

Ben Salah, Mhamed; Chambru, Cédric; Fourati, Maleke

**Working Paper**

## The colonial legacy of education: Evidence from of Tunisia

Graduate Institute of International and Development Studies Working Paper, No. HEIDWP13-2022

**Provided in Cooperation with:**

International Economics Section, The Graduate Institute of International and Development Studies

*Suggested Citation:* Ben Salah, Mhamed; Chambru, Cédric; Fourati, Maleke (2022) : The colonial legacy of education: Evidence from of Tunisia, Graduate Institute of International and Development Studies Working Paper, No. HEIDWP13-2022, Graduate Institute of International and Development Studies, Geneva

This Version is available at:

<https://hdl.handle.net/10419/266127>

**Standard-Nutzungsbedingungen:**

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

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

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

**Terms of use:**

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

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

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



—  
INSTITUT DE HAUTES  
ÉTUDES INTERNATIONALES  
ET DU DÉVELOPPEMENT  
GRADUATE INSTITUTE  
OF INTERNATIONAL AND  
DEVELOPMENT STUDIES

---

**Graduate Institute of International and Development Studies**  
**International Economics Department**  
**Working Paper Series**

Working Paper No. HEIDWP13-2022

# **The Colonial Legacy of Education: Evidence from of Tunisia**

**Mhamed Ben Salah**  
Geneva Graduate Institute

**Cédric Chambru**  
University of Zurich

**Maleke Fourati**  
Mediterranean School of Business, South Mediterranean University

**May 2022**

Chemin Eugène-Rigot 2  
P.O. Box 136  
CH - 1211 Geneva 21  
Switzerland

# The Colonial Legacy of Education: Evidence from of Tunisia\*

MHAMED BEN SALAH<sup>†</sup> 

CÉDRIC CHAMBRU<sup>‡</sup> 

MALEKE FOURATI<sup>§</sup> 

May 2022

## Abstract

We study the effect of exposure to colonial public primary education on contemporary education outcomes in Tunisia. We assemble a new data set on the location of schools with the number of pupils by origin, along with population data during the French protectorate (1881–1956). We match those with contemporary data on education at both district and individual level. We find that the exposure of local population to colonial public primary education has a long-lasting effect on educational outcomes, even when controlling for colonial investments in education. A one per cent increase in Tunisian enrolment rate in 1931 is associated with a 1.69 percentage points increase in literacy rate in 2014. Our results are driven by older generations, namely individuals who attended primary schools before the 1989/91 education reform. We suggest that the efforts undertaken by the Tunisian government after independence to promote schooling finally paid off after 40 years and overturned the effects of history.

**JEL classifications:** D10, N37, N47.

**Keywords:** Colonial investment; Primary education; Tunisia.

\*We thank Denis Cogneau, Rasha Shakra, Lore Vandewalle, and Rémi Viné as well as seminar audiences at the Graduate Institute Geneva, University of Geneva, and MASE-ESSAI Carhage for many helpful comments and suggestions. All errors are our own.

<sup>†</sup>Mhamed Ben Salah, Department of International Economics, Graduate Institute of International and Development Studies, Maison de la paix, Chemin Eugène-Rigot 2A, 1211 Genève, Switzerland. Email: [mhamed.ben@graduateinstitute.ch](mailto:mhamed.ben@graduateinstitute.ch).

<sup>‡</sup>Cédric Chambru, Department of Economics, University of Zurich, Zürichbergstrasse 14, 8032 Zürich, Switzerland. Email: [cedric.chambru@econ.uzh.ch](mailto:cedric.chambru@econ.uzh.ch).

<sup>§</sup>Maleke Fourati, Mediterranean School of Business, South Mediterranean University, Avenue De L'Euro, les Berges du Lac 2, Tunis, Tunisia. Email: [maleke.fourati@msb.tn](mailto:maleke.fourati@msb.tn).

# 1 Introduction

Economic performance has been shown to be persistent and long-lasting (Michalopoulos and Papaioannou 2017; Nunn 2009; Sequeira et al. 2020). In particular, colonial institutions play a major role in explaining current economic outcomes (Acemoglu, Johnson et al. 2001; Banerjee and Iyer 2005). Along this line, this paper investigates the effect of local population exposure to colonial public primary education in Tunisia on current educational outcomes. We assemble a unique data set on the location of colonial schools with the number of pupils and teachers by origin, along with population data including Tunisians and Europeans in 1931 during the French protectorate (1881–1956). We match these data with contemporary census data at the district level. We find that a one per cent increase in exposure of local population to colonial public primary education, measured by enrolment rate in 1931, is associated with a 1.69 percentage points increase in literacy rate in 2014. Results account for European population and colonial investments in education, measured by the number of teachers per 1,000 inhabitants in 1931.<sup>1</sup>

We further take advantage of the 2019 Tunisian Population and Labour Force Survey, a nationally representative sample at the household level, to study the dynamic effect of the legacy of colonial public primary education. We find that a one per cent increase in Tunisian children enrolment in 1931 is associated with an increase of 0.9 percentage point in the probability of being literate in 2019. In addition, we decompose the population in age category using a 10 year gap from 10–19 to 70–79. We document that the older the age category, the higher the association between Tunisian enrolment in 1931 and the probability of being literate in 2019. The relationship becomes no longer significant for generations born after 1990. These findings seem in line with the successive post-independence educational reforms implemented by the Tunisian government to make primary education universally available (Bouttemont 2002; Eibl 2020; Tarifa 1971). The year 1990 coincides with the last important educational reform (1989–91) which made schooling compulsory for children between 6 and 16 year-old. We believe that these reforms overturned the lasting effect of colonial public investment in education 40 years after independence.

This paper is related to the broad literature on the persistent effects of colonial institutions on contemporary outcomes (Acemoglu, Gallego et al. 2014; Banerjee and Iyer 2005; Delleaux 2010; La Porta et al. 1999). In a seminal work, Huillery (2009) investigates the long term impact of colonial investment on education, health, and infrastructure in French West Africa. She highlights the persistence of investment as the main mechanism, meaning that current public investments are predominant in areas where colonial investment was already prevailing at the time. Rocha et al. (2017) study the long-run effect of the settlement policy implemented in Brazil to attract more educated immigrants in some municipalities at the turn of the 20th century. They show that the policy resulted in higher levels of schooling and income per

<sup>1</sup> In a placebo exercise, we find no correlation between Italian enrolment rate in 1931 and literacy rate in Tunisia in 2014. Our findings are also robust to alternative specifications such as varying the variables of interests, and to different district divisions.

capita one century later. Documenting the effect of the partition of Cameroon, Dupraz (2019) find that, in 2005, individuals born after 1970 were more likely to finish high school and have a high-skilled occupation if they were born in the former British part of the country.

Our paper builds on and expands this literature by exploiting the direct exposure of local population to colonial education. To the best of our knowledge, we are the first to investigate the effect of exposure to colonial public primary education while controlling for the level of investment rather than the effect of investment itself. We believe that this is a significant contribution, as local children enrolment in colonial public primary schools may play a more important role than the mere presence of a school itself. Finally, we contribute to the literature by showing that the persistence is driven by older generations, and that the persistence no longer holds for younger ones. Aligned with Chaudhary and Garg (2015) results,<sup>2</sup> we believe that effective policies can counteract the persistent effect of colonial investment in education. Last but not least, the creation of colonial school-level data and district-level population data is also a contribution of this paper to the existing literature on economic history. In particular, we create a repository of historic places, which contains geo-coded information on more than 1,300 localities across Tunisia. These data provide avenue for further research on Tunisia.<sup>3</sup>

The remainder of the paper is organised as follows. Section 2 provides historical background on the French protectorate in Tunisia and discusses the history of education since the early 19th century. Section 3 describes the sources and methodology used to assemble our data set. Section 4 introduces the empirical strategy, quantifies the effect of exposure to colonial education on contemporary education outcomes. Section 5 discusses our findings and Section 6 concludes.

## 2 Historical background

### 2.1 The French protectorate

The French protectorate in Tunisia lasted from 1881 to 1956. Before that, Tunisia was under the Ottoman rule since 1574. Between 1837 and 1855, Ahmed Bey, the Bey of Tunis, delved Tunisia into sustained public expenditures, leaving the country with considerable financial difficulties. These financial difficulties forced the regency of Tunis to contract its first loans. The two successive Beys carried on expansive policies. France took advantage of the country financial crisis combined with the decline of the Ottoman empire to seize power. During the Congress of Berlin, in 1878, Germany and Britain allowed France to invade Tunisia. In February 1881, a conflict between a Tunisian tribe, the *Kroumirs*, and an Algerian one, the *Ouled Nahd*, provided the ground for a French military intervention. In just few months, it resulted in the signing of the Treaty of Bardo. In 12 May 1881, Tunisia became officially a French protectorate. The

<sup>2</sup> Chaudhary and Garg (2015) show that the persistent effect of colonial investment in education holds for a couple of decades in India, but effective policies counterbalanced this long term effect.

<sup>3</sup> There were only a small number of missionary schools in Tunisia during the colonial rule. We do not discuss the rich literature on the long-term consequences of missionary education (e.g. Bauer et al. 2022; Caicedo 2019; Cogneau and Moradi 2014; Jedwab et al. 2022; Waldinger 2017; Wantchekon et al. 2014; Wietzke 2015).

subsequent conventions of La Marsa (8 June 1883) transferred the administrative, judiciary and financial responsibilities from the Bey to the French government.

Online Appendix Table A.1 depicts the evolution of the population living Tunisia by origin between 1881 and 1956. By 1881, Tunisia had a population of about 850,000, including 19,000 Europeans. The presence of Europeans in Tunisia started to be sizeable few decades earlier with the arrival of the Maltese in 1816, followed by the Italians, almost exclusively Sicilians, in the 1820s. As for the French population, immigration mostly began after the end of the conquest of Algeria in 1857 (Planel 2015). During the first decades of the protectorate, roughly 2,000 French colons settled in Tunisia every year.<sup>4</sup> In 1906, Italians represented 63 per cent of the foreign population and were, until 1931, the largest European community in Tunisia (online Appendix Table A.1).<sup>5</sup> Out-migration began in 1954, two years before the independence in 20 March 1956. In 1958, more than half of the French population had already left (Wolkowitsch 1959). The process was somehow slower with the Italian population. Only 20 per cent had left by 1958, and 90 per cent by 1969. Nowadays, only 800 Italians living in Tunisia come from the pre-independence community (Davi 2000).

## 2.2 Colonial education

For centuries, religious spaces were the only source of education in Tunisia. The *Medersa* of the Al Zaytouna Mosque was created in 737, and played a primary role in education. In 1838, the first military and non-religious school was created.<sup>6</sup> Forty years later, in 1875, the Sadiki college was founded following the French schooling system. It was the first modern, non-religious as well as non-military educational structure in the country. In 1893, traditional elementary education given in the Qur'anic schools was however still prevailing with 961 *Kottebs*, including 108 in Tunis, gathering 15,000 children (Bourdand 1893, p. 294).<sup>7</sup>

European education started in Tunisia with the arrival of European migrants in the early 19th century. In 1830, a protestant school was founded by the British for the Maltese population. The same year, an Israelite school was founded by the London Jews Society. The first French school was created in 1840 by mother Émilie de Vialar and was a free girl's school. In 1843, the first Italian school was founded. A dozen of French schools, and a few other European schools were founded after that and until 1883. These were all private and congregational schools, almost exclusively located in Tunis and to some extent in Sousse and Sfax, two major coastal cities (Bannour 2001). There was therefore neither public nor private spread modern

<sup>4</sup> In the mid-1890s, the French protectorate implemented a colonisation policy targeting small land holders. However, contrary to Algeria, French never had access to agricultural land for free, which slowed down the settlement process. Centre des archives diplomatiques de Nantes, Tunisie, protectorat, 1er versement (hereafter CADN-Tun., 1er vers.), 1361 et 1362.

<sup>5</sup> The presence of Italians was a constant source of worries for the French administration that continuously sought for means to increase the presence of French settlers and/or to make sure that the interests of the Italian population were in line with those of France. CADN-Tun., 1er vers., 916 A et 2130 à 2134.

<sup>6</sup> The *École Polytechnique du Bardo* later became known as the *École Militaire du Bardo*.

<sup>7</sup> For an overview of the education system before the protectorate, see Arnoulet (1994), and Foncin (1882). Degorge (2002) provides a discussion of the progresses of public education during the protectorate. On the evolution of non-public schools during this period, see Damis (1974).

education system before the foundation of the *Direction Générale de l'Instruction Publique* in 1883 with Louis Machuel as head of the institution. The main goal of the colonial education system was to give French settlers the opportunity to study in a French schooling system. Historical documents highlight the fact that French education also aimed to attract European children in order to gain cultural influence over Italians, and Tunisian children not only for cultural reasons but also to build a relatively skilled labour force.<sup>8</sup>

In 1881, France enacted Jules Ferry's laws, making French public schools free, compulsory, and secular. After 30 years of presence in Algeria, France could benefit from the experience accumulated to set up their education system in Tunisia. Jules Ferry recommended to give a religious education to both European Catholics and Tunisian Muslims, to teach Arabic alongside with French, and to ensure that Tunisian teachers handled of Arabic and Islamic lessons (Sraieb 1993). These recommendations were born from France's failure in Algeria to bring local children to school.<sup>9</sup> Moreover, catechism were also encouraged for European population specifically in order to teach French to a large and very religious Italian population.

Online Appendix Table A.2 provides an overview of the evolution of colonial investment in public primary education. Between 1891 and 1931, the number of public primary school increased from 76 to 468 and the number of pupils from 8,208 to 67,089. Online Appendix Table A.2 also highlights the geographic diffusion of colonial investment in education. French schools were located in 163 districts over 264 in 1931, compared to only 43 in 1891. The share of Tunisian pupils increased from 32 per cent in 1891 to 52 per cent of the total number of pupils in 1931.

### 2.3 Education after independence

The post-independence government viewed education as a priority. The share of government budget allocated to education grew from 18 per cent in 1958–9 to 32 per cent in 1967 (Eibl 2020; Tarifa 1971). The government aimed a unified educational system throughout the country and to reach full enrolment in primary school by 1966 in a context of high fertility rate. It implies that approximately 980,000 children per year should be enrolled to school by 1966, whereas about 270,631 children attended primary school in 1956–7. One of the biggest constraint was the low number of teachers available, with only 6,195 teachers in 1957, whereas around 20,000 were needed in order to keep a reasonable ratio of 40 pupils per teacher. To reach this goal, the first main education reform after independence was designed and implemented in 1958.

The second main education reform in Tunisia occurred between 1989 and 1991, a few years after the political transition from the rule of president Habib Bourguiba (1956–87) to Zine El Abidine Ben Ali who ruled from 1987 to the revolution of 2011. The reform made school compulsory for children between 6 and 16 years old, and, among other things, reinforced

<sup>8</sup> On the diffusion of Italian schools during the colonial period, see Montalbano (2020). The widespread presence of Italian speaking schools was a concern for the colonial administration that feared the diffusion of the Italian influence and the risk to be overthrown. CADN–Tun., 1er vers., 1356 à 1358.

<sup>9</sup> Jules Ferry's vision was to make local populations feel safe and comfortable, integrating their religion and language to the curriculum so that they do not fear any threat of uprooting linked to French public schools.

secularisation in the schooling system by prohibiting religious veil and making mixed-sex classes compulsory. For the first time, Tunisia reached 4,300 primary schools in 1994–5, a number which remained stable ever since. By 2014, the enrolment rate of children aged 6–11 reached 99 per cent and the overall enrolment rate 95.8 per cent.<sup>10</sup> The government also put substantial efforts in terms of infrastructure and teachers training, which decreased the ratio of pupils per teacher from 42 to 17 between the 1960s and 2014.<sup>11</sup>

### 3 Data

The empirical analysis is based on a new data set on the location of colonial public primary schools in 1931, including the number of pupils and teachers by origin, along with population data, which we match with existing contemporary data on education at the district level. This section describes the sources and methodology used to create these databases.

#### 3.1 Units of analysis

As of 2014, Tunisia was divided into 264 districts (*délégations*), grouped in 24 governorates (*gouvernorats*), themselves grouped into six geographical regions: north east, north west, center east, center west, south east and south west (online Appendix Figure A.1). These administrative boundaries are the results of various reforms since the independence in 1956, and correspond only to some extent to the pre-existing administrative boundaries of the French administration (*contrôle civil*) and the Tunisian administration (*cheikhat* and *caïdat*) that co-existed between 1881 and 1956. In addition, several modern-day large urban areas, such as Gabès, Sfax, Sousse, or Tunis are divided into several districts. Historical sources, however, rarely contain information on the exact postal addresses of school or the spatial distribution of the population within the city. That means that it is impossible to allocate historical data across districts within the same city. Moreover, in tiny urban districts pupils could have easily shuttled from one district to the other within the same large city to attend school. Therefore, we merge these districts to obtain consistent administrative borders. We also decide to exclude the governorate of Tunis from our baseline sample. As the capital of Tunisia, Tunis concentrated an unusually high numbers of European settlers during the protectorate, and continued to attract the bulk of economic activities after independence (Belhedi 1999).<sup>12</sup> Overall, our baseline sample includes a total of 211 districts.<sup>13</sup>

<sup>10</sup> Ministère de l'Éducation, Tunisie, Statistiques scolaires, Données et indicateurs du secteur 2013–2014, [http://www.education.gov.tn/article\\_education/statistiques/stat2013\\_2014/Livre\\_Stat.pdf](http://www.education.gov.tn/article_education/statistiques/stat2013_2014/Livre_Stat.pdf) (Last accessed on 17 March 2022).

<sup>11</sup> For an overview of the the evolution of enrolment rate and literacy rate between 1956 and 2014, see online Appendix Table A.3.

<sup>12</sup> By the late 20th century, Tunis attracted three-fourth of the Tunisian students, accounted for 45 per cent of the car fleet, and concentrated about 40 per cent of the tertiary employment in the public sector and 57 per cent of the industrial employment. Tunis was also the major destination for Tunisian internal migrant workers after independence.

<sup>13</sup> In Table 6, we show that the results are robust to using the full sample of 264 districts, including Tunis.



### 3.2 Contemporary outcomes

We assemble a series of educational outcomes from the population censuses of 2004 and 2014, as well as from the 2019 Tunisian Labour Survey and the official statistics of the Tunisian Ministry for Education. Information on literacy rates at the district level are derived from the *Recensement général de la population et de l'habitat* carried out by the Tunisian national institute of statistics in 2004 and 2014.<sup>14</sup> We rely on the official statistics of the Tunisian Ministry for Education to obtain the number of teachers and pupils by district in 2014–5 for 2014–5.<sup>15</sup> Finally, we use individual level data from the 2019 Tunisian Labour Force Survey (*Enquête nationale sur la population et l'emploi 2019*) to capture age-group specific literacy rates.<sup>16</sup> We divide individuals into age bin with a 10-year increment starting at 10 years old. We code a dummy variable equal to one if the respondent is literate, and zero otherwise. Summary statistics are reported in Table 1.

### 3.3 Historical data

Our analysis is based on a novel city-level data set constructed from the 1931 colonial population censuses.<sup>17</sup> In the censuses, French and European population were counted at the *localité* level and the Tunisian population at the *cheikhhat* level, which were the smallest administrative divisions of the colonial Tunisian administration.

We complete these information with disaggregated data on the location of public primary schools and the number of pupils by gender and nationality in 1931, obtained from the *Statistique Générale de la Tunisie* elaborated by the *Direction générale de l'agriculture, du commerce et de la Colonisation*.<sup>18</sup> These annual reports contain a large amount of statistics on different topics, including weather, health, education, trade, agricultural and industrial production. However the content and/or the level of details varies from one year to another, and detailed statistics on education is no longer reported after 1931.<sup>19</sup>

To the best of our knowledge, there is no available resources to accurately geo-code and match Tunisian historical places with contemporary administrative units.<sup>20</sup> Therefore, we manually created a dedicated database, which encompasses every place mentioned in the

<sup>14</sup> <http://www.ins.tn/enquetes/recensement-general-de-la-population-et-de-lhabitat-2014> (Last accessed on 17 March 2022).

<sup>15</sup> <http://www.education.gov.tn/?p=688&&lang=en> (Last accessed on 17 March 2022).

<sup>16</sup> <https://www.ilo.org/surveyLib/index.php/catalog/LFS#country=217> (Last accessed on 17 March 2022).

<sup>17</sup> Bibliothèque nationale de France (hereafter BnF), NUMM-994559, Dénombrement de la population civile indigène au 22 mars 1931. BnF, 8-SG-14304, Dénombrement de la population civile européenne et tunisienne en Tunisie au 12 mars 1936. We choose the latest available point in time, 1931, in order to estimate the maximal exposure.

<sup>18</sup> BnF, 8-O3I-371, Statistique Générale de la Tunisie, 1931.

<sup>19</sup> We were unable to identify any relevant official publication providing disaggregated information on education after the early 1930s. We searched in the collections of the French National Library, the *Centre des Archives diplomatiques de Nantes* (CADN), which gathers the collections of documents produced by the French Protectorates in Tunisia, and the Tunisian National Archives.

<sup>20</sup> Online tools, such as GeoNames ([www.geonames.org](http://www.geonames.org)), return an extremely high number of false-positive due to (i) errors and approximations made by French colonial administrators in the transliteration of Tunisian places; (ii) the numerous changes in the name of places after 1956.

various population censuses and statistical reports on education. We identify about 1,300 distinct places for which we provides the geographic coordinates. To circumvent the issue of boundaries variations across time and the mismatch between historical and contemporary administrative units, we use this database to match historical data with modern-day districts based on the geographic location. Overall, we are able to geo-locate 98 per cent of the population across our variables (online Appendix Table A.4).<sup>21</sup>

We measure the level of exposure to colonial public primary education as the share of non-adult population (individuals aged 0–18) attending primary schools in 1931. We compute separate enrolment rate for French, Italian and Tunisian pupils. Because most of the Jewish population left Tunisia after 1956 (Messika et al. 2015), we only consider Tunisian children reported as Muslim in the official statistics. To calculate the school-age population, we rely on the information on the age-structure of the population provided in the colonial censuses from 1931.<sup>22</sup> We further calculate the number of teachers and the number of schools per 1,000 Tunisian inhabitants in 1931 by district. Summary statistics are reported in Table 1.

### 3.4 Geographic and socioeconomic controls

We also combine a set of geographic and historical variables at the district level to control for potential covariates that could affect both exposure to colonial education and education outcomes today. This set of variables includes the distance to the coast, the average slope, and the soil suitability for the cultivation of wheat and olive (FAO-GAEZ), the average rainfall per year between 1901 and 1911 (Harris et al. 2020), the number mines in 1905, the number of railway stations in 1926, and precolonial urban centres (André 1981).<sup>23</sup>

## 4 Empirical strategy and results

**Prima facie evidence.** In Figure 1, we document the unconditional relationship between the the exposure of local population to colonial public primary education and current educational outcomes (Figure 1). There is a positive (0.278) and significant relationship between the enrolment rate of Tunisian pupils in 1931 and the literacy rate in 2014.<sup>24</sup>

<sup>21</sup> For an overview of the geographical distribution of these variables, see online Appendix Figure A.2.

<sup>22</sup> The demographic data on the Tunisian age-structure is available at the *caïdat* level ( $n = 33$ ) in the Table V, *Répartition de la population musulmane et israélite d'après le sexe et l'âge* (pp. 102–5). The demographic data on the European age-structure is available at the *circonscription* level ( $n = 20$ ) in the Annex 9, *Répartition par circonscriptions administratives de la population civile européenne, selon l'âge* (p. 94). We assume that there is no significant variations in the age structure within each *caïdat* and *circonscription*, and match each place with its respective administrative unit to derive school-age population before aggregating the data at the district level.

<sup>23</sup> L'Écho des Mines et de la Métallurgie, Lundi 9 janvier 1905, 32e année, numéro 1648, pp. 18–20. BnF, IFN-53081141, Tunisie physique et administrative, dressée par la Compagnie Française des Chemins de fer Tunisiens, 1926.

<sup>24</sup> The figure displays a large number of zeros indicating that, in many districts, there was no school in 1931. In online Appendix Table A.2, we show that the number of districts with at least one school increased significantly between 1891 and 1931, but only reached 163 even after merging small districts. In Table 6, we provide evidence that our results are not driven by fuzzy zeros.

## 4.1 Empirical Strategy

We estimate the long term effect of exposure of local population to colonial public primary education on current educational outcomes by running the following regression:

$$\text{Literacy rate}_{ijk2014} = \alpha + \beta_1 \log(\text{Tunisian enrolment rate}_{ijk1931}) + \beta_2 X_{ijk} + \eta_k + \epsilon_{ijk} \quad (1)$$

where our dependent variable is the literacy rate of the population over 10 years old in 2014 in district  $i$  in governorate  $j$  in region  $k$ . The independent variable is the logarithm of Tunisian enrolment rate in 1931 in district  $i$  in governorate  $j$ .<sup>25</sup> The vector  $X_{ijk}$  captures the set of control variables in district  $i$  in governorate  $j$ . We control for the logarithm of European settlers living in district  $i$  in 1931, to account for French settlement location that may be correlated with schools location choice and long term economic development. We further control for the logarithm of the number of teachers per 1,000 Tunisian inhabitants to account for the possible role of investment in education by French settlers. In doing so, we are able to disentangle the effect of colonial investment in education from the effect of exposure to colonial education. We also add a set of geographical and economic controls, which includes a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, the average rainfall per year between 1901 and 1911, the number of mines in 1905, the number of railway stations in 1926, and a dummy that equals one for pre-colonial major urban centres.<sup>26</sup>  $\eta_k$  are region fixed effects that account for time-invariant region characteristics ( $k = 6$ ). In our baseline estimate, we use robust standard errors clustered at the governorate level ( $j = 23$ ).<sup>27</sup>

## 4.2 Results

**Baseline results.** Table 2 reports the results of our baseline regression for Equation 1. In column 1, we regress the logarithm of Tunisian enrolment rate in 1931 on the literacy rate in 2014. The coefficient indicates that a one per cent increase in enrolment rate in 1931 is associated with a 3.16 points increase in literacy rate in 2014. To account for the fact that the location of French settlements and school location choices were not random, we include, in columns 2 and 3, the logarithms of the European population and the number of teachers per 1,000 inhabitants in 1931. In column 4 we capture geographic and economic disparities across the country by adding our set of geographical and historical controls. Our preferred baseline specification is column 5, which includes region fixed effects to account for region-specific factors. We find that a one per cent increase in enrolment rate in public primary schools in

<sup>25</sup> We add 1 to our variables prior to their log transformation to retain zero-valued observations. Using the inverse hyperbolic sine transformation rather than the log transformation does not alter our results.

<sup>26</sup> Mines and Railways are calculated within a 10 km radius from the district.

<sup>27</sup> In online Appendix Section B.1, we discuss the possibility that spatial autocorrelation leads to a spurious correlation and an underestimation of standard errors (Kelly 2020). We show that our results are robust to using a Bartlett spatial weights correction to compute standard errors.

1931 is associated with a 1.69 percentage points increase in literacy rate in 2014 – that is an additional 850 individuals being literate in 2014.<sup>28</sup>

**Placebo test.** Despite the fact that we account for the settlers’ presence, colonial investment in education and a set of controls, differences in current Tunisian literacy rate may still be explained by omitted variables impacting both enrolment rate during colonial time and current literacy. The history of Tunisia enables us to run a placebo test to address this concern. Indeed, during the protectorate, Italian population was important and outnumbered French population until 1931 (online Appendix Table A.1). Moreover, French schools enrolled not only French and Tunisian children, but also Italian children, most of whom left Tunisia shortly after 1956 (see Section 2). Therefore, one should expect the exposure of Italians to colonial public primary education to have no predictive power on current educational outcomes. Table 3 present the results from estimating Equation 1, but with the logarithm of Italian and French enrolment rate in 1931 as the main explanatory variables. Column 1 reports our baseline result from Table 2 (column 5). Columns 2 and 3 report the results with the logarithm of Italian and French enrolment in 1931, respectively. The coefficients are both close to zero and not statistically significant. We interpret this finding as indicating that exposure of the Tunisian population to colonial public primary education is the main factor explaining our results.

**Persistence of investment.** The literature has pointed out to the persistence of colonial investment in the sense that past colonial investment strongly affects contemporaneous investment (Huillery 2009; Wietzke 2015). However, there is no clear consensus on that matter. For instance, Calvi and Mantovanelli (2018) find that the long lasting effect of access to health care on contemporary health outcomes is not driven by the persistence of health infrastructure. In this paper, we test the persistence of investment by investigating whether the level colonial investment in education may explain contemporaneous one. We regress the number of schools (teachers) per 1,000 Tunisian inhabitants in 2014 on the number of schools (teachers) per 1,000 Tunisian inhabitants in 1931, using the same empirical framework than in column 5 of Table 2. Results are displayed in Table 4. We find that colonial investment is not associated with the current number of schools per 1,000 inhabitants (column 1), and is negatively and significantly associated with the current number teachers per 1,000 inhabitants in 2014 (column 2). It should be noted that this result could be specific to the case of Tunisia. As we discuss in Section 2.3, the post-independence government massively invested in primary education to reach full enrolment rate for primary school age children throughout the country. We believe these investments were significant enough over the last 60 years to offset pre-existing regional disparities at independence

### 4.3 Robustness checks

In this section, we discuss various robustness checks to ensure that our baseline results are robust to a variety of alternative specifications.

<sup>28</sup> In 2014, the mean population in a district is 50,612 inhabitants and the mean literacy rate 76 per cent. Thus, a 1.69 percentage points increase corresponds to  $50,612 \times 1.69 = 855$  individuals.

**Alternative measures of educational outcomes.** We replicate Equation 1 using alternative contemporary educational outcomes and report the results in Table 5. Column 1 replicates our baseline estimate. Columns 2 and 3 report the effect of the exposure to primary education in 1931 on the literacy rate in 2004, and the literacy rate of 15 to 29 years old population in 2014. We find a positive and significant association across all specifications between the exposure to colonial education in 1931 and contemporary educational outcomes. It is noteworthy that the effect is much smaller on the literacy rate of the younger generation (column 2). This suggests that the long lasting effect of colonial public primary education in Tunisia begins to fade away in the recent decades, possibly as a result of the recent reforms undertaken by the Tunisian government.<sup>29</sup>

**Alternative divisions of districts.** To ensure that our results do not depend on our baseline district division, we perform alternative sensitivity tests. Results are reported in Table 6. Column 1 reports the baseline results from column 5 in Table 2. A major concern is that pupils might have been going to a school in neighbouring district, because they lived close to a district border, and/or the school itself is located near a district border. To address this issue, we identify every school lying with 5 km from a district border and re-distribute its pupils across those districts. We first compute a 5 km radius area around colonial schools and calculate the share of that area in each neighbouring districts. We use this share to re-distribute pupils of each school among districts before re-estimating Equation 1 (column 2). Column 3 reports the result from estimating Equation 1 with the region of Tunis.<sup>30</sup> Finally, we run our model using the full sample of 264 contemporary districts, without merging small one and including Tunis. Result is shown in column 4. The coefficient of interest remains positive and statistically significant across all specifications. The fact that it is smaller and less significant in column 4 is not surprising given that, as we argue, using the full sample of districts likely introduces some measurement errors in the explanatory variable.

**Spatial autocorrelation.** In a recent set of papers, Kelly (2019, 2020) points out the risk of inflated  $t$  statistics in historical paper using spatial data.<sup>31</sup> We are confident that using geographic controls and region fixed effect reduce the impact of any potential spatial trend. This, however, does not mean we are not possibly underestimating standard errors (Kelly 2020). We find that Moran's  $I$  statistic of the residuals of our baseline regression is positive and significant, suggesting that geography is driving part of our baseline estimates. However, the  $z$ -score of 2.9 can be considered as relatively low (Kelly 2019). To control for this, we adjust our standard errors to account for spatial correlation using a Conley (1999) procedure but with a matrix allowing for a linear decay of spatial weights with distance (Bartlett correction). Using different distance cut-offs, we show that our standard errors do not vary significantly (online Appendix Table A.5).

<sup>29</sup> In Table 5, we document a positive and significant effect on the literacy rate of age-group 15–29 in 2014– born between 1985 and 1999– at the district level (column 3). Results from columns 8 to 11 in Table 8 indicate that this relationship is driven by individuals born before 1990 (see Section 5).

<sup>30</sup> In our baseline estimate, we omit Tunis because it is likely to be an outlier (see Section 3).

<sup>31</sup> In online Appendix Section B.1, we provide a lengthier discussion on the issue of spatial autocorrelation.

**Ratio as a dependent variable.** A final concern arises from the fact that the dependant variable is a ratio. Linear regressions can lead to misspecification and uncertain statistical validity (see online Appendix Section B.2). To circumvent this caveat, we use a beta regression to estimate our baseline model. Results are reported in online Appendix Table A.6. Coefficient remains positive and statistically significant across all specifications. The magnitude of the effect of exposure to colonial education is also in the same range than our baseline estimate. On average, one percentage point increase in enrolment in primary school in 1931 is associated with 1.25 percentage points increase in the literacy rate in 2014.

## 5 Discussion

Our main findings highlight a robust association between the exposure of local population to colonial public primary education and current educational outcomes. We further use the 2019 Tunisian Population and Labour Force Survey, a nationally representative sample at the household level, to study the dynamic effect of the legacy of colonial public primary education.<sup>32</sup> Once merged with our historical data set, we have a sample of 495,039 respondents of 10 years old and more.<sup>33</sup> We estimate the following regression:

$$\begin{aligned} Literacy\ rate_{hijk\ 2019} = & \alpha + \beta_1 \log(Tunisian\ enrolment\ rate_{ijk\ 1931}) \\ & + \beta_2 X_{ijk} + \beta_3 X_h + \eta_k + \epsilon_{hijk} \end{aligned} \quad (2)$$

where our dependent variable is a dummy variable equal to one if the respondent  $h$  in district  $i$  in governorate  $j$  in region  $k$  reported being literate, and zero otherwise.  $X_h$  is a vector containing the age and gender of the respondent. All other variables are defined as in Equation 1.

Table 7 reports marginal effects from the logit regression for our key explanatory variable. The coefficient is positive and significant at the 5 per cent level, indicating that a one per cent increase in exposure of local population to colonial primary education is associated with a 0.922 percentage point increase in the probability of a respondent being literate. The survey further allows to test our model for different age-group categories. Results for the different age-cohort are reported in Table 8. We observe that the magnitude and significance tend to increase as we move along the age category. For individuals born before 1990, there is a strong and significant relationship between the exposure of Tunisians to colonial public primary education and the probability of being literate (columns 2 to 5). However, for individuals aged between 10 and 30 in 2019, the effect is close to zero and not significant (columns 6 and 7). This means that the effect of exposure to colonial public primary education on literacy rate vanished for individuals born after 1990, when the last significant education reform was

<sup>32</sup> Among other things, the survey contains information on the literacy rate of individuals and their district of residence. The information on the location of respondents is not publicly available. We had to send our files, data set and Stata do-files to the Tunisian Statistical Office, which returned us log files from estimating Equation 2.

<sup>33</sup> In 2019, the population of Tunisia was about 11.7 millions inhabitants, which means that the survey includes 4.2 per cent of the total population.

implemented in Tunisia (see Section 2.3). We believe these results relate to two plausible, non exclusive, alternatives: first, the intergenerational transmission of exposure to colonial public primary education tends to decrease across generations; second, the reforms carried out by the Tunisian government were ultimately successful to mitigate the inheritance of colonial period. By the mid-1990s, more than 85 per cent of an age-cohort attended primary school (online Appendix Table A.3)– 95 per cent in the mid-2000s. It is therefore hardly surprising that the exposure to colonial public primary education no longer explains current literacy level, which reaches almost 100 per cent among younger generation.

In Table 9, we document the relationship between exposure to colonial public primary education and the propensity to have a primary education. Columns 1 to 8 report marginal effects from logit regression. We find that, on average, there is a positive and significant effect of exposure to colonial public primary education and the propensity to have at least a primary school education (column 1). However, we find that the overall effect is driven by age-groups born before 1990, which attended primary school before the last important reform in 1989–91. This is consistent with the evidence on literacy rate shown in Table 8. Although we cannot directly test for it, these results suggest that these policies successfully managed to attenuate the persistent effect of colonial education. These findings are consistent with Chaudhary and Garg (2015) who show how effective education policies in India in the late 1960s nullified the effect British colonial expenditures in education.

## 6 Conclusion

We document a positive and significant long-lasting relationship between local population exposure to colonial public primary education and current literacy. We show that a one per cent increase in enrolment of Tunisian children in French schools in 1931 is associated with a 1.69 percentage points increase in literacy in 2014. We provide evidence that this influence tends to decay over time. There is no longer significant difference in literacy, for generations born after 1990 that can be traced to difference in exposure to colonial public primary education. We believe that the successive education reforms which took place after independence may explain this result, and particularly the 1989–91 reform that made schooling compulsory from 6 to 16 years old. From the early 1990s onward, exposure to colonial public primary education had no significant effect on the propensity to graduate from primary schools. We also find that colonial public investment in education, measured by the number of schools (teachers) per 1,000 Tunisian inhabitants, has no significant effect on the current level of investments in education. The results presented in this study are in line with those of Chaudhary and Garg (2015) who show how effective policies overturned the effect of colonial investments in primary education in India. In the Tunisian case, the efforts made by post colonial governments to provide universal primary education across the country took around 40 years to overcome the spatial disparities inherited from the colonial schooling system for the new generations. The overall effect on the entire population is likely to remain for a couple of decades.

## References

- Acemoglu, Daron, Francisco A. Gallego and James A. Robinson (2014). 'Institutions, Human Capital, and Development'. *Annual Review of Economics* 6.1, pp. 875–912. DOI: [10.1146/annurev-economics-080213-041119](https://doi.org/10.1146/annurev-economics-080213-041119).
- Acemoglu, Daron, Simon Johnson and James A. Robinson (2001). 'The Colonial Origins of Comparative Development: An Empirical Investigation'. *American Economic Review* 91.5, pp. 1369–1401. DOI: [10.1257/aer.91.5.1369](https://doi.org/10.1257/aer.91.5.1369).
- André, Nouschi (1981). 'Observations sur les villes dans le Maghreb précolonial'. *Cahiers de la Méditerranée* 23.1, pp. 3–21. DOI: [10.3406/camed.1981.927](https://doi.org/10.3406/camed.1981.927).
- Arnoulet, François (1994). 'L'enseignement congréganiste en Tunisie aux XIX<sup>e</sup> et XX<sup>e</sup> siècles'. *Revue des mondes musulmans et de la Méditerranée* 72, pp. 26–36. DOI: [10.3406/remmm.1994.1650](https://doi.org/10.3406/remmm.1994.1650).
- Banerjee, Abhijit and Lakshmi Iyer (2005). 'History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India'. *American Economic Review* 95.4, pp. 1190–1213. DOI: [10.1257/0002828054825574](https://doi.org/10.1257/0002828054825574).
- Bannour, Abderrazak (2001). 'Aperçu sur l'enseignement du français en Tunisie de 1830 à 1883'. *Documents pour l'histoire du français langue étrangère ou seconde* 27, pp. 1–13. DOI: [10.4000/dhfles.2604](https://doi.org/10.4000/dhfles.2604).
- Bauer, Vincent, Melina R. Platas and Jeremy M. Weinstein (2022). 'Legacies of Islamic Rule in Africa: Colonial Responses and Contemporary Development'. *World Development* 152, p. 105750. DOI: [10.1016/j.worlddev.2021.105750](https://doi.org/10.1016/j.worlddev.2021.105750).
- Belhedi, Amor (1999). 'Les disparités spatiales en Tunisie, état des lieux et enjeux'. *Méditerranée* 91, pp. 63–91. DOI: [10.3406/medit.1999.3088](https://doi.org/10.3406/medit.1999.3088).
- Bourdand, François (1893). *Tunisie et tunisiens*. Paris: J. Lefort, p. 368.
- Bouttemont, Cécile de (2002). 'Le système éducatif tunisien'. *Revue internationale d'éducation de Sèvres* 29, pp. 129–136. DOI: [10.4000/ries.1928](https://doi.org/10.4000/ries.1928).
- Caicedo, Felipe Valencia (2019). 'The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America'. *The Quarterly Journal of Economics* 134.1, pp. 507–556. DOI: [10.1093/qje/qjy024](https://doi.org/10.1093/qje/qjy024).
- Calvi, Rossella and Federico G. Mantovanelli (2018). 'Long-Term Effects of Access to Health Care: Medical Missions in Colonial India'. *Journal of Development Economics* 135.4, pp. 285–303. DOI: [10.1016/j.jdeveco.2018.07.009](https://doi.org/10.1016/j.jdeveco.2018.07.009).
- Chaudhary, Latika and Manuj Garg (2015). 'Does History Matter? Colonial Education Investments in India'. *The Economic History Review* 68.3, pp. 937–961. DOI: [10.1111/ehr.12078](https://doi.org/10.1111/ehr.12078).
- Cogneau, Denis and Alexander Moradi (2014). 'Borders That Divide: Education and Religion in Ghana and Togo Since Colonial Times'. *The Journal of Economic History* 74.3, pp. 694–729. DOI: [10.1017/S0022050714000576](https://doi.org/10.1017/S0022050714000576).
- Conley, Timothy G. (1999). 'GMM Estimation with Cross Sectional Dependence'. *Journal of Econometrics* 92.1, pp. 1–45. DOI: [10.1016/S0304-4076\(98\)00084-0](https://doi.org/10.1016/S0304-4076(98)00084-0).
- Damis, John (1974). 'The Free-School Phenomenon: The Cases of Tunisia and Algeria'. *International Journal of Middle East Studies* 5.4, pp. 434–449. DOI: [10.1017/S0020743800025526](https://doi.org/10.1017/S0020743800025526).



- Davi, Laura (2000). 'Entre colonisateurs et colonisés : les italiens de Tunisie (XIX<sup>e</sup>–XX<sup>e</sup> siècle)'. In: *La Tunisie mosaïque*. Ed. by Patrick Cabanel and Jacques Alexandropoulos. Toulouse: Presses universitaires du Midi, pp. 99–113. DOI: [10.4000/books.pumi.4859](https://doi.org/10.4000/books.pumi.4859).
- Degorge, Barbara (2002). 'The Modernization of Education: A Case Study of Tunisia and Morocco'. *The European Legacy* 7.5, pp. 579–596. DOI: [10.1080/1084877022000006780](https://doi.org/10.1080/1084877022000006780).
- Delleaux, Fulgence (2010). 'The Persistent Effects of Peru's Mining Mita'. *Econometrica* 78.6, pp. 1863–1903. DOI: [10.3982/ECTA8121](https://doi.org/10.3982/ECTA8121).
- Dupraz, Yannick (2019). 'French and British Colonial Legacies in Education: Evidence from the Partition of Cameroon'. *The Journal of Economic History* 79.3, pp. 628–688. DOI: [10.1017/S0022050719000299](https://doi.org/10.1017/S0022050719000299).
- Eibl, Ferdinand (2020). *Social Dictatorships: The Political Economy of the Welfare State in the Middle East and North Africa*. Oxford: Oxford University Press, p. 380. DOI: [10.1093/oso/9780198834274.001.0001](https://doi.org/10.1093/oso/9780198834274.001.0001).
- Foncin, Pierre (1882). 'L'enseignement en Tunisie'. *Revue internationale de l'enseignement* 4, pp. 401–418.
- Harris, Ian, Timothy J. Osborn, Phil Jones and David Lister (2020). 'Version 4 of the CRU TS Monthly High-Resolution Gridded Multivariate Climate Dataset'. *Scientific Data* 7.109, pp. 1–18. DOI: [10.1038/s41597-020-0453-3](https://doi.org/10.1038/s41597-020-0453-3).
- Huillery, Élise (2009). 'History Matters: The Long-Term Impact of Colonial Public Investments in French West Africa'. *American Economic Journal: Applied Economics* 1.2, pp. 176–215. DOI: [10.1257/app.1.2.176](https://doi.org/10.1257/app.1.2.176).
- Jedwab, Remi, Felix Meier zu Selhausen and Alexander Moradi (2022). 'The Economics of Missionary Expansion: Evidence from Africa and Implications for Development'. *Journal of Economic Growth*. DOI: [10.1007/s10887-022-09202-8](https://doi.org/10.1007/s10887-022-09202-8).
- Kelly, Morgan (2019). *The Standard Errors of Persistence*. CEPR Discussion Paper No DP13783. Centre for Economic Policy Research.
- Kelly, Morgan (2020). *Understanding Persistence*. CEPR Discussion Paper No DP15246. Centre for Economic Policy Research.
- La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny (1999). 'The Quality of Government'. *The Journal of Law, Economics, and Organization* 15.1, pp. 222–279. DOI: [10.1093/jleo/15.1.222](https://doi.org/10.1093/jleo/15.1.222).
- Messika, Martin, Mônica Raisa Schpun and Colette Zytynicki (2015). 'Les Juifs de Tunisie'. *Hommes & migrations* 1312, pp. 143–145. DOI: [10.4000/hommesmigrations.3529](https://doi.org/10.4000/hommesmigrations.3529).
- Michalopoulos, Stelios and Elias Papaioannou (2017). *The Long Economic and Political Shadow of History. Volume I. A Global View*. London: CEPR Press, p. 124.
- Montalbano, Gabriele (2020). 'Le scuole italiane in Tunisia: tra rivalità imperiali e costruzione comunitaria (1861–1910)'. *Altreitalie* 61, pp. 53–70.
- Nunn, Nathan (2009). 'The Importance of History for Economic Development'. *Annual Review of Economics* 1, pp. 65–92. DOI: [10.1146/annurev.economics.050708.143336](https://doi.org/10.1146/annurev.economics.050708.143336).
- Panel, Anne-Marie (2015). *Du comptoir à la colonie. Histoire de la communauté française de Tunisie, 1814–1883*. Paris: Riveneuve éditions, p. 521.

- Rocha, Rudi, Claudio Ferraz and Rodrigo R. Soares (2017). 'Human Capital Persistence and Development'. *American Economic Journal: Applied Economics* 9.4, pp. 105–136. DOI: [10.1257/app.20150532](https://doi.org/10.1257/app.20150532).
- Sequeira, Sandra, Nathan Nunn and Nancy Qian (2020). 'Immigrants and the Making of America'. *The Review of Economic Studies* 87.1, pp. 382–419. DOI: [10.1093/restud/rdz003](https://doi.org/10.1093/restud/rdz003).
- Sraieb, Noureddine (1993). 'L'idéologie de l'école en Tunisie coloniale (1881–1945)'. *Revue des mondes musulmans et de la Méditerranée* 68.1, pp. 239–254. DOI: [10.3406/remmm.1993.2570](https://doi.org/10.3406/remmm.1993.2570).
- Tarifa, Chadli (1971). 'L'enseignement du premier et du second degré en Tunisie'. *Population* 26.1, pp. 149–180.
- Waldinger, Maria (2017). 'The Long-Run Effects of Missionary Orders in Mexico'. *Journal of Development Economics* 127, pp. 355–378. DOI: [10.1016/j.jdeveco.2016.12.010](https://doi.org/10.1016/j.jdeveco.2016.12.010).
- Wantchekon, Leonard, Klašnja Marko and Natalija Novta (2014). 'Education and Human Capital Externalities: Evidence from Colonial Benin'. *The Quarterly Journal of Economics* 130.2, pp. 703–757. DOI: [10.1093/qje/qjv004](https://doi.org/10.1093/qje/qjv004).
- Wietzke, Frank-Borge (2015). 'Long-Term Consequences of Colonial Institutions and Human Capital Investments: Sub-National Evidence from Madagascar'. *World Development* 66.3, pp. 293–307. DOI: [10.1016/j.worlddev.2014.08.010](https://doi.org/10.1016/j.worlddev.2014.08.010).
- Wolkowitsch, Maurice (1959). 'L'émigration des français de Tunisie'. *Annales de Géographie* 68.367, pp. 253–257.

## Figures

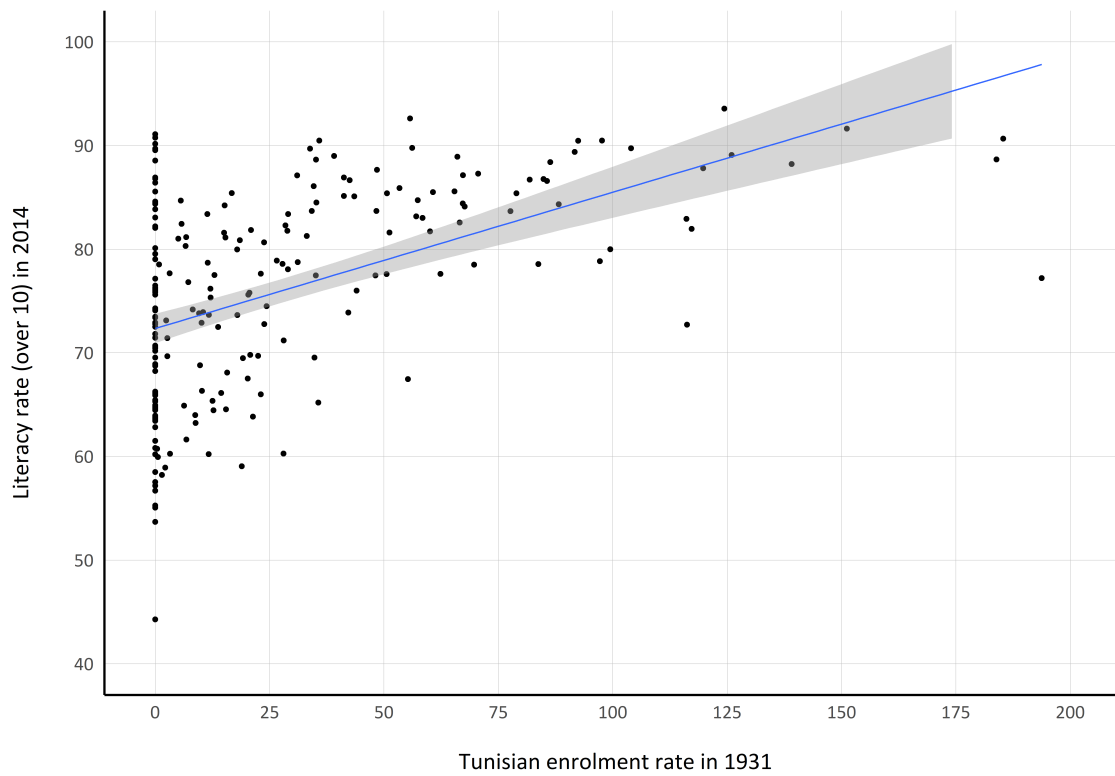


Figure 1: Unconditional relationship between primary school enrolment rate in 1931 and literacy rate in 2014

## Tables

Table 1: Summary Statistics

Variables	Mean	SD	Min	Max
<b>Historical variables</b>				
Log Tunisian enrolment rate 1931	2.12	1.83	0	5.27
Log teachers per 1,000 Tunisian inhabitants 1931	0.29	0.39	0	1.91
Log European population 1931	3.72	2.81	0	11.38
<b>Contemporary variables</b>				
Literacy rate 2014	76.00	9.91	44.29	93.55
Literacy rate 2019*	0.82	0.39	0	1
Primary education 2019*	0.60	0.49	0	1
<b>Control variables</b>				
Number of mines 1905 (10 km radius)	0.47	1.01	0	6
Number of railway stations 1926 (10 km radius)	0.92	1.21	0	5
Soil suitability for olive	2,243	1,835	0	5,379
Soil suitability for wheat	2,703	1,689	0	6,397
Slope index	8,589	1,084	4,886	9,928
Distance to coast/border (km)	54.52	52.57	0.53	219.27
Average rainfall 1901–1910 (mm)	283.74	112.85	65.65	529.96
<b>Other variables</b>				
Log Tunisian population 1931	7.82	3.04	0	11.59
Log schools per 1,000 Tunisian inhabitants 1931	0.17	0.21	0	1.24
Log schools per 1,000 Tunisian inhabitants 2014	1.83	0.46	0.76	2.94
Log teachers per 1,000 Tunisian inhabitants 2014	1.99	0.19	1.42	2.53
Literacy rate 2004	72.65	9.39	42.30	91.10
Literacy rate 2014 (15–29 years old)	93.96	5.49	55.65	99.38
Baccalaureate success rate 2010	67.28	10.80	37.06	100

Notes: \*From an individual survey with 521,706 individuals. All the other variables have 217 observations.

Table 2: Effect of exposure to colonial primary education on literacy in 2014

	(1)	(2)	(3)	(4)	(5)
	Literacy rate 2014				
Log Tunisian enrolment rate 1931	3.1616 (0.4627) [0.0000]	3.2170 (0.5493) [0.0000]	2.3512 (0.6638) [0.0018]	1.2477 (0.5645) [0.0378]	1.6852 (0.3936) [0.0003]
Number of districts	211	211	211	211	211
Dependent Variable Mean	75.7	75.7	75.7	75.7	75.7
European population 1931	No	Yes	Yes	Yes	Yes
Teachers ratio 1931	No	No	Yes	Yes	Yes
Geographic and economic controls	No	No	No	Yes	Yes
Region FE	No	No	No	No	Yes
R <sup>2</sup>	0.326	0.327	0.344	0.627	0.702

*Notes:* OLS estimations. Robust standard errors clustered at the governorate level and p-values in parentheses and brackets respectively. There are 6 regions and 23 governorates. Tunis is excluded. All variables are defined at the district level. The dependent variable is the literacy rate of the population aged over 10 in 2014. The independent variable is the log of the enrolment rate in primary schools of the Tunisian population aged 0–18 in 1931. *European population 1931* is the log of the European population in 1931. *Teachers ratio 1931* is the log of the number of primary school teachers per 1,000 Tunisian inhabitants in 1931. *Controls* include a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, average rainfall per year between 1901 and 1911, the number of mines in 1905, the number of railway stations in 1926, and a dummy for precolonial major urban centres.

Table 3: Effect on literacy: Placebo test

	(1)	(2)	(3)
	Literacy rate 2014		
Log Tunisian enrolment rate 1931	1.6852 (0.3936) [0.0003]		
Log Italian enrolment rate 1931		0.1107 (0.2393) [0.6483]	
Log French enrolment rate 1931			0.0275 (0.2535) [0.9145]
Number of districts	211	211	211
Dependent Variable Mean	75.7	75.7	75.7
European population 1931	Yes	Yes	Yes
Teachers ratio 1931	Yes	Yes	Yes
Geographic and economic controls	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
R <sup>2</sup>	0.702	0.676	0.675

*Notes:* OLS estimations. Robust standard errors clustered at the governorate level and p-values in parentheses and brackets respectively. There are 6 regions and 23 governorates. Tunis is excluded. All variables are defined at the district level. The dependent variable is the literacy rate of the population aged over 10 in 2014. In columns 1 to 3, the independent variable is the log of the enrolment rate in primary schools of the Tunisian, Italian, and French population aged 0–18 in 1931, respectively. *European population 1931* is the log of the European population in 1931. *Teachers ratio 1931* is the log of the number of primary school teachers per 1,000 Tunisian inhabitants in 1931. *Controls* include a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, average rainfall per year between 1901 and 1911, the number of mines in 1905, the number of railway stations in 1926, and a dummy for precolonial major urban centres.

Table 4: Test for the persistence of investment in education

	(1) Log schools per 1,000 inhabitants 2014	(2) Log teachers per 1,000 inhabitants 2014
Log schools per 1,000 inhabitants 1931	-0.0833 (0.0885) [0.3571]	
Log teachers per 1,000 inhabitants 1931		-0.0785 (0.0410) [0.0687]
Number of districts	211	211
Dependent Variable Mean	0.5	2.0
European population 1931	Yes	Yes
Geographic and economic controls	Yes	Yes
Region FE	Yes	Yes
R <sup>2</sup>	0.527	0.572

Notes: OLS estimations. Robust standard errors clustered at the governorate level and p-values in parentheses and brackets respectively. There are 6 regions and 23 governorates. Tunis is excluded. All variables are defined at the district level. The dependent variable is the log of the number of primary schools (teachers) per 1,000 inhabitants in 2014. The independent variable is the log of the number of primary schools (teachers) per 1,000 Tunisian inhabitants in 1931. *European population 1931* is the log of the European population in 1931. *Controls* include a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, average rainfall per year between 1901 and 1911, the number of mines in 1905, the number of railway stations in 1926, and a dummy for precolonial major urban centres.

Table 5: Effect on literacy: Alternative outcomes

	(1) Literacy rate 2014	(2) Literacy rate 2004	(3) Literacy rate age-group 15–29
Log Tunisian enrolment rate 1931	1.6852 (0.3936) [0.0003]	1.3607 (0.4810) [0.0098]	0.6019 (0.1970) [0.0058]
Number of districts	211	211	211
Dependent Variable Mean	75.7	75.7	75.7
European population 1931	Yes	Yes	Yes
Teachers ratio 1931	Yes	Yes	Yes
Geographic and economic controls	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
R <sup>2</sup>	0.702	0.654	0.537

Notes: OLS estimations. Robust standard errors clustered at the governorate level and p-values in parentheses and brackets respectively. There are 6 regions and 23 governorates. Tunis is excluded. All variables are defined at the district level. In columns 1 to 4, the dependent variable is respectively the literacy rate of the population aged over 10 in 2014, the literacy rate of the population aged over 10 in 2004, and the literacy rate of the population aged 15–29 in 2014. The independent variable is the log of the enrolment rate in primary schools of the Tunisian population aged 0–18 in 1931. *European population 1931* is the log of the European population in 1931. *Teachers ratio 1931* is the log of the number of primary school teachers per 1,000 Tunisian inhabitants in 1931. *Controls* include a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, average rainfall per year between 1901 and 1911, the number of mines in 1905, the number of railway stations in 1926, and a dummy for precolonial major urban centres.

Table 6: Effect on literacy: Alternative specifications

	(1)	(2)	(3)	(4)
	Literacy rate 2014			
Log Tunisian enrolment rate 1931	1.6852 (0.3936) [0.0003]	1.8036 (0.4882) [0.0013]	1.5851 (0.3682) [0.0003]	0.8095 (0.3406) [0.0262]
Merge	Yes	Yes	Yes	Yes
Tunis	No	No	Yes	Yes
5 km	No	Yes	No	No
Number of districts	211	211	217	264
Dependent Variable Mean	75.7	75.7	76.0	78.1
European population 1931	Yes	Yes	Yes	Yes
Teachers ratio 1931	Yes	Yes	Yes	Yes
Geographic and economic controls	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.702	0.672	0.698	0.645

*Notes:* OLS estimations. Robust standard errors clustered at the governorate level and p-values in parentheses and brackets respectively. There are 6 regions and 24 governorates. Tunis is excluded. All variables are defined at the district level. The dependent variable is the literacy rate of the population aged over 10 in 2014. Column 1 reproduces the baseline estimate. In column 2, we redistribute pupils between districts within 5 km around borders. Column 3 includes the governorate of Tunis (6 districts). Column 4 uses the full sample of 264 districts. The independent variable is the log of the enrolment rate in primary schools of the Tunisian population aged 0–18 in 1931. *European population 1931* is the log of the European population in 1931. *Teachers ratio 1931* is the log of the number of primary school teachers per 1,000 Tunisian inhabitants in 1931. *Controls* include a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, average rainfall per year between 1901 and 1911, the number of mines in 1905, the number of railway stations in 1926, and a dummy for precolonial major urban centres.

Table 7: Effect on literacy: 2019 Population and Labour Force Survey

	(1)	(2)	(3)	(4)	(5)
	Literacy rate 2019				
Log Tunisian enrolment rate 1931	0.0250 (0.00415) [0.0000]	0.0232 (0.00553) [0.0000]	0.0141 (0.00719) [0.049]	0.0839 (0.00604) [0.165]	0.00922 (0.00438) [0.035]
Number of individuals	495,039	495,039	495,039	495,039	495,039
European population 1931	No	Yes	Yes	Yes	Yes
Teachers ratio 1931	No	No	Yes	Yes	Yes
Geographic and economic controls	No	No	No	Yes	Yes
Region FE	No	No	No	No	Yes

*Notes:* Logit regression. Marginal effect is reported with robust standard errors clustered at the governorate level and p-values in parentheses and brackets respectively. There are 6 regions and 23 governorates. Tunis is excluded. The dependent variable is the literacy rate in 2019 as reported by individuals during the survey. It is a dummy variable equal to one if the respondent in district  $i$  in governorate  $j$  reported being literate, and zero otherwise. We restrict our sample to the respondents older than 10 years-old. The independent variable is the log of the enrolment rate in primary schools of the Tunisian population aged 0–18 in 1931 in district  $i$  in governorate  $j$ . *European population 1931* is the log of the European population in 1931. *Teachers ratio 1931* is the log of the number of primary school teachers per 1,000 Tunisian inhabitants in 1931. *Controls* include a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, average rainfall per year between 1901 and 1911, the number of mines in 1905, and the number of railway stations in 1926. We also control for age and gender in all regressions.

Table 8: Effect on literacy by age-group: 2019 Population and Labour Force Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Literacy rate 2019						
Birth decade	All	1950s	1960s	1970s	1980s	1990s	2000s
Log Tunisian enrolment rate 1931	0.00922 (0.00438) [0.035]	0.0217 (0.00995) [0.029]	0.0206 (0.00833) [0.013]	0.0137 (0.00654) [0.036]	0.00783 (0.00383) [0.041]	0.000998 (0.000152) [0.512]	0.000415 (0.0000841) [0.621]
Number of individuals	495,039	51,630	66,847	79,611	86,271	80,425	90,112
European population 1931	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teachers ratio 1931	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic and economic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	(8)	(9)	(10)	(11)			
Birth year	1980–4	1985–9	1990–4	1995–9			
Log Tunisian enrolment rate 1931	0.0090972 (0.0044405) [0.040]	0.0065395 (0.0110) [0.056]	0.0020762 (0.0034183) [0.351]	-0.0000966 (0.0012549) [0.036]			
Number of individuals	43,357	42,914	41,040	39,385			
European population 1931	Yes	Yes	Yes	Yes			
Teachers ratio 1931	Yes	Yes	Yes	Yes			
Geographic and economic controls	Yes	Yes	Yes	Yes			
Region FE	Yes	Yes	Yes	Yes			

*Notes:* Logit regression. Marginal effect is reported with robust standard errors clustered at the governorate level and p-values in parentheses and brackets respectively. There are 6 regions and 23 governorates. Tunis is excluded. The dependent variable is the literacy rate in 2019 as reported by individuals during the survey. It is a dummy variable equal to one if the respondent in district  $i$  in governorate  $j$  reported being literate, and zero otherwise. We restrict our sample to the respondents older than 10 years-old. The column labels *birth decade* correspond to specific age-groups, e.g. 1940s stands for age-group 70–9. The independent variable is the log of the enrolment rate in primary schools of the Tunisian population aged 0–18 in 1931 in district  $i$  in governorate  $j$ . *European population 1931* is the log of the European population in 1931. *Teachers ratio 1931* is the log of the number of primary school teachers per 1,000 Tunisian inhabitants in 1931. *Controls* include a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, average rainfall per year between 1901 and 1911, the number of mines in 1905, and the number of railway stations in 1926. We also control for age and gender in all regressions.



Table 9: Effect on education level by age-group: 2019 Population and Labour Force Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth decade	All	1940s	1950s	Finished primary school				
				1960s	1970s	1980s	1990s	2000s
Log Tunisian enrolment rate 1931	0.0106262 (0.004373) [0.015]	0.0169372 (0.0089244) [0.058]	0.0205378 (0.0094753) [0.030]	0.0202643 (0.0080858) [0.012]	0.0131825 (0.0065593) [0.044]	0.0079457 (0.0043101) [0.065]	0.0052079 (0.0043293) [0.229]	0.0018859 (0.0035956) [0.600]
Number of individuals	495,039	25,395	51,630	66,847	79,611	86,271	80,425	90,112
European population 1931	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teachers ratio 1931	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic and economic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes:* Logit regression. Marginal effect is reported with robust standard errors clustered at the governorate level and p-values in parentheses and brackets respectively. There are 6 regions and 23 governorates. Tunis is excluded. The dependent variable is a dummy variable equal to one if the respondent in district  $i$  in governorate  $j$  reports finishing at least primary school during the survey, and zero otherwise. We restrict our sample to the respondents older than 10 years-old. The column labels *birth decade* correspond to specific age-groups, e.g. 1940s stands for age-group 70–9. The independent variable is the log of the enrolment rate in primary schools of the Tunisian population aged 0–18 in 1931 in district  $i$  in governorate  $j$ . *European population 1931* is the log of the European population in 1931. *Teachers ratio 1931* is the log of the number of primary school teachers per 1,000 Tunisian inhabitants in 1931. *Controls* include a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, average rainfall per year between 1901 and 1911, the number of mines in 1905, and the number of railway stations in 1926. We also control for age and gender in all regressions.

## **Online appendix**

### The Colonial Legacy of Education: Evidence from Tunisia

MHAMED BEN SALAH

*Graduate Institute of International and Development Studies*

CÉDRIC CHAMBRU

*University of Zurich*

MALEKE FOURATI

*South Mediterranean University*

## Contents

<b>A Additional information on data</b>	<b>3</b>
<b>B Additional Results</b>	<b>6</b>
B.1 Spatial autocorrelation . . . . .	6
B.2 Beta Regression . . . . .	8
<b>Online Appendix References</b>	<b>9</b>

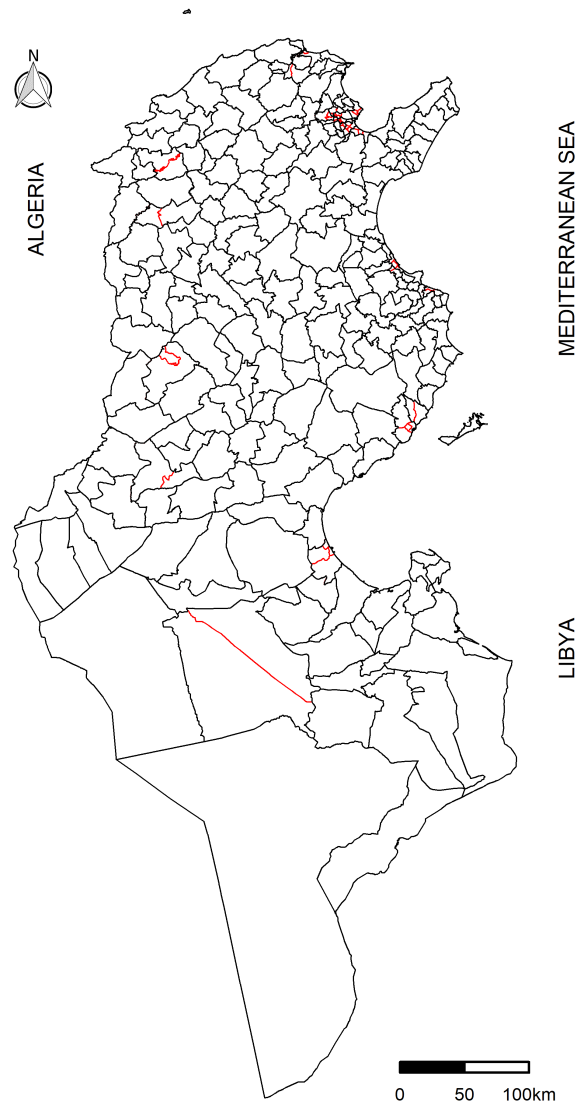
## List of Figures

A.1 The borders of districts in Tunisia in 2014 . . . . .	3
A.2 Population, schools, and enrolment rate in 1931, and literacy rate in 2014 . . . . .	4
A.3 Moran’s I scatterplots of residuals . . . . .	7

## List of Tables

A.1 Population in Tunisia during the French protectorate, 1881–1956 . . . . .	5
A.2 Colonial investment in education in Tunisia, 1891–1931 . . . . .	5
A.3 Post-colonial education indicators, 1956–2014 . . . . .	5
A.4 Share of population and education data without coordinates in 1931 . . . . .	5
A.5 Correction of standard errors using a Bartlett spatial weights correction . . . . .	7
A.6 Beta regressions . . . . .	8

## A Additional information on data



Notes: The black lines represent the borders of the 211 districts (*délégations*) in our sample. The red lines represent the initial borders of districts we merge to get our sample.

Figure A.1: The borders of districts in Tunisia in 2014

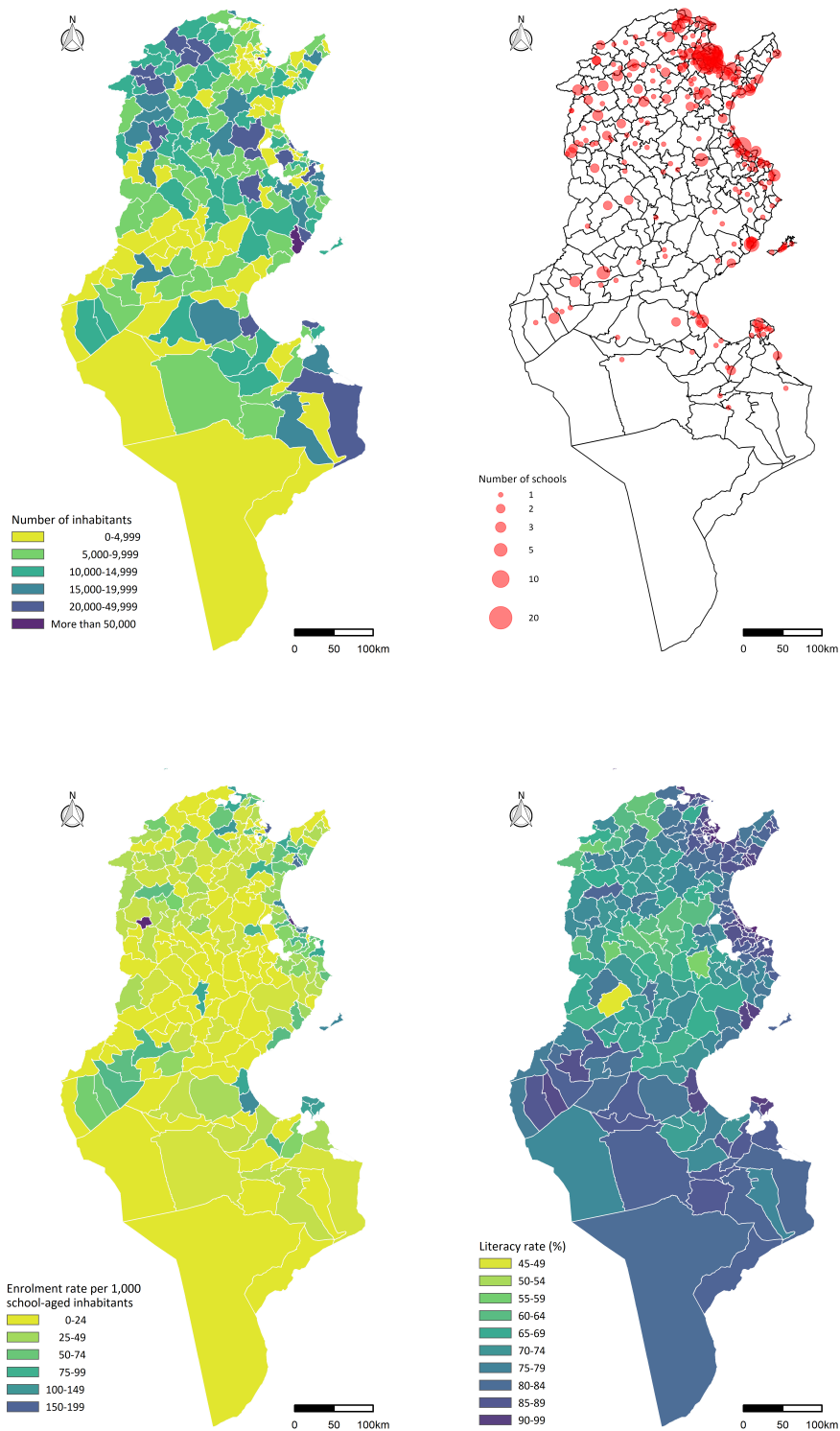


Figure A.2: Population, schools, and enrolment rate in 1931, and literacy rate in 2014

Table A.1: Population in Tunisia during the French protectorate, 1881–1956

	Europeans				Total	Tunisians		
	French	Italians	Maltese	Others		Muslims	Israelites	Total
1881	300*	11 206**	-	-	-	-	-	-
1886	-	-	-	-	-	-	-	876 861*
1889	-	-	-	-	-	-	-	907 642*
1891	9 973	21 016**	11 706***	-	-	-	-	-
1896	16 207	-	-	-	-	-	-	-
1901	24 201	-	-	-	-	-	-	-
1906	34 610	81 156	10 330	2 799	128 895	-	-	-
1911	46 044	88 082	11 300	3 050	148 476	1 740 144*	50 467*	1 790 611*
1921	54 476	84 799	13 520	3 320	156 115	1 826 545	48 436	1 874 981
1926	71 000	89 216	-	13 065	173 281	1 864 908	54 243	1 919 151
1931	91 427	91 178	8 643	4 145	195 393	2 086 762	56 248	2 143 010
1936	108 068	94 289	7 279	3 569	213 205	2 265 750	59 585	2 325 335
1946	143 977	84 935	6 459	4 178	239 549	2 832 978	71 543	2 904 521
1956	180 450	66 909	-	-	255 332	3 383 213	57 786	3 440 999

Notes: \*Do not come from a census. Estimates made by the French administration, which can underestimate the actual numbers. The number of French in 1881 before the colonisation comes from an estimation found in the *Statistique générale de la Tunisie* from of 1881–92. The 1889 statistics were compiled by the Secretary General of the Tunisian Government. The 1896 statistics were compiled by civilian controllers and millet commanders. The statistics of 1911 comes from the *Statistique générale de la Tunisie* of 1913. \*\*The number of Italians before 1906 is given by the Italian government. \*\*\*These numbers correspond to the number of British and Maltese as drawn from a census made by the British General Consulate.

Table A.2: Colonial investment in education in Tunisia, 1891–1931

	1891	1901	1911	1921	1931
Number of schools	76	132	244	336	468
Number of 2014 districts with a school	43	66	107	136	163
Total number of pupils	8,208	14,272	24,497	36,639	67,089
Number of Tunisian pupils	2,651	3,226	7,487	13,026	34,556
Share of Tunisian pupils	32%	21%	31%	36%	52%
Number of teachers	-	434	706	1,174	1,833
Number of Tunisian teachers	-	37	80	141	272

Sources: see Section 3.3.

Table A.3: Post-colonial education indicators, 1956–2014

	1956	1966	1975	1984	1994	2004	2014
Literacy rate	15.3	32.1	45.1	53.8	68.3	77.1	81.2
Enrolment rate (6–14)	-	-	59.9	75.9	86.2	95.1	95.8

Sources: Table 1 (p. 9) and Table 19 (p. 25) from the 2014 Tunisian population census (*Recensement Général de la Population et de l'Habitat 2014*).

Table A.4: Share of population and education data without coordinates in 1931

	Population	Missing	% missing
French	91,500	1,898	2.07
Italian	91,192	1,645	1.80
Tunisian	2,142,012	33,928	1.58
French pupils	7,375	103	1.40
Italian pupils	4,437	25	0.56
Tunisian pupils	31,158	259	0.83

## B Additional Results

### B.1 Spatial autocorrelation

This paper relies on various spatial regressions to study the effect of the exposure to colonial education in a given district on the level of literacy, in the same district, about 80 years later. The first law of geography states that ‘everything is related to everything else, but near things are more related than distant things’ (Tobler 1970), which can translate in the fact that spatial data tend to show an important autocorrelation, based on the distance between points. Kelly (2019, 2020) argues that it might generate both a spatial trend leading to spurious correlation and an underestimation of standard errors causing inflated  $t$  statistics. However, there is no consensus among researchers on the importance of spatial autocorrelation in the persistence literature in particular (Voth 2021). There is, therefore, no standard procedure to address this potential issue.

We are confident about the fact that our geographic control variables (latitude, distance to the coast), along with the region fixed effects, are reducing substantially the impact of the spatial trend. Our results are also robust to including longitude, squared latitude or longitude and an interaction of latitude and longitude to control for non-linearities. On the other hand, our residuals are likely to be spatially correlated. In our baseline specification, we try to control for the spatial correlation by clustering our standard errors at the governorate level, which is one administrative level above our observations. However, for clustered standard errors to be consistent it requires residuals to be uncorrelated between clusters, a condition that is not necessarily satisfied when we think, for instance, at neighbouring districts on the opposite side of governorate borders.

Following Kelly (2019), we begin by testing for the spatial autocorrelation of the residuals for our main regression (column 5 in Table 2), by computing their Moran’s I statistic. We find a positive and significant I statistic with a  $z$ -score of 2.9, suggesting that our residuals are spatially correlated (online Appendix Figure A.3). However, according to Kelly, our  $z$ -score is relatively low and should not cause a significant underestimation of our standard errors.<sup>1</sup> We are therefore confident that despite a presence of spatial autocorrelation, neither a spatial trend nor an underestimation of standard errors are driving our coefficient nor their significance.

We further use a Conley (1999) adjustment of our standard errors in order to account for spatial correlation. Classic Conley adjustment is very sensitive to the choice of distance cut-offs, we use a modified Bartlett matrix to allow for a linear decay of spatial weights with distance. We also use three different distance cut-offs (25 km, 50 km, and 75 km). Results reported in online Appendix Table A.5 indicate that our baseline estimate remains significant even after adjusting our standard errors for spatial correlation with different cut-offs. This finding is robust to using different distance cut-offs and a classic Conley adjustment.

<sup>1</sup> The probability of spatial noise outperforming the original independent variable is low for regressions with residuals getting a  $z$ -score of their Moran’s I statistic lower than 4.

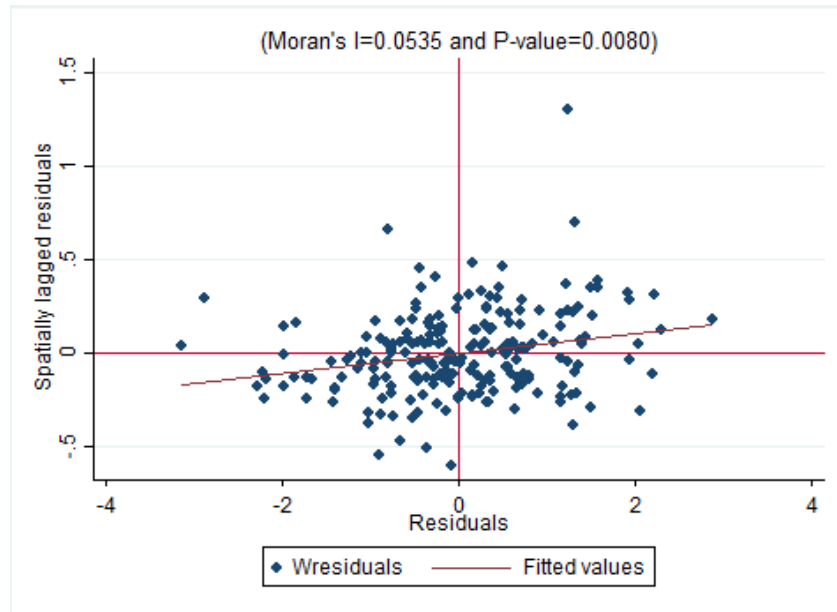


Figure A.3: Moran's I scatterplots of residuals

Table A.5: Correction of standard errors using a Bartlett spatial weights correction

	(1)	(2)	(3)	(4)
	Literacy rate 2014			
Log Tunisian enrolment rate 1931	1.5611 (0.3824) [0.0005]	1.5611 (0.3948) [0.0001]	1.5611 (0.3840) [0.0000]	1.5611 (0.3617) [0.0000]
Distance cutoff (km)	(baseline)	25	50	75
Number of districts	211	211	211	211
Dependent Variable Mean	76.0	76.0	76.0	76.0
European population 1931	Yes	Yes	Yes	Yes
Teachers ratio 1931	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.699	0.699	0.699	0.699

*Notes:* OLS estimations. In column 1, robust standard errors clustered at the governorate level. In columns 2 to 4, standard errors are estimated using a spatial correction at three different thresholds: 25 km, 50 km and 75 km. We use the modified Bartlett spatial weights that decay linearly with distance. P-values in brackets. There are 6 regions and 23 governorates. Tunisia is excluded. All variables are defined at the district level. The dependent variable is the literacy rate of the population aged over 10 in 2014. The independent variable is the log of the enrolment rate in primary schools of the Tunisian population aged 0–18 in 1931. *European population 1931* is the log of the European population in 1931. *Teachers ratio 1931* is the log of the number of primary school teachers per 1,000 Tunisian inhabitants in 1931. *Controls* include a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, average rainfall per year between 1901 and 1911, the number of mines in 1905, and the number of railway stations in 1926.



## B.2 Beta Regression

Since our main dependent variable, literacy rate, is a fraction, linear regressions can lead to misspecification and uncertain statistical validity. Indeed, the fitted values for the model could fall outside the unit interval (0, 1). In order to correct for this, and since our dependent variable never takes the exact value of 0 or 1, we use a Beta regression to estimate our model 1 (Ferrari and Cribari-Neto 2004; Smithson and Verkuilen 2006). In other words, we compute a maximum likelihood estimator, based on a Beta distribution using a logit as link function. The logit function will take us from our bounded space to real numbers, which allows us then to perform a regression, assuming our data is Beta distributed, by maximising its likelihood. This will allow us to better specify our estimators so they do not lead to predictions with negative or higher than 100% literacy rates. The result is reported in online Appendix Table A.6. As we can see, our coefficient is highly significant, with a similar magnitude as in our main regression. Indeed, in order to perform a Beta regression, we divide our literacy rate by 100 so it takes a value between 0 and 1. Therefore, the coefficient of 0.0125 corresponds to an increase of 1.25 percentage points in 2014 literacy rate for a one per cent increase in 1931 Tunisian enrolment.

Table A.6: Beta regressions

	(1)	(2)	(3)	(4)	(5)
	Literacy rate 2014				
Literacy rate in 2014					
Log Tunisian enrolment rate 1931	0.0296 (0.0032) [0.0000]	0.0297 (0.0040) [0.0000]	0.0194 (0.0049) [0.0001]	0.0103 (0.0043) [0.0178]	0.0148 (0.0043) [0.0007]
Number of districts	211	211	211	211	211
European population 1931	No	Yes	Yes	Yes	Yes
Teachers ratio 1931	No	No	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes
Region FE	No	No	No	No	Yes

*Notes:* Beta regression using a logit as link function. Marginal effect is reported with robust standard errors clustered at the governorate level and p-values in parentheses and brackets respectively. There are 6 regions and 23 governorates. Tunis is excluded. All variables are defined at the district level. The dependent variable is the literacy rate of the population aged over 10 in 2014. To be restricted to unit interval (0, 1), we divide it by 100. The independent variable is the log of the enrolment rate in primary schools of the Tunisian population aged 0–18 in 1931. *European population 1931* is the log of the European population in 1931. *Teachers ratio 1931* is the log of the number of primary school teachers per 1,000 Tunisian inhabitants in 1931. *Controls* include a measure of terrain slope, latitude, distance from the coast, soil suitability for wheat and olive, average rainfall per year between 1901 and 1911, the number of mines in 1905, and the number of railway stations in 1926.

## Online Appendix References

- Conley, Timothy G. (1999). 'GMM Estimation with Cross Sectional Dependence'. *Journal of Econometrics* 92.1, pp. 1–45. DOI: [10.1016/S0304-4076\(98\)00084-0](https://doi.org/10.1016/S0304-4076(98)00084-0).
- Ferrari, Silvia and Francisco Cribari-Neto (2004). 'Beta Regression for Modelling Rates and Proportions'. *Journal of Applied Statistics* 31.7, pp. 799–815. DOI: [10.1080/0266476042000214501](https://doi.org/10.1080/0266476042000214501).
- Kelly, Morgan (2019). *The Standard Errors of Persistence*. CEPR Discussion Paper No DP13783. Centre for Economic Policy Research.
- Kelly, Morgan (2020). *Understanding Persistence*. CEPR Discussion Paper No DP15246. Centre for Economic Policy Research.
- Smithson, Michael and Jay Verkuilen (2006). 'A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables'. *Psychological Methods* 11.1, pp. 54–71. DOI: [10.1037/1082-989X.11.1.54](https://doi.org/10.1037/1082-989X.11.1.54).
- Tobler, Waldo R. (1970). 'A Computer Movie Simulating Urban Growth in the Detroit Region'. *Economic Geography* 46.supl. Pp. 234–240. DOI: [10.2307/143141](https://doi.org/10.2307/143141).
- Voth, Hans-Joachim (2021). 'Persistence-Myth and Mystery'. In: *The Handbook of Historical Economics*. Ed. by Alberto Bisin and Giovanni Federico. London: Academic Press, pp. 243–267. DOI: [10.1016/C2017-0-04055-7](https://doi.org/10.1016/C2017-0-04055-7).