# Literacy and primary school expansion in Portugal: 1940–62

Pedro Gomes[*] Matilde Pinto Machado[†]

Birkbeck, University of London Universidad Carlos III de Madrid

[*] p.gomes@bbk.ac.uk
[†] matilde.machado@uc3m.es

# Literacy and primary school expansion in Portugal: 1940–62


## Abstract

In 1940, the Portuguese government approved a massive primary school construction plan that projected a 60% increase in the number of primary schools. Based on the collection of a new dataset, we describe literacy levels in Portugal prior to the plan as well as the plan's strategy regarding the location of schools. We then estimate the causal impact of the increase in the number of schools between 1940 and the early 60s on enrolment and literacy, all at the county level. We conclude the increase in the number of schools was responsible for 80% of the increase in enrolment and 13% of the increase in the literacy rate of the affected cohorts.

En 1940, el Gobierno Portugués aprobó un ambicioso plan de construcción de infraestructuras con el objetivo de aumentar el número de escuelas primarias en 60%. Basándonos en una nueva base de datos, empezamos por describir la situación de la escolarización en Portugal anterior al plan, así como la estrategia de localización de las nuevas escuelas proyectadas por el plan. A continuación, estimamos el impacto que tuvo el aumento del número de escuelas primarias en cada provincia entre 1940 y el inicio de los años 60 en la matriculación y alfabetización de los niños. Concluimos que el incremento en el número de escuelas primarias fue responsable del 80% en el aumento de la matriculación y del 13% en el aumento de la alfabetización.



*Keywords:* school construction, investment in education, historical expansion of education, literacy growth, schooling expansion, primary school.

*JEL:* I28, I25, I26, N34, N64

## 1. Introduction

In December 1940, the Portuguese government approved a national plan for the construction of primary schools called the *Centenários* plan (*Plano dos Centenários* in Portuguese). The original plan aimed at building 6,082 schools across the mainland—corresponding to 9,314 classrooms—over a 10-year period. The endeavour would mean approximately one new school for every 90 children aged 7–10 years and a 60% increase in the number of primary schools, providing a remarkable example of schooling expansion. The plan was updated in 1955 and 1961, extending the construction of new primary schools until the late 1960s.

The objective of this paper is threefold. First, we describe the state of literacy in Portugal at the time of the beginning of the *Centenários* plan at the county level. Second, we estimate the factors behind the geographic distribution of primary schools in the plan. Third, we use geographical variation in the number of schools built (and planned) to estimate causal effects of the increase in the supply of primary schools between 1940 and 1962. We first show this increase can be well instrumented by the number of planned schools, and then use an IV regression to estimate its effects on the enrolment of children and on the literacy rate of the cohort affected by the plan.

Prior to the *Centenários* plan, Portugal had a poorly educated population: In 1940, only 40% of the population knew how to read (see Figure 1a); the average years of schooling for people aged 25–64 was only 1.6, compared to close to four years in Spain, Italy, or France (see Figure 1b).

[Figure 1 here]

In this context, the government assumed an educational policy of mass literacy, using a single and simplified educational program that sought the aquisition of a "minimal" level of knowledge (Magalhães, 2010). This new policy began with the creation of the Ministry of National Education in 1936 and the educational reform of 1938 (see Palma and Reis, 2018, for a description of the more immediate measures), and it was followed by a centralized plan for the construction of school infrastructures throughout the country—the *Centenários* plan. The initial guidelines of the plan were approved in December 1940 and the original layout of schools in July 1941. In the five decades after the plan, Portugal's literacy rate grew by 50 percentage points and the average years of schooling jumped to seven.

For a perspective on the scale of the *Centenários* plan, we can compare it with other school expansions. For example, the Rosenwald schools iniciative in the southern United States in the first third of the XX century increased the number of existing schools by almost 5,000 in 16 years (Aaronson and Mazumder, 2011). Other examples, described in Aghion et al. (2019) for the last quarter of the XIX century, include France where 17,320 new schools were built and Japan where 12,500 primary schools were established in the space of 1 year. More recently, there is the school construction plan in Indonesia in the 1970s, which aimed to build 61,000 primary schools (Duflo, 2001). The Centenários plan can be directly compared to the latter example. In Portugal, more than four schools were planned per 1,000 children aged 5–14, compared to two schools per 1,000 children of the same age group in the Indonesian case.

Regarding the execution of the plan, we do not have information regarding which and when schools were built. Some evidence shows the government was slow to implement the plan, and only a small fraction of it was carried out before 1950 (Beja et

al., 1996), most likely due to the plan's requirement of matching funds from local authorities to build the schools.[1] However, in 1962, mainland Portugal had 7,417 more operating schools than it had in 1940—equivalent to 120% of the number of schools projected in the plan. Some of this increase was due to an increase in private schools, mainly in Lisbon and Porto.[2]

We use three main sources to construct our county level dataset. The first source is the *Education Statistics Yearbook* for 1940–41 and 1962–63 (from the Portuguese Estatísticas da Educação, Ano Lectivo 1940–41 and Ano Lectivo 1962–63), which we use to characterize the schooling situation at the beginning of the period under review and the change in enrolment rates between school years 1940–41 and 1962–63. The second source is the Portuguese Census for 1940 and 1960, from which we draw demographic and literacy data for 1940, and literacy outcomes for different cohorts in 1960. The third source has information about schools and classrooms planned by the *Centenários* plan in 1941; it is drawn directly from the document *Mapa Definitivo das Obras De Escolas Primárias* (1943), which consists of the list of parishes affected and the number of schools and classrooms set to be built in each. More details about the construction of the dataset and the institutional setting are given in section 2.

Section 3 describes the state of schooling in Portugal in 1940. At that time, the country had a deficit of primary schools, with only one primary educational establishment per 700 inhabitants. The literacy rate of the population aged 7–19

[1] Decreto Lei 38968 (Diário do Governo, I Série, 1952) reports that by July 1952, a total of 1,390 schools were constructed within the *Centenários* plan and 292 were under construction.
[2] According to the *Education Statistics Year Book* for 1962, Portugal had 382 primary private schools and 379 schools that taught several years including primary. The total adds to 761 private schools in the mainland in 1962 versus 474 in 1940.

varied between 22% and 79% across counties.[3] We show these differences were mainly related to the number of schools per resident, the quality of education measured by attendance and exam pass rates, and the population density, as well as other demographic factors.

Section 4 analyzes the *Centenários* plan. We conclude that the plan envisaged the construction of more schools in counties with lower literacy rates. At the county level, a 10-percentage-point difference in the percentage of the population aged 7–19 who knew how to read translated into 0.2 fewer schools planned per 1,000 residents. Given the close to 60-percentage-point difference in this group's literacy rates between the least and the most literate counties at the time, our estimate implies a difference of 1.2 schools planned per 1,000 residents between these counties.

In section 5, we analyse the effects of the plan on student enrolment and the literacy rate of the affected cohort in the early 1960s. We start by showing the number of planned schools by county is a strong predictor of the subsequent variation in the number of schools from 1940 to 1962, even after controlling for several county characteristics in 1940. Therefore, the *Centenários* plan can be used to empirically identify the causal effects of school construction on education, similar to Duflo (2001, 2004), Breierova and Duflo (2004), Martinez-Bravo (2017) or Somanathan (2008), using the Indonesia case. More precisely, we estimate the causal effect of the change in the number of schools per resident, instrumented by the number of schools planned per resident, in the change in student enrolment from 1940 to 1962 and the literacy rate of the relevant cohorts in 1960, respectively. We find that an additional school per 1,000 residents increased enrolment by 20 students per 1,000 residents older than 7

[3] Literacy is defined hereafter according to the information available in the 1940 census as "able to read."

and increased 5 percentage points the literacy rate of the population aged 7 to 39 in 1960, relative to that of the population older than 40 in 1960.

Our paper contributes to a multidisciplinary literature studying Portuguese education during the 20th century, including Beja et al. (1996), de Carvalho (1986), Nóvoa (1992), or Mónica (1978) and Ramos (1988) in history; Mendonça (2007) in sociology; Cordeiro et al. (2013) or Santos et al. (2013) in geography; Candeias (1998) Candeias et al. (2004), Magalhães (2010, 2018) and Martins (2009) in educational sciences; or Amaral (2002), Reis (1993) and Palma and Reis (2018) in economics. The latter paper is particularly relevant to our analysis, although it refers to a period that just precedes ours. Using the sample of military records of 20-year-old individuals in 1924, 1931, 1941 and 1950, its authors compare individuals' literacy rates between the republic and the dictatorship prior to the *Centenários* plan. They argue the increase in the literacy rates can be attributed to the education reform early in the dictatorship that included a more simplified national program imprinted with a Catholic identity and gender separation in schools.[4] Their results support the belief among historians (e.g., Reis, 1993; Ramos, 1988) that the increase in the literacy prior to the *Centenários* plan was mostly due to demand-side factors just as others have alluded to in other contexts (e.g., Cinnirella and Hornung, 2016). Amaral (2002) hints at the joint role of supply and demand factors in the increase in enrolment during Estado Novo. His empirical analysis with macro data on the number of students supports the view that supply factors preceded demand factors. Our paper complements these authors' analysis by showing the causal impact of the increase in the supply of schools in explaining the subsequent increase in enrolment and literacy.

[4] Cappelli and Vasta (2019) also argue that centralisation in Italy had benefits in terms of literacy.

The nation-building literature describes several historical examples where public funds were used to strengthen the support for incumbent governments. Cinnirella and Schueler (2018), for example, show how an increase in the share of centralised funding of schools, particularly the one used to pay for teachers' wages, increased the share of votes in nationalist parties in Prussia at the turn of the 20th century. Aghion et al. (2019) show, and illustrate with several historical examples (France, Japan and China), how military threats have encouraged investment in education since the 19th century. In another example, Bandiera et al. (2018) describe how mandatory education was used in the US to transmit and reinforce the national identity to migrant children. Moreover, the received literature mantains that authoritarian regimes are more prone to use education policy with the purpose of indoctrination or to strengthen the support for the regime (see, e.g., Alesina and Bryony, 2019, Lott, 1999, Ramirez and Boli, 1987, and the references therein). In the Portuguese context, Mónica (1977) describes how the dominant conservative elite realized the convenience of disciplining the masses through education—which proved crucial inthe transmission of patriotic values to the young generations—, thus abandoning the belief that education was useless and unwanted by the people. The regime's identity building was further reinforced, particularly during the first years, with the help of state-sponsored propaganda (e.g., Ribeiro, 2005; Corkill and Pina, 2009). The embarrassment associated with an international image of backwardness portrayed by the staggering number of iliterates may have also played a role in the inception of the *Centenários* plan. As some claim, the literacy results from the 1940 census were purposely put on hold not to darken the image of the country the regime wished to transmit during the international exposition of 1940 (as cited in Candeias et al., 2004, page 162).

Our paper, however, cannot provide an answer to whether the *Centenários* plan was part of a long-term strategy to ensure the continuation of Salazar's regime, or whether its purpose was simply to replicate the mass education policies that had proved successful in other countries during the 19th and early 20th centuries (e.g., Aghion et al., 2019). Instead, we focus on the actual effects of the plan on school enrolment and literacy rates. As such, this paper is also related to a literature on the supply determinants of human-capital accumulation to which many of the previously cited papers belong.

## 2. Construction of the database and institutional setting

We build a database for mainland Portugal with data aggregated at the county level.[5] Mainland Portugal has 18 districts, 272 counties and 3,791 parishes (the equivalent to *Distrito*, *Concelho* and *Freguesia* in Portuguese). Table 1 lists the sources of the datasets we used, the variables of interest, and the year to which they refer. Table 2 gives the basic statistics of the main variables, used in the regressions. Figure 2 maps the Portuguese districts for the unfamiliar reader.

[Table 1 and 2 here]

[Figure 2]

The first dataset was extracted from the *Education Statistics Yearbook* for 1940-1941, which gives a snapshot of the formal education system. In the 1940-1941, mainland Portugal had 9,661 primary schools, 12,726 teachers and regents and 506,086 students. Out of the 9,661 primary schools, 7,264 were regular schools and

[5] In this paper, we only focus on mainland Portugal and disregard Azores and Madeira, as well as the overseas colonies.

2,397 were school outposts ("postos escolares"). School outposts were created in 1931 as an "additional instrument to deal with the so called problem of illiteracy" (Decreto Lei 20604). Primary education was provided by teachers in schools and by regents ("regentes") in school outposts. The academic requirements needed to be a regent were less demanding than those required to be a primary teacher; therefore, academics view teaching at school outposts to be of lower quality (Magalhães, 2010, 2018). These numbers include both public and private schools. Private schools were scarce in primary education in 1940: The number of private schools with any primary teaching, 474, represented less than 5% of the total number of schools and schools outposts in the country, and their students represented only 9.3% of the total number of primary students (according to Estatisticas da Educação, Ano lectivo 1940–41). Private schools were mostly concentrated in Lisbon—with 30% of the primary students in private education—Évora (11.5%), Porto (8.6%), Setúbal (8.4%) and Coimbra (8.2%). To ensure the results are not influenced by the presence of private schools, in our regressions, we include specifications with a dummy for the capital of the district, where most private schools were located, and one specification excluding the cities of Lisbon and Porto.

In 1930, the four years of primary education were divided into two groups: the first three years of school (mandatory) and a fourth year that was optional (Decreto Lei 18140). By the end of the third and fourth grades, students took an exam and upon passing were granted a certificate.

Concurrently with the *Centenários* plan, the government created special tertiary schools to qualify teachers who were needed for the school expansion. As described in the introduction to the 1952 Law (Decreto Lei 38968), 14 such schools were either

created or reopened between 1942 and 1948 throughout the country, mainly in district capitals: Lisboa, Porto, Coimbra, Braga, Viseu, Évora, Faro, Guarda, Bragança and Vila Real (as well as in the islands: Funchal, Ponta Delgada, Horta, e Angra do Heroísmo).

Our second dataset is taken from the 1940 census. The statistics are contained in books—one for each Portuguese district—that have been digitized by the Portuguese National Institute of Statistics. The books include data on the total population in each Portuguese county by gender, civil status, and religion. They also include other information such as surface area, a detailed age decomposition of the population and, for different age groups, a self-reported account on primary school attendance and completion and whether they knew how to read. Aggregate literacy rates, as calculated by the official statistics, include everyone who knows how to read, divided by total population, including children younger than 7 that have not yet attended school. As such, we prefer to have two alternative measures of literacy for a young cohort aged 7–19 and an older cohort aged 20 or older.

We draw information on schools and classrooms planned in 1941 directly from the document *Mapa Definitivo das Obras De Escolas Primárias, 1943*, which consists of the listing of the parishes affected by the plan and the corresponding number of schools and classrooms to be built.[6] The document also includes the principles that guided the plan: (i) Children should not have to travel more than 3 km to attend school; (ii) boys and girls should attend separate schools (exceptions allowed in areas of very low population density); (iii) class size should be limited to 40 students; (iv) schools should have a maximum of four classrooms; and (v) the national government would finance

[6] The listing only refers to "schools" and does not mention the construction of school outposts even though many of the planned educational establishments had only one classroom.

up to 50% of the construction costs, and local administration should cover the remaining costs. The plan included all 18 districts, 269 counties, and 2,831 parishes in mainland Portugal. Only three counties—Constância and the cities of Lisbon and Porto—and 960 (of the 3791) parishes were unaffected by the plan.

Although the natural end point of our analysis would be just prior to the end of the dictatorship, that is, the early 1970s, we chose to measure the effects of the plan by the early 1960s for several reasons. First, the 1970 census contains only preliminary results for a subset of the population, and no comprehensive questions of literacy by county. Second, using the 1960s seemed to us more appropriate considering that the major jump in literacy among children had occurred by 1960 as reported by Candeias et al. (2004). Third, to avoid confounding effects from large population movements (due to the colonial war and emigration), as well as the increase to six years of mandatory education, approved in 1964 (Decreto Lei, 45810, July 1964), that could have diverted funds from primary education as well as teachers from some schools.

The information on the number of schools and school outposts, as well as student enrolment by county in 1962–63 is taken from the *Education Statistics Yearbook* for 1962–63. Finally, we use the 1960 census for the literacy rates of different cohorts. Our definition of literacy rate is taken from the available information in the census. In 1940 and 1950, the census asked about the "ability to read". As Candeias et al. (2004) point out, many of those who were able to read (and write) had never received formal education. In 1960, the census question changed such that all those who did not

attend school were assumed to be illiterate. Therefore, the literacy rates in 1960 are biased downwards, compared to the 1940 census.[7]

## 3. Literacy in Portugal in 1940

The literacy numbers for Portugal in 1940 are staggering. Candeias et al. (2004) argues that Portugal at the time was one century behind other European countries in terms of literacy rates. Earlier education policy had limited effects on literacy rates, which led historians to point out as the main causes, not so much the lack of schools, but the high opportunity costs of education and the absence of clear benefits of learning how to read or write, particularly among the rural poor (e.g., Mónica, 1977; Reis, 1993; Ramos, 1988). Without detailed individual data, we cannot quantify which of these factors are more important. For this reason, in this section, we only aim to identify the main aggregate factors associated with cross-county differences in the literacy rate of the population aged 7–19 in 1940.

The literacy rate varied between 22% and 79% across counties, and its geographical distribution, together with the literacy rates of the older cohort, can be seen in the maps of Figure 3. The literacy rates of the younger cohort are, on average, 14 percentage points higher than the literacy rates of the older cohort, which might reflect the effects of the education reform early in the dictatorship, discussed in Palma and Reis (2018).

[Figure 3 here]

[7] In a companion paper, in Portuguese, we provide a more descriptive analysis of the three first datasets by district (Gomes and Machado, 2018).

We present results from a linear regression of the literacy rate of the population aged 7–19 by county, on a series of explanatory variables. These results, although not causal, identify the variables that are more associated with the literacy rate.

The main variable of interest is the total number of school establishments per 1,000 residents, which reflects the quantity aspect of the supply side. Ideally, we would distinguish between schools and school outposts, but this information is not available at the county level. The map on the left in Figure 4 shows the geographical distribution of the school establishments (i.e., schools and school outposts) per 1,000 residents. On average, we find 1.5 schools per 1,000 residents, with a large range across counties, from 0.18 to 2.8.

[Figures 4 here]

We then include a group of variables related to school attendance and achievement, proxies for the quality of school supply. The first of these variables is the ratio between the number of children aged 7–13 who attended primary school, according to census data, and those who were enrolled, according to Education Statistics, which gives a measure of effective attendance. Although this measure is an imperfect measure of real attendance, it has the advantage of being measured at similar time periods across the country—because both the enrolment period and the census surveys were common to all regions in the country. This variable may also reflect the capacity of the state to enforce compulsory schooling at different counties, and/or other variables, such as the prevalence of child labour or the (opportunity) cost of attending school. The second variable reflecting school quality is the number of students who passed the third-grade exam, relative to the total number of students.

The third and fourth groups of variables are composed of geographic and sociodemographic variables, respectively. We include dummy variables of district and district capitals, the surface area of the county (in logarithm), and the number of parishes in the county as geographical variables. These variables should account for differences across counties for which we do not have a direct control, for example, land inequality and political influence of regional governments (Go and Lindert, 2010). For sociodemographic variables, we use the resident population (in logarithm), the age structure as measured by the percentage of children aged 0–13 on the total population, the ratio of men to women, the percentage of people who were married, and the percentage of residents who were not Catholic.

Table 3 shows the results of linear regressions for the literacy rate of the population aged 7–19. In column (1), we consider only the number of educational establishments per 1,000 residents. One additional school per 1,000 residents in the county is associated with a 4.2-percentage-point increase in the literacy rate of the population aged 7-19 years in 1940, that is, a 8.1% increase in the literacy rate.

In columns (2) and (3), we added the measures of attendance and exam pass rates, without and with district dummies, which prove to be important in explaining the literacy rate. With the inclusion of these variables, the R-squared increases from 0.044 to 0.36 and 0.53, respectively. Higher attendance and pass rates are both statistically significantly related to higher literacy rates.

[Table 3 here]

In columns (4) and (5), we include the geographic and demographic variables, and in columns (6) and (7), we include all the controls. Note the coefficients of the

logarithm of the surface area and the logarithm of the population have opposite signs although they are virtually of the same magnitude, which means the ratio of the resident population per unit area, that is, the population density, is what matters for the literacy rate. A 1% increase in the population density, either by a larger population or by a smaller area, is associated with increases of the literacy rate of around 0.03 to 0.05 percentage points. The capitals of districts had, on average, literacy rates that were 5 percentage points higher than the rates in other counties. Regarding demographic variables, counties with a higher share of children aged 0-13 or a higher male-female ratio had, on average, lower literacy rates.

In columns (3), (5) and (7), we include district dummies, which capture permanent differences between counties in different districts, and in column (8), we excluded the cities of Lisbon and Porto. When we include all variables, the model explains 78% of the variation in the dependent variable, and the coefficient of schools per 1,000 residents doubles in size, corresponding to 8.9 percentage points or 17%.

## 4. The *Centenários* plan: Determinants of schools by county

In this subsection, we estimate a linear regression of the number of schools planned per 1,000 residents at the county level over a series of explanatory variables. The map on the right in Figure 3 shows the distribution of schools planned per 1,000 residents in 1940 by county. On average, close to 1 school was planned for each 1,000 residents, varying from 0 to 2.3. Two large regions stand out with a high concentration of new schools per county, namely, the south spanning Setubal, Évora, Beja and Faro, and the central northern region of Vila Real and Viseu.

In addition to the variables used in section 3, we also include the literacy rate of the population aged 7–19 to assess whether the plan aimed to provide more schools in areas where the percentage of illiterates was higher. Table 4 shows the linear regression results for the number of schools planned per 1,000 residents. In column (1), we consider as explanatory variables only the literacy rates, and in (2), we add the number of existing schools. In column (3), we include the geographic and demographic controls, and in column (4), we add the attendance and exam pass rate. In column (5), (6) and (7), we add different sets of variables together with district dummies. In column (8), we drop Lisbon and Porto to check the robustness.

[Tables 4 here]

The number of planned schools is negatively associated with the literacy rate of the population aged 7–19, and this association is not only statistically significantly, but also persists and even becomes stronger as we increase the number of controls.[8] A 10-percentage-point increase in the percentage of people aged 7–19 who knew how to read in the county reduces by roughly 0.2 the number of schools planned per 1,000 residents. Given the close to 60-percentage-point difference in literacy rates between the least and the most literate counties at the time, our estimate implies a difference of 1.2 schools planned per 1,000 residents between them. This difference is economically significant, representing 120% of the average number of schools planned per 1,000 residents (see Table 2). Additionally, we find that, conditional on the literacy rate, the greater the number of existing schools per 1,000 residents in the county, the greater the number of schools planned. For each additional existing school per 1,000

[8] Similar results are obtained with the literacy rate of the entire population.

residents, the plan projected an increase of 0.3 schools. These two effects are statistically significant in all specifications and explain 37% of the variability of the planned schools.

The percentage of students enrolled who actually attended primary school (i.e., the "attendance rate") is associated with a larger number of schools planned. This variable likely reflects the pressure or congestion of existing schools. With regards to the sociodemographic variables, we find that fewer schools were planned in regions with a higher percentage of non-Catholics, traditionally the interior of the Alentejo region in south-central Portugal, and fewer schools were planned in counties with a larger proportion of married residents; both variables are likely indicators of wealth.

## 5. The effect of the number of schools on school enrolment and literacy

### 5.1. The school landscape 20 years after the plan

We start by analysing whether the actual change in the number of schools at the county level between 1940 and 1962 can be predicted by the plan. The geographic distribution of the variation in schools between 1940 and 1962 is shown in the right-hand side map in Figure 5. In 1962, mainland Portugal had 7,417 more operating schools than it had in 1940—120% of the number of schools projected in the plan. The average variation of schools per county was 1.06, compared to the 0.98 predicted by the plan.

[Figure 5 here]

We run a linear regression where the dependent variable is the variation in the number of schools between 1940 and 1962 per 1,000 residents in 1940, against the

number of planned schools, also normalized by 1,000 residents in 1940 (see column (1) of Table 5). This regression is shown graphically in the scatter plot in Figure 7. We find that one aditional planned school predicts an increase of 0.8 schools. This variable alone explains 30% of the cross-county variation in our dependent variable. The reason the coefficient is statisticaly different from 1 may be due to changes in the demographic characteristics that led to the updates to the plan in 1955 or to the inability to fulfill the plan.

[Table 5 here]

[Figure 7 here]

The inclusion of other variables in the regression decreases the coefficient of the planned schools, but it remains statistically significant and relatively high at around 0.6. In subsequent columns, we add different controls, so that the regressions reflect the first stage of the IV regressions in the following subsection. Only a few other variables are statistically significant: the surface area of the county, which is positively associated with the increase in the number of schools; the population, which is negatively associated with the increase in the number of schools; and, perhaps more interesting, the existing number of schools per 1,000 residents, which has a negative coefficient of around -0.3. Notice that although the literacy rate in 1940 of the population aged 20 and above is not significant, we find a statistically positive effect of the literacy gap between this cohort and the younger cohort (aged 7–19).

The number of planned schools by county is strongly correlated with the number of schools built. Nonetheless, to be a good instrument for the latter, we need to make sure the number of planned schools is not correlated with underlying

characteristics that affect our dependent variables: enrolment and literacy. For example, if more schools were planned in counties where individuals would do better even without the school construction, our IV estimates could be biased upwards. In fact, we show the opposite is true; that is, more schools were planned where literacy was lowest, and thus we should not expect a positive bias. Moreover, notice that although the number of schools built is endogenous because it depended on the availability of matching local funds, this is not the case for the number of planned schools, which were exogenous to the availability of these funds.[9] In the regressions below, we include specifications where we control for existing literacy rates and the number of existing shools in order to control for variables that show some correlation with our instrument.

### 5.2. The effect of more schools on enrolment and literacy

We proceed to estimate the causal effect of the change in the supply of schools between 1940 and 1962 on two variables: the change in students enrolled in primary education divided by the population aged 7 or above, between 1940 and 1962, and the difference between literacy rates of the young and old cohorts in 1960. Figure 6 shows the literacy rates in 1960 of the population aged 7–39 and the population older than 40.[10] The average literacy rate of the younger cohort is 80% (see Table 2), which is double the literacy rate of the older cohort. This difference between the literacy rates

[9] Aaronson and Mazumder (2011), for example, discuss the potential selection bias coming from matching funds on the actual building of schools.

[10] The 1960 census only gives literacy outcomes at the county level for two cohorts: people aged 7–39 and 40 or older. Throughout the paper, we try to refer to the same cohort. For example, in Table 3, our dependent variable is the literacy rate in 1940 of those aged 7–19, which would be 27 to 39 years old in 1960, and hence included in the 7–39 cohort.

of the two cohorts varied from 15 to 53 percentage points across counties. Enrolment increased by 26 students per 1,000 residents older than 7, varying from -46 to 277.

[Figure 6 here]

We point out some notes on these outcome variables. First, the variation in the number of students enrolled could have been influenced by legislation passed in 1952 (Decreto lei 39698) that recognized the inoperability and inefficacy of the existing measures aimed at fostering enrolment and school attendance. The new legislation not only increased sanctions to parents whose children did not attend school but, more importantly, modified the way they were put into practice. Together with the reinforcement of fines and penalties, the collection of social benefits such as family allowance was made conditional on the children's school attendance. Second, in 1960, the census definition regarding literacy changed relative to previous years such that all those who did not attend school were assumed to be illiterate. The new definitiom amplifies the differences in literacy across cohorts, because a higher share of the older cohort knew how to read but had not attended school. On the other hand, the 1952 legislation introduced strong incentives for the education of adults, which could have had the opposite effect i.e. the reduction of differences across cohorts.[11] Third, although the *Centenários* plan clearly did not affect the 40-and-older cohort in 1960, because those individuals were 20 or above when the plan was approved, it did affect some individuals of the younger cohort, namely, children aged 7–13 when the plan

[11] Among the incentives introduced to foster adult literacy was the requirement of a primary-education exam certificate to be able to obtain a driver's license or to become a public servant or to hold many other jobs. Illiteracy also affected men's extension of and status within during military service as well as individuals' permission to emigrate.

was approved. We loosely denote the younger cohort aged 7–39 in 1960 as the "affected cohort". Despite these limitations, we believe using information exclusively collected in 1960 is preferable in order to ensure the same definition of literacy is applied to both cohorts. A more precise interpretation of our second outcome variable is the change in the fraction of people who attended school.

Because the change in the number of schools between 1940 and 1962 by county is endogenous, we instrument it using the number of planned schools by county in 1940. The identifying assumption is that the number of planned schools in 1940 affects the change in literacy between cohorts in 1960 only through its effect on the real change in the number of schools, conditional on characteristics of the population and the county in 1940, as well as district fixed effects. Table 5 shows the first stage regressions, and Tables 6 and 7 show the results of the second stage, both IV and OLS. In all regressions, we reject the underidentification and the weak instruments tests, which validate our approach (see tests at the bottom of columns 4–8).

[Tables 6 and 7 here]

The effect of the change in the number of schools per resident on the change in student enrolment is always positive and statistically significant both for OLS and IV regressions. The OLS results are systematically smaller than the IV. Our IV estimates imply one additional school per 1,000 residents—the average increase was 1.06—increases enrolment by around 20 students per 1,000 residents older than 7 (see column (5) results in Table 6), which corresponds to close to 80% of the average increase in enrolment.

The effect of the change in the number of schools per resident on the variationof literacy rates between cohorts is positive and statistically significant both for OLS and IV regressions. Because more schools were built in less favored counties, the OLS results are systematically smaller than the IV. Our estimates imply one additional school per 1,000 residents increases the difference in the literacy rate by around 4.3 percentage points, which corresponds to 10.7% of the average difference in the literacy rates.

To further address endogeneity concerns about our instrument, we further control for the difference in literacy rates of the same cohorts in 1940. Results are even larger, around 5.3 percentage points, that is, around 13% of the average difference. Results show that the difference in literacy rates across cohorts is larger in smaller counties, counties with a larger proportion of children and married people, and with a larger fraction of Catholics. The latter result, statistically significant in all but the last two specifications, is consistent with the theory which attributs a role to the curriculum's catholic imprint in persuading parents to send their children to school (Palma and Reis, 2018).

## 6. Conclusion

In 1940, the Portuguese population was mostly illiterate, due in part to a scarce and unequal network of primary schools. With this background, the government launched the *Centenários* plan, with the objective of increasing the number of primary schools in the country by 60% over a 10-year period.

We collect data from three main sources to characterize the state of formal education in Portugal at the time of the beginning of the *Centenários* plan, the strategy

the plan followed regarding the distribution of new schools, and, finally, the effects of the increase in the number of schools 20 years after the plan. The quality and detail of the data used allow a rigorous statistical analysis at the county levels.

From our analysis, we conclude that the variation in literacy rates among counties in 1940 was mainly related to the number of schools, the quality of education, measured both by the percentage of enrolled students attending class and by the passing rate in the third-grade exam, and the population density, as well as other demographic factors.

As for the strategy followed by the *Centenários* plan, we conclude the construction of schools privileged counties with low literacy rates. For example, a 10-percentage-point difference in the number of children aged 7–19 years who knew how to read translated into a difference of approximately 0.2 schools per 1,000 residents, that is, close to 20% of the average number of schools planned per county.

The variation in the number of schools 20 years after the approval of the *Centenários* plan is highly positively correlated with the initial number of schools planned, conditional on other county characteristics.

This strong correlation allows us to use the *Centenários* plan as a policy experiment to help identify the effects of schooling. Hence, in the last part of the paper, we estimate the causal impact of the variation in the supply of schools per resident and conclude it was responsible for the 80% increase in enrolment and the 13% increase in the literacy rate of the affected cohorts in the beginning of the 60s.

Our results contribute to the debate regarding the causes of the slow growth in Portuguese literacy. We establish that supply-side factors, in particular, the massive

school construction during part of the Estado Novo, proved relevant for the increase in literacy during the 1950s and early 1960s.

The analysis carried out in this paper is the start of a more ambitious project in which we will study the long-term effects of the expansion of access to primary education on the Portuguese society.

## Acknowledgements

We thank Nuno Palma and Jaime Reis for comments on a previous version of the paper, Agustin Casas, Julio Cáceres, Luigi Minale and Jaime Millán, as well as three anonymous referees. We would also like to extend our thanks to Françoise Le Cunff, from the archive of the Ministry of Education, for supporting us in the collection of some of the data.

# Tables and Figures

**Table 1: Datasets and variables**

| Source and variables | Year |
|---|---|
| **Education statistics, yearbook 1940-1941** | |
| Schools and school outposts | 1940-41 |
| Students registered (male and female; 1st - 4th year) | 1940-41 |
| Students proposed to 3rd year exam | 1940-41 |
| Students approved in 3rd year exam | 1940-41 |
| **1940 Census** | |
| Surface area | 1940 |
| Total population | 1940 |
| Population by gender | 1940 |
| Population by literacy status and age groups | 1940 |
| Population by civil status | 1940 |
| Population by religion | 1940 |
| ***Centenários* Plan** | |
| Schools planned | 1941 |
| **Education statistics, yearbook 1962-1963** | |
| Schools and school outposts | 1962-63 |
| Students registered (male and female; 1st - 4th year) | 1962-63 |
| **1960 Census** | |
| Population by literacy status and age groups | 1960 |

## Table 2: Variables constructed and summary statistics

| Variable and description | Mean | St. Dev. | Min. | Max. |
|---|---|---|---|---|
| **1940s (county level, 272 observations)** | | | | |
| *Literacy rate (total pop.):* Population that knows how to read divided by total population (1940 Census) | 35.6 | 8.1 | 16.7 | 70.0 |
| *Literacy rate (pop. aged 7-19) :* Population aged 7-19 that knows how to read divided by the respective population (1940 Census) | 51.7 | 10.9 | 21.6 | 79.0 |
| *Literacy rate (pop. aged 20+) :* Population aged 20 or above that knows how to read divided by the respective population (1940 Census) | 37.7 | 9.0 | 17.1 | 76.0 |
| *Schools per 1,000 residents:* Primary schools and school outpost (Education Statistics, 1940-41) per 1,000 residents (1940 Census) | 1.55 | 0.54 | 0.19 | 2.83 |
| *Planned schools per 1,000 residents:* Number of schools planned ("Plano dos Centenários" ) divided by 1,000 residents (1940 Census) | 0.98 | 0.40 | 0.00 | 2.33 |
| *Attendance rate:* Population aged 7-13 whose parents claim to attend primary school (1940 Census) divided by the students enrolled in the county (Education Statistics, 1940-41) | 0.51 | 0.11 | 0.22 | 1.02 |
| *Exam pass rate:* Students passing the 3rd-grade exam divided by the students enrolled in the county (Education Statistics, 1940-41). | 0.12 | 0.02 | 0.04 | 0.25 |
| *Enrollement:* Students enrolled in primary education divided by population aged 7 or above (Education Statistics, 1940-41 and 1940 Census). | 91.5 | 17.3 | 46.5 | 134.7 |
| *Population (log):* Logarithm of the resident population (1940 Census) | 9.87 | 0.69 | 8.15 | 13.45 |
| *Surface (log) :* Logarithm of the surface area in $km^2$ (1940 Census) | 5.47 | 0.83 | 1.87 | 7.45 |
| *Number of parishes:* Number of parishes in the county (1940 Census) | 13.78 | 12.79 | 1.00 | 89.00 |
| *District capital:* Dummy variable if the county is district capital | 0.07 | 0.25 | 0.00 | 1.00 |
| *Coastal county:* Dummy variable if the county has direct access to the sea | 0.19 | 0.39 | 0.00 | 1.00 |
| *% non-Catholic:* Percentage of non-Catholics of the total population (1940 Census) | 0.07 | 0.10 | 0.00 | 0.59 |
| *% pop. aged 0-13:* Percentage of children aged 0-13 out of the total population (1940 Census) | 0.31 | 0.03 | 0.19 | 0.38 |
| *Male-female ratio:* Ratio of men and women in the resident population (1940 Census) | 0.95 | 0.08 | 0.69 | 1.12 |
| *% married:* Percentage of married population out of the total population (1940 Census) | 0.35 | 0.04 | 0.21 | 0.46 |
| **1960s (county level, 272 observations)** | | | | |
| *Schools per 1,000 1940 residents:* Primary schools and school outpost (Education Statistics, 1962-63) per 1,000 residents (1940 Census) | 2.60 | 0.83 | 0.62 | 5.12 |
| *Enrollement:* Students enrolled in primary education divided by population aged 7 or above (Education Statistics, 1962-63 and 1960 Census). | 117.3 | 28.5 | 29.8 | 351.7 |
| *Literacy rate (pop. aged 7-39) :* Population aged 7-39 that knows how to read divided by the respective population (1960 Census) | 79.9 | 5.3 | 59.7 | 91.9 |
| *Literacy rate (pop. aged 40+) :* Population aged 40 or above that knows how to read divided by the respective population (1960 Census) | 39.8 | 9.1 | 20.1 | 76.6 |
| *Variation schools per 1,000 residents (1940-1962):* Schools and School Outposts in 1962 minus Schools and School Outposts in 1940 divided by 1,000 1940's residents | 1.06 | 0.61 | -0.54 | 3.65 |
| *Variation of enrolled students (1940-1962):* Students enrolled per 1000 residents older than 7 in 1960s minus Students enrolled per 1000 residents older than 7 in 1940, divided by 1,000 1940's residents | 25.7 | 27.6 | -45.8 | 277.2 |

**Table 3: Determinants of literacy rates of population aged 7-19, by county**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Schools per 1,000 residents | 4.226*** | 6.449*** | 4.307*** | 8.265*** | 8.404*** | 9.058*** | 8.977*** | 8.958*** |
| | (1.204) | (1.013) | (1.482) | (1.085) | (1.239) | (0.912) | (1.100) | (1.101) |
| Population (logs) | | | | 4.916*** | 4.517*** | 2.704*** | 3.064*** | 3.757*** |
| | | | | (0.945) | (1.089) | (0.857) | (0.951) | (1.083) |
| Surface (logs) | | | | -5.078*** | -4.552*** | -3.415*** | -3.559*** | -3.826*** |
| | | | | (0.696) | (0.859) | (0.723) | (0.823) | (0.853) |
| Number of parishes | | | | -0.0447 | -0.0193 | -0.0300 | -0.00724 | -0.0331 |
| | | | | (0.0491) | (0.0600) | (0.0447) | (0.0508) | (0.0563) |
| Capital of district | | | | 5.760*** | 4.611*** | 4.881*** | 3.821*** | 4.149*** |
| | | | | (1.645) | (1.518) | (1.360) | (1.288) | (1.292) |
| Coastal county | | | | 1.133 | 1.621 | 1.427 | 1.631 | 1.477 |
| | | | | (1.115) | (1.231) | (0.987) | (1.118) | (1.118) |
| % non Catholic | | | | 6.955 | 4.457 | 10.22** | 10.47** | 9.986** |
| | | | | (5.344) | (5.026) | (4.269) | (4.324) | (4.330) |
| % pop. aged 0-13 | | | | -107.6*** | -146.4*** | -96.97*** | -123.1*** | -130.4*** |
| | | | | (21.72) | (23.83) | (17.93) | (19.56) | (20.17) |
| Male-female ratio | | | | -38.05*** | -43.46*** | -42.35*** | -42.94*** | -43.04*** |
| | | | | (6.600) | (9.636) | (5.737) | (9.186) | (9.071) |
| % married | | | | 30.66** | 23.31 | 28.44** | 20.67 | 16.86 |
| | | | | (14.04) | (17.37) | (12.73) | (14.68) | (14.81) |
| Attendance rate | | 49.07*** | 45.23*** | | | 30.85*** | 29.08*** | 28.23*** |
| | | (5.252) | (5.016) | | | (3.941) | (3.801) | (3.844) |
| Exam pass rate | | 95.53*** | 94.32*** | | | 67.38*** | 52.07*** | 50.93*** |
| | | (26.54) | (27.91) | | | (17.99) | (16.52) | (16.55) |
| Constant | 45.17*** | 5.112 | 21.37*** | 75.88*** | 101.5*** | 64.69*** | 82.11*** | 81.08*** |
| | (1.974) | (3.792) | (4.662) | (13.67) | (17.10) | (12.39) | (15.24) | (15.46) |
| District dummies | NO | NO | YES | NO | YES | NO | YES | YES |
| Observations | 272 | 272 | 272 | 272 | 272 | 272 | 272 | 270 |
| R-squared | 0.044 | 0.356 | 0.530 | 0.599 | 0.706 | 0.704 | 0.775 | 0.767 |

Note: Standard deviation in parentheses. The dependent variable is the literacy rate for the population aged 7-19, taken from the 1940 Census. See Table 2 for description of the construction of the variables. Column (8) excludes the cities of Lisbon and Porto. Statistical significance: *** p <0.01, ** p <0.05, * p <0.1. Robust standard errors used.

## Table 4: Determinants of planned schools, per county

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Literacy rate (pop. aged 7-19) | -0.0166*** | -0.0198*** | | -0.0248*** | -0.0140*** | -0.0233*** | -0.0177*** | -0.0179*** |
| | (0.00200) | (0.00188) | | (0.00216) | (0.00261) | (0.00285) | (0.00288) | (0.00292) |
| Schools per 1,000 residents | | 0.312*** | 0.146*** | 0.357*** | 0.221*** | 0.350*** | 0.258*** | 0.260*** |
| | | (0.0384) | (0.0531) | (0.0393) | (0.0622) | (0.0558) | (0.0626) | (0.0628) |
| Population (logs) | | | -0.000184 | | 0.0366 | 0.0954** | 0.0393 | 0.0539 |
| | | | (0.0442) | | (0.0470) | (0.0432) | (0.0463) | (0.0552) |
| Surface (logs) | | | 0.0690*** | | 0.0447 | -0.0136 | 0.0433 | 0.0377 |
| | | | (0.0255) | | (0.0315) | (0.0258) | (0.0310) | (0.0334) |
| Number of parishes | | | -0.00240 | | -0.00204 | -0.00455* | -0.00234 | -0.00281 |
| | | | (0.00248) | | (0.00249) | (0.00232) | (0.00254) | (0.00277) |
| Capital of district | | | -0.237*** | | -0.0887 | -0.113* | -0.0870 | -0.0824 |
| | | | (0.0763) | | (0.0670) | (0.0657) | (0.0643) | (0.0648) |
| Coastal county | | | 0.00956 | | -0.116** | 0.0230 | -0.113** | -0.116** |
| | | | (0.0565) | | (0.0487) | (0.0505) | (0.0491) | (0.0498) |
| % non Catholic | | | -0.831*** | | -0.608** | -0.654*** | -0.529* | -0.540** |
| | | | (0.261) | | (0.274) | (0.242) | (0.274) | (0.273) |
| % pop. aged 0-13 | | | 0.903 | | 0.406 | -1.189 | 0.247 | 0.139 |
| | | | (0.941) | | (1.104) | (0.926) | (1.081) | (1.115) |
| Male-female ratio | | | 1.335*** | | 1.010** | 0.390 | 0.840** | 0.824* |
| | | | (0.355) | | (0.429) | (0.321) | (0.425) | (0.427) |
| % married | | | -3.891*** | | -2.718*** | -2.961*** | -2.668*** | -2.716*** |
| | | | (0.712) | | (0.821) | (0.656) | (0.812) | (0.827) |
| Attendance rate | | | | 0.994*** | | 0.752*** | 0.544*** | 0.536*** |
| | | | | (0.189) | | (0.182) | (0.190) | (0.192) |
| Exam pass rate | | | | -0.699 | | -0.649 | 0.0803 | 0.0756 |
| | | | | (0.920) | | (0.944) | (0.935) | (0.937) |
| Constant | 1.840*** | 1.523*** | 0.307 | 1.281*** | 0.712 | 1.608** | 0.764 | 0.727 |
| | (0.106) | (0.117) | (0.728) | (0.146) | (0.786) | (0.653) | (0.795) | (0.802) |
| District dummies | NO | NO | NO | NO | YES | NO | YES | YES |
| Observations | 272 | 272 | 272 | 272 | 272 | 272 | 272 | 270 |
| R-squared | 0.204 | 0.371 | 0.367 | 0.424 | 0.591 | 0.500 | 0.602 | 0.584 |

Note: Standard deviation in parentheses. The dependent variable is the number of schools planned per 1,000 residents per county, according to the Plano dos Centenários. See Table 2 for description of the construction of the variables. Column (8) excludes the cities of Lisbon and Porto. Statistical significance: *** p <0.01, ** p <0.05, * p <0.1. Robust standard errors used.

**Table 5: Determinants of the variation in the number of schools between 1940 and 1962, by county (first stage)**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Planned schools per 1,000 residents | 0.836*** | 0.617*** | 0.597*** | 0.592*** | 0.616*** | 0.591*** |
| | (0.0770) | (0.0853) | (0.0897) | (0.0872) | (0.0845) | (0.0866) |
| Population (logs) | | -0.182** | -0.163** | -0.187*** | -0.210** | -0.211** |
| | | (0.0719) | (0.0811) | (0.0710) | (0.0850) | (0.0833) |
| Surface (logs) | | 0.287*** | 0.268*** | 0.296*** | 0.297*** | 0.305*** |
| | | (0.0465) | (0.0564) | (0.0467) | (0.0475) | (0.0477) |
| Number of parishes | | 0.000393 | 0.00025 | 0.00044 | 0.00150 | 0.00143 |
| | | (0.00280) | (0.0028) | (0.0027) | (0.0031) | (0.0030) |
| Capital of district | | 0.101 | 0.118 | 0.0926 | 0.0860 | 0.0782 |
| | | (0.103) | (0.105) | (0.102) | (0.100) | (0.0997) |
| Coastal county | | 0.0159 | 0.0215 | 0.00751 | 0.0217 | 0.0120 |
| | | (0.0888) | (0.0886) | (0.0890) | (0.0876) | (0.0876) |
| % non Catholic | | 0.0980 | 0.123 | 0.0178 | 0.108 | 0.0238 |
| | | (0.351) | (0.348) | (0.353) | (0.355) | (0.356) |
| % pop. aged 0-13 | | 1.343 | 0.880 | 1.086 | 1.664 | 1.402 |
| | | (1.612) | (1.739) | (1.604) | (1.714) | (1.697) |
| Male-female ratio | | 0.716 | 0.655 | 0.473 | 0.715 | 0.469 |
| | | (0.593) | (0.579) | (0.621) | (0.593) | (0.620) |
| % married | | -1.465 | -1.567 | -1.100 | -1.286 | -0.930 |
| | | (1.068) | (1.083) | (1.095) | (1.101) | (1.131) |
| Schools per 1,000 residents | | -0.339*** | -0.329*** | -0.249*** | -0.338*** | -0.248*** |
| | | (0.0807) | (0.0823) | (0.0926) | (0.0809) | (0.0930) |
| Literacy rate 1940 (pop 20 and over) | | | -0.00419 | | | |
| | | | (0.00535) | | | |
| Literacy rate gap 1940 | | | | 0.0134* | | 0.0134* |
| (pop 20 and over minus pop 7-10) | | | | (0.00687) | | (0.00684) |
| Constant | 0.239*** | 1.005 | 1.325 | 1.277 | 1.053 | 1.304 |
| | (0.0815) | (1.042) | (1.047) | (1.075) | (1.056) | (1.088) |
| District dummies | NO | YES | YES | YES | YES | YES |
| Observations | 272 | 272 | 272 | 272 | 270 | 270 |
| R-squared | 0.304 | 0.612 | 0.613 | 0.619 | 0.610 | 0.617 |

Note: Standard deviation in parentheses. The dependent variable is the variation of schools between 1940 and 1962 divided by 1,000 1940 residents. All dependent variables are from 1940. See Table 2 for description of the construction of the variables. Columns (7) and (8) exclude the cities of Lisbon and Porto. Results represent first stage regressions of IV estimates presented in the next Section (Tables 8-10). Statistical significance: *** p <0.01, ** p <0.05, * p <0.1. Robust standard errors used

## Table 6: Determinants of the change in gross enrolment between 1940 and 1962, by county

| | OLS (1) | OLS (2) | OLS (3) | OLS (4) | IV (5) | IV (6) | IV (7) | IV (8) |
|---|---|---|---|---|---|---|---|---|
| Variation of schools between 1940 | 24.40*** | 14.46*** | 13.59*** | 14.49*** | 27.17*** | 21.28*** | 18.02*** | 21.44*** |
| and 1962 per 1,000 residents | (2.334) | (2.210) | (2.308) | (2.253) | (3.055) | (4.865) | (5.167) | (4.886) |
| Population (logs) | | -6.124* | -4.033 | -6.735 | | -4.769 | -3.420 | -5.169 |
| | | (3.455) | (3.365) | (4.303) | | (3.479) | (3.313) | (4.284) |
| Surface (logs) | | 1.230 | -0.837 | 1.390 | | -1.181 | -2.122 | -1.139 |
| | | (2.486) | (2.319) | (2.739) | | (2.906) | (2.723) | (3.126) |
| Number of parishes | | 0.319** | 0.307** | 0.335** | | 0.324** | 0.311** | 0.332** |
| | | (0.138) | (0.134) | (0.165) | | (0.133) | (0.128) | (0.159) |
| Capital of district | | -11.87** | -9.549 | -11.79** | | -11.92** | -9.850* | -11.72** |
| | | (5.623) | (5.879) | (5.882) | | (5.349) | (5.525) | (5.597) |
| Coastal county | | 5.977 | 6.853 | 6.236 | | 6.452 | 7.051 | 6.683 |
| | | (5.914) | (6.033) | (6.031) | | (5.709) | (5.739) | (5.816) |
| % non Catholic | | -45.59 | -41.54 | -44.67 | | -43.44 | -40.66 | -42.53 |
| | | (48.95) | (48.19) | (48.76) | | (46.62) | (45.78) | (46.44) |
| % pop. aged 0-13 | | 218.9*** | 163.6*** | 216.7*** | | 199.4*** | 157.8*** | 194.3*** |
| | | (55.68) | (56.47) | (57.35) | | (52.67) | (53.51) | (54.42) |
| Male-female ratio | | 24.86 | 15.86 | 25.99 | | 13.16 | 9.501 | 14.12 |
| | | (27.17) | (26.51) | (27.44) | | (27.48) | (26.59) | (27.63) |
| % married | | 9.841 | 2.378 | 9.618 | | 32.64 | 17.73 | 31.51 |
| | | (41.14) | (41.00) | (42.28) | | (41.87) | (40.34) | (42.54) |
| Schools per 1,000 residents | | -18.98*** | -18.32*** | -19.05*** | | -17.11*** | -17.21*** | -17.15*** |
| | | (3.918) | (3.824) | (3.937) | | (4.079) | (4.006) | (4.085) |
| Literacy rate 1940 | | | -0.471*** | | | | -0.415** | |
| (pop 20 and over) | | | (0.174) | | | | (0.176) | |
| Constant | -0.114 | 1.920 | 40.05 | 6.812 | -3.045 | -1.947 | 33.08 | 2.653 |
| | (2.846) | (40.11) | (41.32) | (41.15) | (3.648) | (38.20) | (39.94) | (39.17) |
| District dummies | NO | YES | YES | YES | NO | YES | YES | YES |
| Observations | 272 | 272 | 272 | 270 | 272 | 272 | 272 | 270 |
| R-squared | 0.288 | 0.549 | 0.556 | 0.546 | 0.284 | 0.539 | 0.552 | 0.535 |
| Underidentification test: Kleibergen-Paag (p-value) | | | | | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Cragg-Donald Wald F statistic | | | | | 117.86 | 46.84 | 40.79 | 46.46 |
| Kleibergen-Paap Wald F | | | | | 124.24 | 52.41 | 44.32 | 53.15 |
| Montiel-Pflueger critical values | | | | | 23.109 | 23.109 | 23.109 | 23.109 |

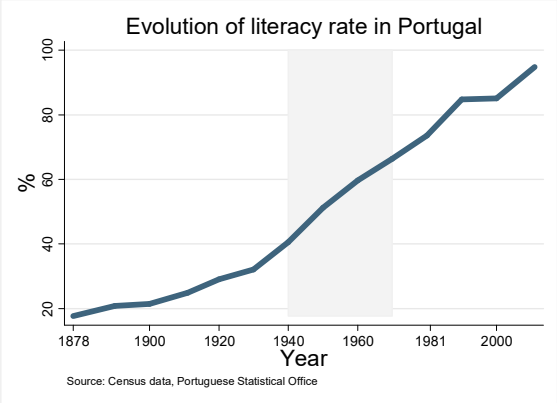
Note: Robust standard deviation in parentheses. The dependent variable is the variation in the number of students enrolled in primary schools between school years 1940-41 and 1962-63 per 1000 residents aged 7 and above. "Variation of schools between 1940 and 1962 per 1,000 residents" is the endogenous variable of change in the number of schools per thousand of 1940 residents between school year 1962-63 and school year 1940-41, instrumented by the number of planned schools per 1000 residents in 1940. The remaining variables are from 1940 (Table 2). Columns (4) and (8) exclude the cities of Lisbon and Porto. Statistical significance: *** p <0.01, ** p <0.05, * p <0.1. Robust standard errors used. Montiel-Pflueger critical values for tau=10% and alpha=5%.

## Table 7: Determinants of the variation in literacy between 7-39 and 40+ cohorts in 1962, by county

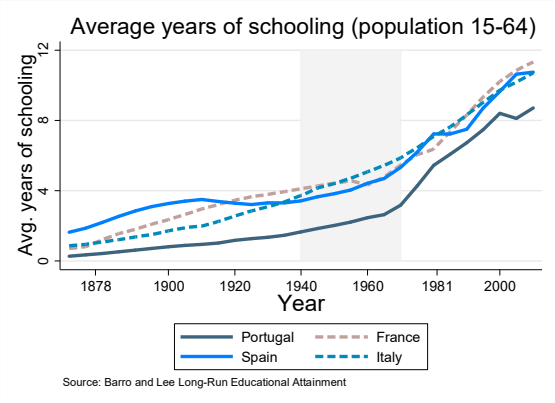
| | OLS (1) | OLS (2) | OLS (3) | OLS (4) | IV (5) | IV (6) | IV (7) | IV (8) |
|---|---|---|---|---|---|---|---|---|
| Variation of schools between 1940 | 0.957 | 1.229** | 1.670*** | 1.622*** | 6.642*** | 4.266*** | 5.275*** | 5.289*** |
| and 1962 per 1,000 residents | (0.627) | (0.515) | (0.506) | (0.502) | (1.356) | (1.158) | (1.178) | (1.195) |
| Population (logs) | | -1.992*** | -1.830*** | -2.337*** | | -1.388** | -1.094* | -1.510** |
| | | (0.590) | (0.583) | (0.671) | | (0.632) | (0.630) | (0.737) |
| Surface (logs) | | 1.657*** | 1.375*** | 1.572*** | | 0.583 | 0.0678 | 0.210 |
| | | (0.512) | (0.474) | (0.499) | | (0.606) | (0.595) | (0.621) |
| Number of parishes | | 0.0241 | 0.0227 | 0.0419 | | 0.0262 | 0.0248 | 0.0405 |
| | | (0.0283) | (0.0301) | (0.0333) | | (0.0284) | (0.0309) | (0.0346) |
| Capital of district | | -1.925** | -1.846** | -2.066*** | | -1.944** | -1.847** | -2.011** |
| | | (0.832) | (0.802) | (0.779) | | (0.865) | (0.868) | (0.849) |
| Coastal county | | -1.489* | -1.363 | -1.244 | | -1.277 | -1.087 | -0.977 |
| | | (0.857) | (0.827) | (0.824) | | (0.906) | (0.902) | (0.895) |
| % non Catholic | | -8.448** | -7.078* | -6.818* | | -7.489** | -5.622 | -5.344 |
| | | (3.849) | (3.673) | (3.693) | | (3.735) | (3.663) | (3.666) |
| % pop. aged 0-13 | | 80.94*** | 85.92*** | 90.95*** | | 72.26*** | 77.22*** | 80.86*** |
| | | (13.38) | (12.75) | (13.21) | | (13.01) | (13.27) | (14.08) |
| Male-female ratio | | 2.671 | 7.454 | 7.683 | | -2.537 | 2.699 | 2.897 |
| | | (5.956) | (5.556) | (5.517) | | (6.262) | (5.964) | (5.935) |
| % married | | 24.62** | 17.48 | 20.08* | | 34.78*** | 27.31** | 29.40** |
| | | (11.44) | (11.13) | (11.04) | | (12.37) | (12.05) | (12.05) |
| Schools per 1,000 residents | | 2.236*** | 0.660 | 0.632 | | 3.071*** | 1.214 | 1.190 |
| | | (0.661) | (0.680) | (0.692) | | (0.723) | (0.742) | (0.750) |
| Literacy rate gap 1940 (pop 7-19 | | | -0.261*** | -0.264*** | | | -0.328*** | -0.332*** |
| minus pop 20 and over) | | | (0.0599) | (0.0608) | | | (0.0690) | (0.0702) |
| Constant | 39.07*** | 7.832 | 1.928 | 3.198 | 33.05*** | 6.110 | -1.573 | -0.380 |
| | (0.794) | (10.70) | (10.37) | (10.49) | (1.562) | (10.56) | (10.55) | (10.67) |
| District dummies | NO | YES | YES | YES | NO | YES | YES | YES |
| Observations | 272 | 272 | 272 | 270 | 272 | 272 | 272 | 270 |
| R-square | 0.010 | 0.695 | 0.724 | 0.694 | -0.328 | 0.650 | 0.663 | 0.623 |
| Underidentification test: Kleibergen- | | | | | | | | |
| Paag (p-value) | | | | | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Cragg-Donald Wald F statistic | | | | | 117.86 | 46.84 | 40.79 | 46.46 |
| Kleibergen-Paap Wald F | | | | | 124.24 | 52.41 | 44.32 | 53.15 |
| Montiel-Pflueger critical values | | | | | 23.109 | 23.109 | 23.109 | 23.109 |

Note: Robust standard deviation in parentheses. The dependent variable is the diffference in the literacy rate of cohorts aged 7-39 in 1960 and cohorts aged 40+ in 1960. "Variation of schools between 1940 and 1962 per 1,000 residents" is the endogenous variable of change in the number of schools per thousand of 1940 residents between school year 1962-63 and school year 1940-41, instrumented by the number of planned schools per 1000 residents in 1940. The remaining variables are from 1940 (Table 2). Columns (4) and (8) exclude the cities of Lisbon and Porto. Statistical significance: *** p <0.01, ** p <0.05, * p <0.1. Robust standard errors used. Montiel-Pflueger critical values for tau=10% and alpha=5%.

## Figure 1

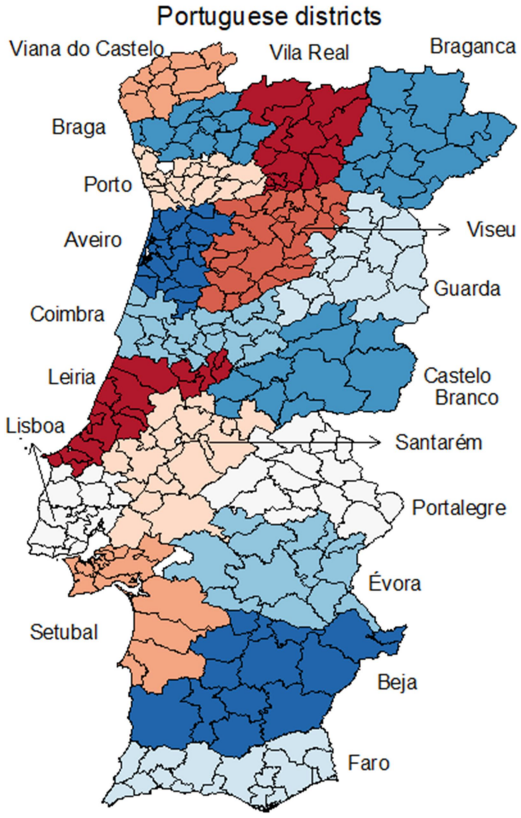



1a



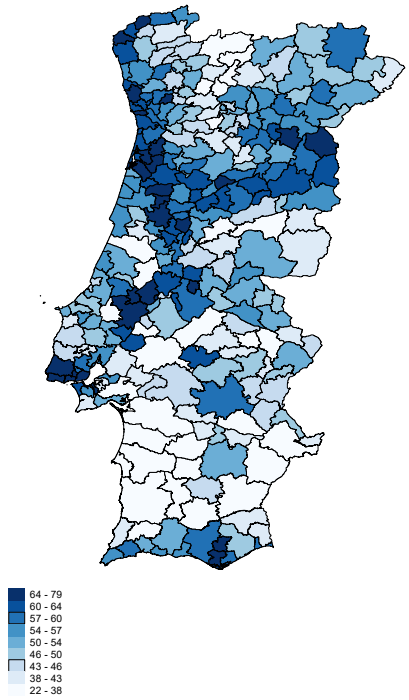

1b

## Figure 2



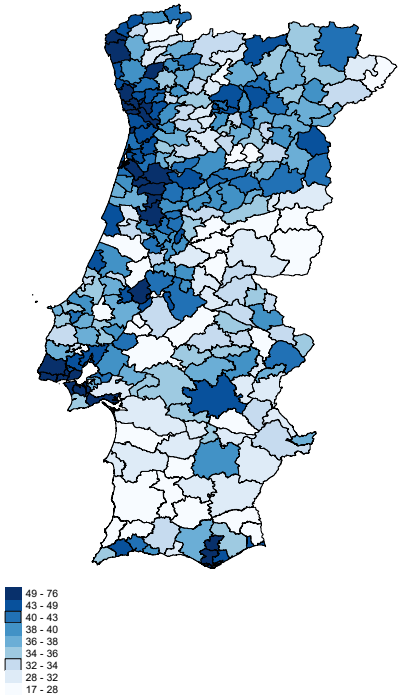
## Figure 3: Literacy rates, 1940

Literacy rate, 1940 (population age 7-19)




Source: 1940 Census

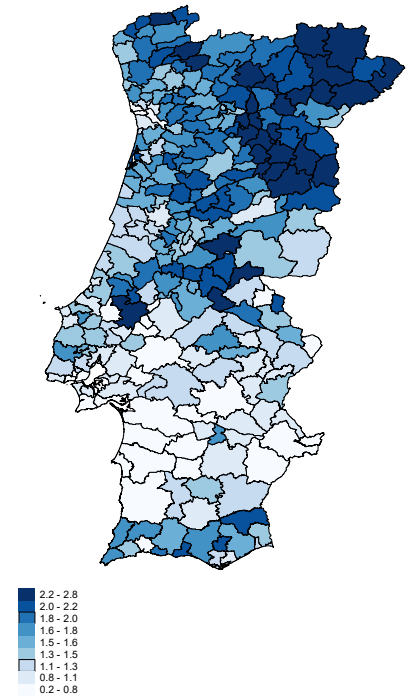
Literacy rate, 1940 (population age 20+)




Source: 1940 Census

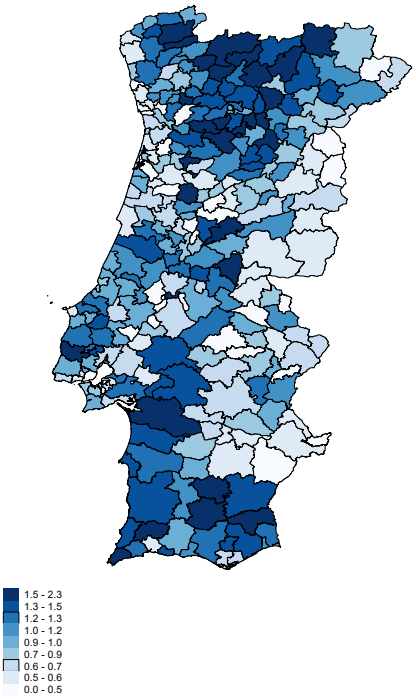
## Figure 4: Existing and planned school, 1940-1941

Schools per 1000 residents, 1940




Source: Education Statistics Yearbook 1940-1941 and 1940 Census

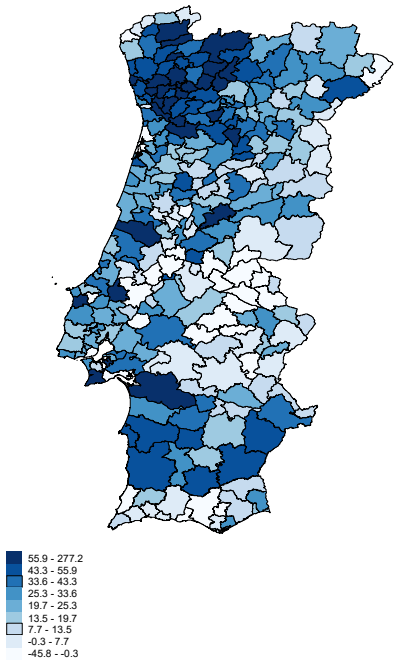
Planned schools per 1000 residents, 1941




Source: Mapa Definitivo das Obras De Escolas Primárias and 1940 Census

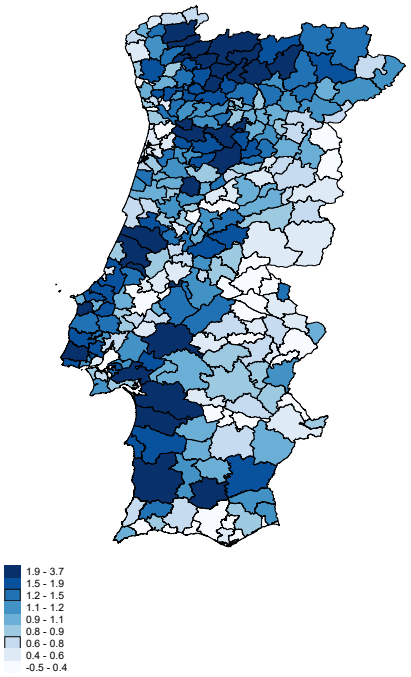
## Figure 5: Variation of schools and enrolment between 1940 and 1962

Variation between 1940 and 1962 in enrolled students per 1000 residents (7+)




Source: Education Statistics Yearbook 1940-1941 and 1962-1963 and 1940 and 1960 Census

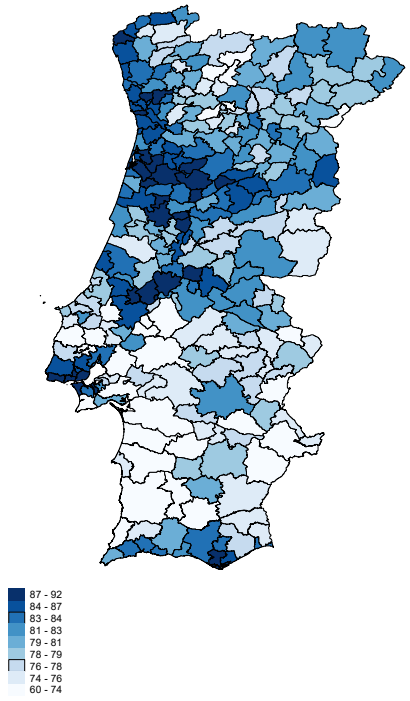
Variation between 1940 and 1962 in schools per 1940's 1000 residents




Source: Education Statistics Yearbook 1940-41 and 1962-63, 1940 Census

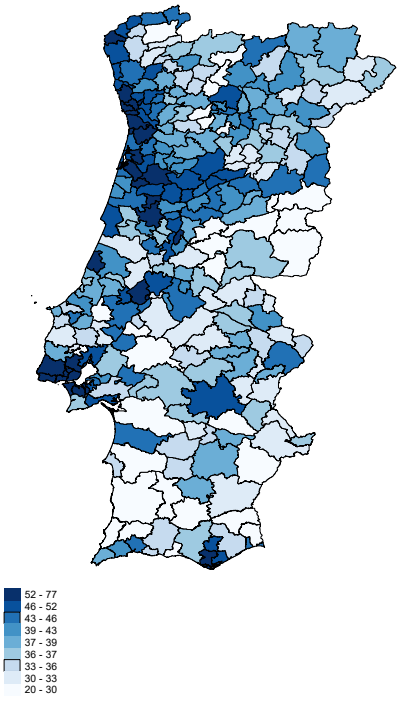
## Figure 6: Literacy rates, 1960

Literacy rate, 1960 (population age 7-39)



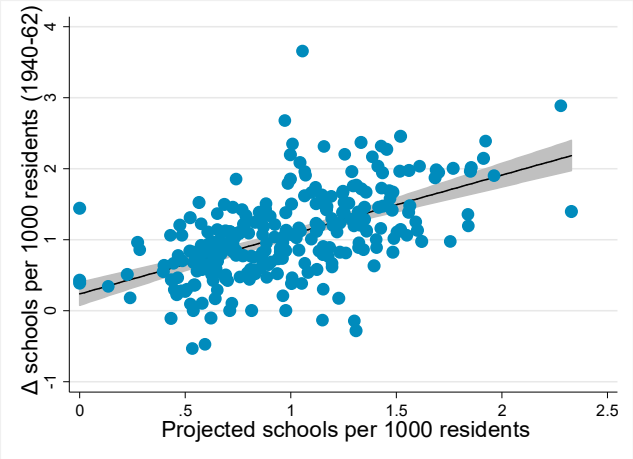

Source: 1940 Census

Literacy rate, 1960 (population age 40+)




Source: 1940 Census

**Figure 7: Projected and actual increase in schools**



## Sources and Official Publications


1. Diário do Governo I Série (1930), Número 72, 20 Março: Decreto Lei 18141
2. Diário do Governo I Série (1931). Número 283, 9 de Dezembro: Decreto Lei 20604
3. Estatísticas da Educação, Ano Lectivo 1940-1941. Instituto Nacional de Estatística, Portugal. Lisboa: Bertrand (irmãos).
4. Mapa Definitivo das Obras de Escolas Primárias: Plano dos Centenários (1943), Imprensa Nacional, 1943, Arquivo Ministério da Educação.
5. Diário do Governo I Série (1952). Número 241, 27, Outubro: Decreto Lei 38968
6. Programa de Construções Escolares do Ano de 1955. Arquivo Ministério da Educação.
7. Plano de Construções para o ensino primário (actualização do Plano dos Centenários), Lei 2107 de 5 de abril de 1961. Diário do Governo I Série, Nº 107, Decreto nº 43674.


## References


1. AGHION, P.; JARAVEL, X.; PERSSON, T. and ROUZET, D. (2019): "Education and Military Rivalry." *Journal of the European Economic Association*, 17 (2), 376-412.
2. ALESINA, A. and BRYONY, R. (2019): "Nation Building and Education." NBER Working Paper No. 18839.
3. AMARAL, L. (2002): "How a country catches up: Explaining Economic Growth in Portugal in the Post War Period (1950s to 1973)." Doctoral Dissertation European University Institute, Florence.
4. AARONSON, D. and MAZUMDER, B. (2011): "The Impact of Rosenwald Schools on Black Achievement." *Journal of Political Economy*, 119(5), 821-888.
5. BANDIERA, O.; MOHNEN, M.; RASUL, I. and VIARENGO, M (2019): "Nation-Building Through Compulsory Schooling During the Age of Mass Migration." *The Economic Journal*, doi:10.1111/ecoj.12624
6. BEJA, F.; SERRA, J; MACHÁS, E. and SALDANHA, I. (1996): "Muitos Anos de Escola. Volume II, Anos 40-Anos 70." Lisboa: Ministério de Educação.
7. BREIEROVA, L. and DUFLO, E. (2004) "The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less than Mothers?" NBER Working Paper No. 10513.
8. CANDEIAS, A.; PAZ, A. L. and ROCHA, M. (2004): "Alfabetização e Escola em Portugal nos Séculos XIX e XX: os Censos e as Estatísticas." Lisboa: Fundação Calouste Gulbenkian.
9. CAPPELLI, G. and VASTA, M. (2019): "Can School Centralization Foster Human Capital Accumulation? A Quasi-Experiment from Early XX Century Italy." *The Economic History Review*, (forthcoming).
10. CINNIRELLA, F. and SCHUELER, R. (2018): "Nation Building: The Role of Central Spending in Education." *Explorations in Economic History,* 67, 18-39.
11. CINNIRELLA, F. and HORNUNG, E. (2016): "Landownership Concentration and the Expansion of Education." *Journal of Economic Development*, 121, 135-152.
12. CORDEIRO, A. M. R.; SANTOS, L. and CARIDADE, P. (2013): "A Reorganização do Parque Escolar Nacional e a Importância das Componentes Geográficas na Metodologia Adotada." *Cadernos de Geografia*, nº 32, Coimbra.
13. CORKILL, D. and PINA ALMEIDA, J. C. (2009): "Commemoration and Propaganda in Salazar's Portugal: The Mundo Português Exposition of 1940." *Journal of Contemporary History*, 44 (3), 381-399.
14. DE CARVALHO, R.. (1986). "História do Ensino em Portugal: desde a Fundação da Nacionalidade até ao Fim do Regime de Salazar-Caetano." Lisboa.
15. DUFLO, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*, 91(4), 795-813.

16. DUFLO, E. (2004): “The Medium Run Consequences of Educational Expansion: Evidence from a Large School Construction Program in Indonesia.” *Journal of Development Economics*, 74(1): 163-197
17. GO, S. and LINDERT, P. (2010): “The Uneven Rise of American Public Schools to 1850.” *Journal of Economic History*, 70(1), 1-26.
18. GOMES, P. and MACHADO, M. P. (2018): “A Escolarização em Portugal em 1940 e o Plano dos Centenários: uma Analise Quantitativa.” mimeo.
19. LOTT JR, J. R. (1999): “Public Schooling, Indoctrination, and Totalitarianism.” *Journal of Political Economy*, 107 S6 : S127-S157.
20. MAGALHÃES, J. (2010): “Da Cadeira ao Banco. Escola e Modernização (Séculos XVIII-XX).” Lisboa: Educa/Unidade de I&D de Ciências da Educação.
21. MAGALHÃES, J. (2018): “Escola Única e Educação em Rural no Estado Novo em Portugal.” *Historia y Memoria de la Educación*, 7, pp 269-298 .
22. MARTINEZ-BRAVO, M. (2017): “The Local Political Economy Effects of School Construction in Indonesia.” *American Economic Journal: Applied Economics*, 9(2): 256–289.
23. MARTINS, E. C. (2009): “A Política Portuguesa de Educação e de Assistência Social no Período do Estado Novo (1930-1974).” Série-Estudos - Periódico do Programa de Pós-Graduação em Educação da UCDB, n. 27, jan./jun.
24. MENDONÇA, A. (2018) “O Insucesso Escolar: Políticas Educativas e Práticas Sociais: um Estudo de Caso sobre o Arquipélago da Madeira.” Coleção: Educação e Formação. Edições Pedago.
25. MÓNICA, M. F. (1977): “«Deve-se Ensinar o Povo a Ler?»: a Questão do Analfabetismo (1926-39).” *Análise Social*, Jan 1977, 321-353.
26. MÓNICA, M. F. (1978): “Educação e Sociedade no Portugal de Salazar (A Escola Primária Salazarista 1926-1939).” Lisboa: Presença, Gabinete de Investigações Sociais.
27. NÓVOA, A. (1992): “A Educação Nacional”, in Fernando Rosas (coord), “Portugal e o Estado Novo (1930-1960), Nova Historia de Portugal” (dir, Joel Serrão and António Henrique de Oliveira Marques), Lisboa: Presença, 456-520.
28. PALMA, N. and REIS, J. (2018): “Can Autocracy Promote Literacy? Evidence from a Cultural Alignment Success Story.” The University of Manchester Economics Discussion Paper 1805.
29. PIMENTA, P. (2006): “Do ‘Plano dos Centenários’ à Construção da Rede Escolar no Distrito de Vila Real.” Tese de Mestrado, Universidade do Minho, Instituto de Educação e Psicologia.
30. RAMIREZ, F. O. and BOLI, J. (1987): “The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization.” *Sociology of Education*, 60(1), 2-17.
31. RAMOS, R. (1988), “Culturas da Alfabetização em Portugal: uma Introdução à História da Alfabetização no Portugal Contemporâneo.” *Análise Social*, Vol. XXIV (103-104).
32. REIS, J. (1993): “O Analfabetismo em Portugal no Século XIX: uma Interpretação." *Colóquio Educação e Sociedade*, 1(2), 13-40.
33. RIBEIRO, N. (2005): "A Emissora Nacional nos Primeiros Anos do Estado Novo." Lisboa: Quimera Editores.
34. SANTOS, L.; CORDEIRO, A. M. R. and ALCOFORADO, L. (2013): “Evolução da Rede Escolar em Portugal: Análise dos Fatores que Historicamente têm Influenciado o Seu Planeamento.” IX Congresso da Geografia Portuguesa - Geografia: Espaço, Natureza, Sociedade e Ciência.
35. SOMANATHAN, A. (2008): “Use of Modern Medical Care for Pregnancy and Childbirth Care: Does Female Schooling Matter?" The World Bank Policy Research Working Paper No. 4625.