804, 0.9787]. Number of groups: 8

Apart from the matching estimator, there are other estimators to compute treatment effects. To complete the robustness analysis, we present the results of the inverse-probability-weighted regression adjustment (IPWRA) estimator. This estimator uses two models, one to predict the treatment status and other to predict outcomes, and one of its advantages is that only one of the models must be correctly specified in order to get a consistent estimator (see the Appendix for a brief description of this estimator). As before we estimate the treatment model respect to the geographical variables (both first- and second- nature) and the outcome model that includes all the control variables. Table 5 shows the results, which confirm the existence of a language mismatch in 1930. Nevertheless, unlike the previous results, with this procedure this mismatch was already statistically significant in 1900.

---

<sup>27</sup> We use the Stata command “pscore” (StataCorp, 2021b).

Table 5: Average treatment effect of the treated  
 Estimator: IPWRA  
 Outcome model: linear  
 Treatment model: logit  
 Number of observations: 140

| Year | ATET    | Robust SE | Z            |
|------|---------|-----------|--------------|
| 1860 | 1.1635  | 2.4425    | 0.48         |
| 1900 | -6.8871 | 3.2450    | <u>-2.12</u> |
| 1930 | -9.9130 | 4.0137    | <u>-2.47</u> |

The results presented in Tables 3-5 support the existence of a negative effect on the percentage of male literacy in the municipalities treated, which is noted in the case of the use of the IPWRA estimator in 1900 and which is robustly observed in the time cut-off corresponding to 1930. In this case, the effect is robust regardless of the estimation method used and ranges from 5.9% when the Kernel method is used to determine the matched observations to 9.9% in the case of an IPWRA estimation of the treatment effect. That is to say, it could be concluded that the Valencian municipalities that form part of the Catalan linguistic domain reached a significantly lower level of male literacy in 1930 than the municipalities that can be considered to have similar attributes but which belonged to the Castilian linguistic domain, thus proving the presence of a language mismatch effect.<sup>28</sup>

## 5. Concluding remarks

Using the case of Valencia, in this article we have analysed the impact caused by the introduction of an education system that sometimes generated a linguistic distance between the language of schooling and the language of use in the literacy levels of the municipalities. In this respect, we find that there is a negative effect between the presence of a linguistic distance between the language of use and the language of schooling and the population's level of education. These results endorse those found by Cinnirella and Schueler (2016) and Jain (2017) regarding the importance of the language mismatch effect on results in education.

Nevertheless, we have shown that, in the case of the Region of Valencia, this effect only appears at the end of the analysed period. In this sense, if we consider that the regulations imposing Spanish as the language of instruction in primary schools originated in the Ancien Régime, the absence of effects on educational levels would be an indication of the lack of effectiveness of

---

<sup>28</sup> Jain (2017) provides values for the mismatch effect in Indian districts that would range between 18 and 22%.

the regulations. Only in the period that begins with the entry into the twentieth century is it possible to identify the expected effects of the implementation of a nation-building policy such as the one described.

This temporal evolution could be related to different aspects that deserve further consideration. The first is that the effectiveness of this type of policy is related to the capacity of the state to force its compliance. Throughout the text it has been shown that in the Spanish case this registered a key change around 1900, when the state assumed the financing of public primary schools, thus legitimizing the forced compliance with the regulations in spite of the opposition from social segments such as the Church or some local authorities. The results obtained in this research would support the hypothesis according to which state capacity acted as a contingent on the effective implementation of nation-building policies. In the Spanish case, this capacity was very limited until the first third of the twentieth century, endorsing a gradual and slow vision of the deployment of the liberal state throughout the nineteenth century and the first third of the twentieth century.

Second, considering that in the Spanish case universal male suffrage was introduced in 1890, the identification of this effect in 1930 could be linked to the growing interest of the state in achieving linguistic unification. In this sense, Alesina and Reich (2013) and Alesina, Giuliano and Reich (2019) put forward a theoretical model from which it is derived that the advance of democracy would drive the use of homogenization policies in the face of the growing risk of social unrest. The results obtained in this paper provide partial evidence in favor of the fulfillment of this hypothesis in the context of the incipient advance of democracy in Spain at the turn of the century.

Linking both aspects, thirdly, this result would be in line with the thesis sustained in Alesina, Giuliano and Reich (2019), which would point to the advance of democracy as a relevant element for the understanding of the growth of state capacity, necessary for the effective implementation of public policies of national homogenization.

Finally, the evidence presented contributes to the knowledge of the determinants of the unequal progress of mass literacy in the Spanish territory. The confluence in the Spanish territory of different languages and the compulsory use of Spanish as the only language of instruction would have meant a differential educational cost for those citizens and territories that did not have Spanish as their mother tongue. In this sense, the consideration of the unequal distribution of

the cost associated with the implementation of this nation-building policy can contribute to a better understanding of the origins of the economic and territorial inequality that characterizes Spanish society.

### **Acknowledgements**

The authors would like to thank the Bank of Spain for its financial support through the “*Programa de Ayudas a la Investigación en Macroeconomía, Economía Monetaria, Financiera y Bancaria, e Historia Económica*”. Financial support from the Spanish Ministry of Science and Innovation, Projects ECO2015-65049-C12-1-P, ECO2017-90848-REDT, HAR2015-64076-P, PGC2018-095821-B-I00, PGC2018-095529-B-I00 (MCIU/AEI/FEDER, UE), the Generalitat Valenciana (AICO/2018/1930 and PROMETEO/2020/083), FBBVA (*Ayudas a equipos de investigación en Humanidades Digitales*) and the Norwegian Research Council, Project 249723, is also gratefully acknowledged.

## References

- Alesina, A. and La Ferrara, E. 2005. "Ethnic diversity and economic performance", *Journal of Economic Literature*, 43(3), 762-800.
- Alesina, A. and Reich, B. 2013. Nation building *NBER Working Paper*, 18839, Cambridge, MA.
- Alesina, A., Baqir, R. and Easterly, W. 1999. "Public goods and ethnic divisions", *Quarterly Journal of Economics*, 114, 1243-1284.
- Alesina, A., Giuliano, P. and Reich, B. 2019. "Nation-building and education", *NBER Working Paper*, 18839, Cambridge, MA.
- Alesina, A., Reich, B. and Riboni, A. 2020. "Nation-building, nationalism, and wars", *Journal of Economic Growth*, 25, 381-430.
- Bandiera, O., Mohnen, M., Rasul, I. and Viarengo, M. 2018. "Nation-building through compulsory schooling during the age of mass migration", *Economic Journal*, 129(617), 62-109.
- Beltrán Tapia, F.J, Díez-Minguela, A., Martínez-Galarraga, J. and Tirado-Fabregat, D.A. 2019. *Capital humano y desigualdad territorial. El proceso de alfabetización en los municipios españoles desde la Ley Moyano hasta la Guerra Civil* (Banco de España: Estudios de Historia Económica), 74.
- Beltrán Tapia, F.J, Díez-Minguela, A., Gómez-Tello, A., Martínez-Galarraga, J. and Tirado-Fabregat, D.A. 2020. "Lordships, state capacity and beyond: literacy rates in mid-nineteenth-century Valencia", *EHES Working Papers*, 196, September.
- Bleakley, H. and Chin, A. 2004. "Language skills and earnings: evidence from childhood emigrants", *Review of Economics and Statistics*, 86(2), 267-298.
- Caliendo, M. and Kopeining, S. 2008. "Some practical guidance for the implementation of Propensity Score Matching", *Journal of Economic Surveys*, 22(1), 31-72.
- Cappelli, G. and Vasta M. 2020. "Can school centralization foster human capital accumulation? A quasi-experiment from early twentieth-century Italy", *Economic History Review*, 73(1), 159-184.
- Casanova, E. 2001. "La frontera lingüística castellano-catalana en el País Valenciano", *Revista de Filología Románica*, 18, 213-260.
- Cinnirella, F. and Schueler, M. 2018. "Nation building. The role of central spending in education", *Explorations in Economic History*, 67, 18-39.

- Cinnirella, F. and Schueler, M. 2016. “The cost of decentralization: linguistic polarization and the provision of education”, *CESifo Working Paper*, 5894.
- Clots-Figueras, I. and Masella, P. 2013. “Education, language and identity”, *Economic Journal*, 123, 332-357.
- Cvrcek, T. and Zajicek, M. 2013. “School, what is it good for? Useful human capital and the history of public education in Central Europe”, *NBER Working Paper*, 19690.
- De Gabriel, N. 1994. “La formación del magisterio”, in J.L. Guereña, A. Tiana and J. Ruiz (coords.), *Historia de la educación en la España contemporánea: diez años de investigación* (Madrid: Centro de Investigación y Documentación Educativa), 215-266
- Escolano A. 1997. *Historia ilustrada del libro escolar en España: del Antiguo Régimen a la Segunda República* (Madrid: Fundación Germán Sánchez Ruipérez).
- Escolano A. 2012. *Historia de la educación (Edad Contemporánea)* (Madrid: Universidad Nacional de Educación a Distancia).
- Fouret, F. and Ouzof, J. 1980. *Lire et écrire: l’alphabétisation des Français de Calvin à Jules Ferry* (Paris: Editions de Minuit).
- Guinot, E. 1999. *Els fundadors del Regne de València, vol. I* (València: Tres i Quatre).
- Heinrich, C., Maffioli, A. and Vázquez, G. 2010. *A primer applying propensity-score matching*, Inter-American Development Bank IDB-TN-161.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. and Jarvis, A., 2005. “Very high resolution interpolated climate surfaces for global land areas”, *International Journal of Climatology*, 25(15), 1965-1978.
- Instituto Nacional de Estadística. 1991. *Censo de 1787 “Floridablanca”, Tomo VI* (Madrid: Instituto Nacional de Estadística).
- Jain, T. 2017. “Common tongue: the impact of language on educational outcomes”, *Journal of Economic History*, 77, 473-509.
- Kersting, F. and Wolf, N. 2019. “On the origins of national identity”, *Rationality and Competition, Discussion paper 217*, Munich, Germany.
- Lapeyre, H. 1959. *Géographie de l’Espagne morisque* (París: SEVPEN).
- Lapeyre, H. 1986. *Geografía de la España morisca* (Valencia: Ediciones Alfonso el Magnánimo).
- Livi Bacci, M. 1987. “Il Censimento di Floridablanca nel contesto dei censimenti europei”, *Genus*, 43, 137-151.

- Miguel, E. 2004. "Tribe or nation? Nation building and public goods in Kenya versus Tanzania", *World Politics*, 56, 327-362.
- Miguel, E. and Gugerty, M.K. 2005. "Ethnic diversity, social sanctions and public goods in Kenya", *Journal of Public Economics*, 89(11-12), 2325-2368.
- Núñez, C.E. 1992. *La fuente de la riqueza. Educación y desarrollo económico en la España contemporánea* (Madrid: Alianza).
- Prados de la Escosura, L. 2017. *Spanish economic growth, 1850-2015* (Cham: Palgrave- Macmillan).
- Sanchis Guarner, M. 1973. "La frontera lingüística entre Alicante y Murcia", *Cuadernos de Geografía*, 13, 15-29.
- StataCorp 2021a. Stata treatment-effects reference manual: Potential outcomes/counterfactual outcomes. Release 17. College Station, TX: Stata Press.
- StataCorp 2021b. Stata: Release 17. Statistical Software. College Station, TX: StataCorp LLC.
- Swee, E. L. 2005. "Together or separate? Post-conflict partition, ethnic homogenization, and the provision of public schooling", *Journal of Public Economics*, 64, 415-424.
- Williams, E. and Cooke, J. 2002. "Pathways and labyrinths: language and education in development", *TESOL Quarterly*, 36(3), 297-322.

## Appendix A

### Conditional Independence Assumption or Unconfoundedness assumption

This appendix provides a brief explanation about the importance of the conditional independence assumption or unconfoundedness assumption. For further details, see Caliendo and Kopeining (2008) and Heinrich et al (2010).

In a random assignment, the treatment status ( $T$ ) is uncorrelated with any other variable (both observable and unobservable) and, as result, the potential outcomes will be statistically independent of the treatment status. In mathematic terms:  $(Y_1, Y_0) \perp T$ , where  $Y_1$  is the potential outcome of a treated unit and  $Y_0$  is the potential outcome of an untreated unit.

Without random assignment, treatment may be correlated with factors influencing the potential outcomes. Then, treated and untreated units not only differ in the treatment status but also in other characteristics. If the researchers are not able to control for the characteristics that determine the selection process, the impact of the program will not be correctly estimated.

We assume that the characteristics on which the treated and untreated units differ are observable. Given the propensity score (that depends on these observable variables), the potential outcomes are independent of the treatment status, namely the treatment status is as good as random:  $(Y_1, Y_0) \perp T | p(X)$ .

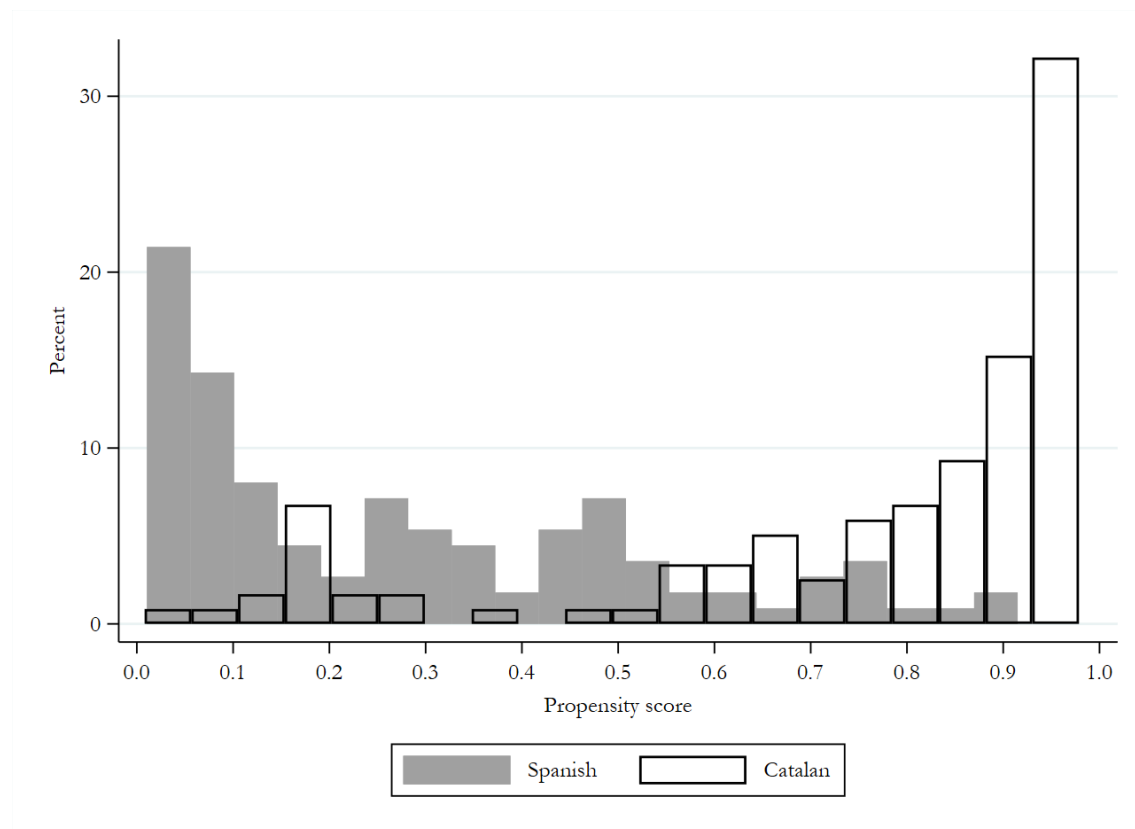
This assumption allows us to correctly identify the impact of the program because it ensures that the differences between the treated and untreated units are taken into account diminishing, in this way, the selection bias. Then, the untreated observations can be used as counterfactual for the treated observations.



Table A1: Core probit model for estimating propensity scores, Catalan

|                                   | Coeff.                  | Maginal effects        |
|-----------------------------------|-------------------------|------------------------|
| Temperature (degrees)             | 1.9352***<br>(0.3345)   | 0.1962***<br>(0.0290)  |
| Rainfall (mm)                     | 0.0421***<br>(0.0043)   | 0.0043***<br>(0.0002)  |
| Altitude (m)                      | 0.0054***<br>(0.0019)   | 0.0005***<br>(0.0002)  |
| Ruggedness (u)                    | -0.0047*<br>(0.0026)    | -0.0005*<br>(0.0003)   |
| ln(Distance to the coast) (km)    | -0.9634***<br>(0.2900)  | -0.0977***<br>(0.0282) |
| ln(Distance to capital city) (km) | 0.2637<br>(0.3148)      | 0.0267**<br>(0.0319)   |
| ln(distance to main road) (km)    | 0.1766*<br>(0.0925)     | 0.0179*<br>(0.0093)    |
| Intercept                         | -48.0418***<br>(7.1136) | 0.1962***<br>(0.0290)  |
| Observations                      |                         | 524                    |

Figure A1: Visual analysis for common support for estimated propensity scores



Notes: Common support: [0.00804, 0.9787].

### **Inverse-probability-weighted regression adjustment (IPWRA) estimator**

This Appendix provides a brief explanation about the IPWRA estimator. For further details, see StataCorp (2021a).

The IPWRA combines elements of the inverse-probability weighted (IPW) estimator and the regressions adjustment (RA) estimator. The advantage of the IPWRA estimator is that it is more robust to misspecification since only one of the models must be correctly specified in order to obtain a consistent estimator.

One of the most difficult challenges to analyze the impact of treatment is to construct a credible counterfactual, above all when the treatment status is not random and depends on some characteristics that also affect the outcome.

The IPW estimator consider these counterfactuals as missing values and weights the observed units by the inverse of the probability of being in a specific group (treatment or control). The objective is to correct the estimates of the treated and untreated sample means for the missing data (the counterfactuals). For instance, in order to compute the male literacy rate of Catalan-municipalities we had to apply more weight to the Catalan-municipalities further to the coast. On the other hand, the aim of the regression adjustment is to predict potential outcomes after considering the effect of relevant covariates.

The IPWRA estimators uses a three-step approach (although the Stata output only present the final step) to estimate treatment effects:

- 1) They estimate the treatment model in order to compute the inverse probability weights.
- 2) For each treatment level, they estimate the weighted regression models of the outcome and, for each subject, obtain the treatment-specific predicted outcomes.
- 3) They compute the means of the treatment-specific predicted outcomes. The differences of these averages provide the estimates of ATEs (considering both treated and untreated subjects) and ATETs (considering treated subject only).

## Appendix B

### The sample

Our sample is based on the number of municipalities located in the Valencia region today. According to the latest count, there were a total of 542 in the Comunitat Valenciana. However, there have been various changes and alterations, so we have had to make some adjustments in order to have consistent municipalities over time.

### Endogenous variable: literacy (homogenization of municipalities in Population Census from 1860 to 2011)

The Population Census of 1860 included a total of 570 municipalities in the Valencia region.<sup>29</sup> This means that we have to convert this historical figure into the current 542 municipalities.<sup>30</sup> To homogenize the data we assigned a current INE code to all the municipalities that appear in the census. First of all, following Goerlich et al. (2006: *Apéndice 2. Alteraciones de los municipios entre los censos de 1900 y 2001*) and using information from the document entitled “*Variaciones de los municipios de España desde 1842*” (Ministerio de Administraciones Públicas, 2008), the 570 municipalities of 1860 were converted into today’s 542. These sources enabled us to assign a current INE code to the vast majority of the historical municipalities in the population census, following their historical pathways. However, in some cases complications arose. Most changes in municipalities between these dates are the result of mergers or separations, although the causes vary: towns absorbing neighbouring towns, towns that were merged together, towns that split from each other, etc.<sup>31</sup> In 1860, for example, today’s municipality of l’Eliana was part of La Pobla de Vallbona. Hence no information on literacy rates for l’Eliana is available for 1860 (in fact the municipality of l’Eliana separated from la Pobla de Vallbona in 1955). To address this problem we created a «pseudo-municipality» known as Pobla de Vallbona - l’Eliana. Thus the municipalities included within a pseudo-municipality form a new single entity and have the same literacy rate, or to be more precise, a joint literacy rate. We worked in a similar way with the other cases. Overall, due to border changes, there is no information on literacy in the population

---

<sup>29</sup> These 570 municipalities were distributed among the three provinces that make up the region as follows: 142 (Alicante), 144 (Castellón) and 284 (Valencia). They were further divided into 46 *partidos judiciales* or judicial districts: 14 (Alicante), 10 (Castellón) and 22 (Valencia). Since there were 4 districts in the city of Valencia, 43 cities and towns were district capitals.

<sup>30</sup> Benicull de Xúquer became the 542nd municipality in the early 2000s, when it split from Polinyà de Xúquer.

<sup>31</sup> There were also numerous changes in the names of the municipalities, with many of them applying to express the name of the municipality in the language of the region.

census of 1860 for 18 out of the 542 municipalities. To solve this problem, 17 artificial pseudo-municipalities (covering 35 municipalities) were created (see Table A2). If the current total number of municipalities in the region of Valencia is 542, once the pseudo-municipalities were created this figure was reduced by 18. Thus our sample consists of 524 municipalities, 17 of which are our artificial pseudo-municipalities.

Table A2. Pseudo-municipalities

| INE code | Municipality                 | Province  |
|----------|------------------------------|-----------|
| 3005     | <b>Albatera</b>              | Alicante  |
| 3904     | San Isidro                   | Alicante  |
| 3014     | <b>Alacant/Alicante</b>      | Alicante  |
| 3050     | Campello, El                 | Alicante  |
| 3015     | <b>Almoradí</b>              | Alicante  |
| 3903     | Montesinos, Los              | Alicante  |
| 3077     | <b>Fondó de las Neus, El</b> | Alicante  |
| 3078     | Hondón de los Frailes        | Alicante  |
| 3093     | <b>Novelda</b>               | Alicante  |
| 3114     | Romana, La                   | Alicante  |
| 3099     | <b>Orihuela</b>              | Alicante  |
| 3902     | Pilar de la Horadada         | Alicante  |
| 3013     | Algueña                      | Alicante  |
| 3105     | <b>Pinós, El/Pinoso</b>      | Alicante  |
| 12101    | San Rafael del Río           | Castellón |
| 12121    | <b>Traiguera</b>             | Castellón |
| 12124    | Vall d'Alba                  | Castellón |
| 12128    | <b>Vilafamés</b>             | Castellón |
| 12902    | Sant Joan de Moró            | Castellón |
| 12135    | <b>Villareal</b>             | Castellón |
| 12901    | Alquerías del Niño Perdido   | Castellón |
| 12068    | Herbés                       | Castellón |
| 12080    | <b>Morella</b>               | Castellón |
| 46116    | Eliana, l'                   | Valencia  |
| 46202    | <b>Pobla de Vallbona, la</b> | Valencia  |
| 46124    | Fontanars dels Alforins      | Valencia  |
| 46184    | <b>Ontinyent</b>             | Valencia  |
| 46190    | <b>Paterna</b>               | Valencia  |
| 46903    | San Antonio de Benagéber     | Valencia  |
| 46058    | Benifairó de les Valls       | Valencia  |
| 46122    | <b>Faura</b>                 | Valencia  |
| 46007    | <b>Albal</b>                 | Valencia  |
| 46065    | Beniparrell                  | Valencia  |
| 46197    | <b>Polinyà de Xúquer</b>     | Valencia  |
| 46904    | Benicull de Xúquer           | Valencia  |

Note: The municipalities in bold are those that existed in 1860.

## Census of Floridablanca (Census of 1787)

The Census of Floridablanca contains information on the ancien regime. With data for 1787, it includes a huge amount of information covering the total population by town (or population entity) and by gender, its structure by age groups and its distribution by professions. It also includes information on the administrative characteristics of the population entities, such as category (*ciudad, villa, lugar, aldea,...*), person in authority (*alcalde mayor, alcalde ordinario, gobernador, ...*) and jurisdiction (royal or lordship).<sup>32</sup> On the former kingdom of Valencia, it provides information on a total of 550 population entities.<sup>33</sup> These include 9 that today belong to the Comunitat Valenciana but were then part of other administrations.<sup>34</sup>

As regards one of our main explanatory variables – jurisdiction – the census carried no information of any kind for 43 of today’s municipalities. These mainly involve two types of cases: a) 15 of them are included in our pseudo-municipalities (see above), so we assume the same jurisdiction as their pseudo-municipality partner/s (see Table A3), and b) for the remaining 28 municipalities the strategy is twofold: in some cases we can identify the municipality they belonged to before becoming independent and then apply the same jurisdiction, otherwise we search for historical information from different sources to discover the type of jurisdiction in the past, assuming they existed in 1787 (see Table A4).<sup>35</sup> Once this is done, and taking into account the 17 artificial pseudo-municipalities, we have information on type of jurisdiction for all 524 municipalities in the sample.

---

<sup>32</sup> *Realengo, señorío secular, señorío religioso* and *órdenes militares* are the main types of jurisdiction.

<sup>33</sup> There were 129 in the today’s province of Alicante, 141 in Castellón and 280 in Valencia. These towns or entities were in turn grouped into larger administrative entities called *partidos* or districts (13) and *intendencias* or regions (the kingdom of Valencia was an *intendencia*). Moreover, some entities were «free» or exempted.

<sup>34</sup> Sax and Villena (Murcia); Camporrobles, Caudete de las Fuentes, Fuenterrobles, Requena, Utiel, Venta del Moro and Villagordo de Cabriel (Cuenca) joined Valencia in the mid-nineteenth century.

<sup>35</sup> Torrevieja, which was created from lands belonging to Orihuela, Guardamar, Rojals and Almoradí (all of them *realengos*), is considered to be royal jurisdiction, although before the nineteenth century there was no population, just surveillance towers (and salt mines). In the cases of Geldo and Serra d’En Galceran, for some reason the census does not supply any information about jurisdiction. Based on information obtained from alternative sources we assign to these two towns a lordship jurisdiction.

Table A3. Municipalities without jurisdiction in the Census of Population of 1787 which belong to a pseudo-municipality

| INE code | Municipality               | Same jurisdiction as pseudo-municipality partner |
|----------|----------------------------|--|
| 3013     | Algueña                    | <b>3105 Pinós, El/Pinoso (3089 Monòver)</b>      |
| 3050     | el Campello                | 3014 Alicante/Alacant                            |
| 3078     | Hondón de los Frailes      | <b>3077 Fondó de les Neus (3019 Asp)</b>         |
| 3114     | la Romana                  | 3093 Novelda                                     |
| 3902     | Pilar de la Horadada       | 3099 Orihuela                                    |
| 3903     | Los Montesinos             | 3015 Almoradí                                    |
| 3904     | San Isidro                 | 3005 Albaterra                                   |
| 12101    | San Rafael del Río         | 12121 Traiguera                                  |
| 12124    | Vall d'Alba                | 12128 Vilafamés                                  |
| 12901    | Alquerías del Niño Perdido | 12135 Vila-Real                                  |
| 12902    | Sant Joan de Moró          | 12128 Vilafamés                                  |
| 46116    | l'Elia                     | 46202 la Pobla de Vallbona                       |
| 46124    | Fontanars dels Alforins    | 46184 Ontinyent                                  |
| 46903    | San Antonio de Benagéber   | 46190 Paterna                                    |
| 46904    | Benicull de Xúquer         | 46197 Polinyà de Xúquer                          |

Table A4. Municipalities without jurisdiction in the Census of Population of 1787 which do not belong to a pseudo-municipality

| INE code    | Municipality                                      | INE code | Municipality              |
|-------------|---|----------|---------------------------|
| 3004        | Aigües  | 12049    | Costur                    |
| 3011        | Alfàs del Pi                                      | 12069    | Higuera                   |
| 3012        | Algorfa   | 12102    | Santa Magdalena de Pulpis |
| 3016        | Almudaina   | 12114    | Torás                     |
| 3051        | Campo de Mirra/Camp de Mirra, El                  | 46046    | Barx                      |
| 3052        | Cañada  | 46048    | Bellreguard               |
| 3062        | Daya Vieja  | 46082    | Canet d'En Berenguer      |
| <b>3077</b> | <b>Fondó de las Neus, El/Hondón de las Nieves</b> | 46087    | Casas Altas               |
| <b>3105</b> | <b>Pinós, El/Pinoso</b>                           | 46088    | Casas Bajas               |
| 3120        | San Miguel de Salinas                             | 46089    | Casinos                   |
| 3121        | Santa Pola  | 46108    | Chera                     |
| 3122        | San Vicente del Raspeig/Sant Vicent del Raspeig   | 46141    | Higuera                   |
| 3133        | Torre Vieja                                       | 46149    | Losa del Obispo           |
| 3138        | Verger, El  | 46224    | Segart                    |

Note: Both El Pinós and El Fondó de les Neus are the main population entities in their respective pseudo-municipalities. However, these two municipalities did not exist back in 1787, since they were created by breaking away from Monòver (1826) and Asp (1839) respectively.

As for the other variables included in our analysis apart from the administrative ones (i.e. total population, population by gender, population by age group and by profession), there are two different situations to be considered. On the one hand, we have the same problem as before given that the census has no information of any kind for the 28 municipalities shown in Table A4. Although we have been able to compile past administrative information for these places, we cannot do this for the demographic variables. In addition, there are 21 municipalities with administrative information (category, authority, jurisdiction, district, regional authority), but no information on population and its structure and professions (see Table A5). This means that with some econometric specifications we are left with a minimum sample of  $(524-28-21) = 475$  municipalities.

Table A5. Municipalities in the Census of 1787 with incomplete information

| INE code | Municipality         | INE code | Municipality         |
|----------|----------------------|----------|----------------------|
| 3064     | Dolores              | 46054    | Benetússer           |
| 3118     | San Fulgencio        | 46078    | Burjassot            |
| 12028    | Benicasim/Benicàssim | 46098    | Corbera              |
| 12060    | Figueroles           | 46125    | Fortaleny            |
| 12097    | Sacañet              | 46152    | Llocnou de la Corona |
| 12110    | Teresa               | 46186    | Paiporta             |
| 46002    | Ador                 | 46196    | Pinet                |
| 46013    | Alboraia             | 46197    | Polinyà de Xúquer    |
| 46014    | Albuixech            | 46223    | Sedaví               |
| 46022    | Alfagar              | 46237    | Tavernes Blanques    |
| 46032    | Almàspera            |          |                      |

### Census of 1887

The census of 1887 contains information not only about the population of the municipalities but also its distribution between the core entity (city, town) and the minor entities (village, hamlet, homestead, mills and isolated buildings). In this case there is information on all 524 municipalities in our sample.

Using data from the *Nomenclator* of Spain for 1887, we created a variable that measures the percentage of population living in the core entity. In the computation of this variable it is important to bear two things in mind.

(i) Pseudo-municipalities. In order to compute this dispersion measure, only the core entity of the main municipality is taken into account (see Table A2).

(ii) Municipalities that were independent in 1887 but today belong to other municipalities (see Table A6). In order to compute the dispersion measure, the core entities of both municipalities in 1887 are taken into account.

Table A6. Completing the information on jurisdiction

| Independent municipalities in 1887  | INE code     | Current municipality   |
|---|--------------|------------------------|
| Adsubia<br>Forna  | <b>3001</b>  | Adsubia                |
| Daya Nueva o Vieja<br>Puebla de Rocamora  | <b>3061</b>  | Daya Nueva             |
| Sela y Mirarrosa<br>Miraflor  | <b>3901</b>  | Poblets, Els           |
| Alcudia de Veo<br>Veo   | <b>12006</b> | Alcudia de Veo         |
| Montanejos<br>Campos de Arenoso   | <b>12079</b> | Montanejos             |
| Pobla de Benifassà, La<br>Ballestar<br>Bojar<br>Corachar<br>Fredes                  | <b>12093</b> | Pobla de Benifassà, La |
| Rosell<br>Bel   | <b>12096</b> | Rosell                 |
| Albaida<br>Aljorj   | <b>46006</b> | Albaida                |
| Albuixech<br>Mahuella   | <b>46014</b> | Albuixech              |
| Gandia<br>Beniopa<br>Benipeixcar  | <b>46131</b> | Gandia                 |
| Valencia<br>Benifaraig<br>Borbotó<br>Campanar<br>Masarrochos<br>Villanueva del Grau | <b>46250</b> | Valencia               |



### Other relevant information

As well as the information on literacy levels, type of jurisdiction, whether or not the municipality was Morisco and the official language, our data set also includes information on other demographic and geographical variables for the 524 municipalities of Valencia, taken mainly from the population censuses. Table A7 summarizes the sample of the data set.

Table A7. Summary

|                                   | <b>Number of municipalities</b> |
|-----------------------------------|---------------------------------|
| <b>Dependent variables</b>        |                                 |
| Literacy (1860)                   | 524                             |
| Literacy (1900)                   |                                 |
| Literacy (1930)                   |                                 |
| <b>Variables of interest</b>      |                                 |
| Jurisdiction                      | 524                             |
| Morisco or non-Morisco            | 524                             |
| Official language                 | 524                             |
| <b>Control variables</b>          |                                 |
| Population and professions (1787) | 474                             |
| Population (1860)                 | 524                             |
| Settlement patterns (1887)        | 524                             |
| Geographical information          | 524                             |