Culture, Institutions and School Achievement in Brazil

Daniel A. F. Lopes, Geraldo A. Silva Filho
and Leonardo Monasterio *

Resumo

Este trabalho estima o impacto da cultura no desempenho acadêmico de estudantes brasileiros em testes padronizados. A partir de dados com identificação por aluno, aplicamos um algoritmo de classificação de ancestralidade que atribui ao aluno, com base nos sobrenomes dos pais, um dos seguintes grupos: ibéricos, japoneses, italianos, germânicos, europeus do leste e sírio-libaneses. Mostramos que os alunos com ancestralidade europeia não-ibérica e japonesa obtêm notas, na prova de Matemática da Avaliação Nacional da Alfabetização e na Prova Brasil, estatística e substantivamente mais elevadas, mesmo com um amplo conjunto de controles individuais, familiares e municipais. Testamos ainda, por meio de proxies, a hipótese de persistência das instituições locais, influenciadas pela imigração em massa no Brasil no século XIX e XX, e mostramos que os mecanismos de transmissão familiar da cultura permanecem robustos para os alunos com ancestralidades japonesa ou italiana.

Palavras-chaves: avaliação educacional, imigração, cultura, ancestralidade.

Área Anpec: Área 12 - Economia Social e Demografia Econômica.

Jel codes: I2, F22

Abstract

This paper estimates the impact of culture on the academic performance of Brazilian students in standardized tests. Based on data with student identification, we apply an algorithm of surname classification that assigns the student, based on the surnames of his/her parents, to one of the following ancestry groups: Iberian, Japanese, Italian, Germanic, Eastern European and Syrian-Lebanese. We show that students with non-Iberian and Japanese European ancestry obtain statistically and substantively higher scores on 3rd and 5th grade standard Math tests, even with a large set of individual, family and municipal controls. We also tested the hypothesis of persistence of local institutions, established during the era of mass immigration in Brazil in the 19th and 20th centuries, and we showed that the mechanisms of family transmission of culture remain robust for students with Japanese and Italian ancestry.

Keywords: educational evaluation, immigration, culture, ancestry.

*Lopes: Universidade Católica de Brasília (UCB); Silva Filho: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP); Monasterio: Instituto de Pesquisa Econômica Aplicada (IPEA) and UCB.

1 Introduction

Can culture explain the diversity of societal outcomes as opposed to institutions and economic factors? This paper follows the literature on cultural transmission and explores the importance of a distinct cultural trait transmitted from parents to children as an alternative and complementary source of persistence in educational outcomes across generations. Several studies find a persistent correlation between individual educational achievement and family socioeconomic background (BLACK; DEVEREUX; SALVANES, 2005; JR.; LEVITT, 2004; HERTZ et al., 2008). This research has found at most moderately-sized (and often zero) causal effects, suggesting that much of the correlation between parents' and children's educational outcomes must be due to other family characteristics, including access to high quality schools (ROUSE; BARROW, 2006), or inherited abilities and traits (KRAPOHL et al., 2014).

We will be primarily interested in differences in culture which we define, for our purposes, as systematic differences in preferences and beliefs across either socially or spatially differentiated groups. Parents transmit to their children not only human capital, income, wealth, and genetic traits but also a specific set of cultural values (BISIN; VERDIER, 2000). Vertical transmission denotes transmission within the family – its intensity may be purposefully chosen by parents, and tends to induce persistence. Horizontal transmission refers to the mechanism through which values and beliefs are transmitted via social interactions with peers or with adults external to the family.

In this paper, we study the educational outcomes of higher than second-generation descendants[1] of immigrants.These children, born and raised in Brazil, face the institutions but they potentially differ in their cultural heritage as reflected in their ancestors' country of origin. The major concern our analysis needs to address is how to separate the effects of culture from the effects of strictly economic factors and institutions. To circumvent this problem we exploit the difference in the "portability" of culture relative to economic and institutional conditions. When individuals emigrate, they may take some aspects of their culture with them and transmit them intergenerationally, while living in the economic and formal institutional environment of the host country.This suggests that studying immigrants or their descendants may be a useful strategy for isolating some aspects of culture. Relying on the experience of descendants of immigrants from various countries of ancestry who live in the same country – in order to assess the effect of culture on economic or other social outcomes – is often referred to as the *epidemiological approach* to the analysis of culture.

Following Fernández & Fogli (2009), Figlio et al. (2016) and SØrensen et al. (2016), our identification strategy relies on the opportunity to observe children whose ancestors migrated to Brazil in the same location (same classroom), thus distinguishing between the cultural factors from other institutional and economic factors. We explore the educational outcomes of higher than second-generation Brazilian students in a dataset that contain individual-level administrative data from the *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP) on 3rd and 5th graders. First, this paper presents the novel use of administrative data to study the relevance of cultural traits in explaining educational outcomes of descendants of immigrants in the very long run. INEP data allow us to observe the entire population of public school students, and to control for classroom and school fixed effects and several socioeconomic characteristics. Second, we are able to follow these students over time during their primary education years, measuring not only their educational achievement at one point in time, but also the change over time. Cross-

[1] Studying high-order generation immigrants rather than immigrants has the benefit of minimizing shocks to "normal" behavior resulting from immigration (e.g., language barriers). On the other hand, the fact that they are second or higher-order generation descendants might weaken the strength of cultural effects on economic actions since cultural transmission is restricted mostly to parents and ethnic social networks rather than operating in society at large.

section comparisons can confound cultural transmission with unmeasured shared correlates between parents and children, but longitudinal data permit the opportunity to explore both levels and trajectories of outcomes. To analyze the importance of culture, we link each student within groups of countries based on her forebears' country of origin.

Another main problem our analysis has to deal with is how to disentangle the effects of culture from those of local institutions[2]. Immigrants might have influenced the design and evolution of local institutions, as suggested by Kreutz (2000). Hence, cultural transmission might be occurring horizontally. We explore data on colony settlement geographical distribution in Rio Grande do Sul and São Paulo as instrument for local institutions. That is, if the human capital of one's ancestry group is an important input in the formation of one's own human capital (as argued by Borjas, Bronars & Trejo (1992), Borjas (1995), then systematic differences in human capital across ancestry groups may be responsible for our results.

2 Literature review

Our paper speaks to three sets of literature. Besides being related to the literature on cultural transmission (ALESINA; GIULIANO; NUNN, 2013; ALESINA; GIULIANO, 2015; ALGAN; CAHUC, 2010; GALOR; MOAV, 2002; GALOR; MICHALOPOULOS, 2012; GUISO; SAPIENZA; ZINGALES, 2006; NUNN; WANTCHEKON, 2011; SACERDOTE, 2005; TABELLINI, 2008; VOIGTLÄNDER; VOTH, 2012), our paper relates to the inter-generational mobility literature and to the research on immigrants' assimilation. Chetty, Hendren & Katz (2016) find that local conditions matter less for immigrants consistently with the conjecture that culture, rather than neighborhood's characteristics, can play an important role for immigrants. The literature on immigrants has systematically identified an "advantage" of some immigrant groups but, as far as we know, no paper has identified which cultural factors may be responsible for these findings in the very long run (CARD; DINARDO; ESTES, 1998; ABRAMITZKY; BOUSTAN; ERIKSSON, 2014).

This study also relates to the literature attempting to identify the long-term, slow-moving determinants of economic and social outcomes, and how they are shaped by endogenous processes and historical accidents (ENGERMAN; SOKOLOFF, 1997; ACEMOGLU; JOHNSON; ROBINSON, 2002; NUNN, 2008). It also relates to a growing literature on the historical determinants of local institutions and economic performance in Brazil. Naritomi, Soares & Assunção (2012) study the colonial origins of institutions in Brazil looking back to the sugar-cane and gold cycles of the seventeenth and eighteenth centuries. Musacchio, Martinez & Viarengo (2014) study the political economy of education in Brazil during the period of increased decentralization of revenues after the proclamation of the Republic. Carvalho Filho & Colistete (2010) study the connections between the coffee cycle, European immigration of farm laborers and the establishment of public instruction institutions in the state of São Paulo. Rocha, Ferraz & Soares (forthcoming) also analyzed the case of the colonies in São Paulo and found evidence suggesting the role of human capital of immigrants in the long-term development of such areas. Carvalho Filho & Monasterio (2012) study the long run impact of state-sponsored settlements of European immigrants in Southern Brazil finding evidence for cultural spillovers.

[2] Despite accounting for both vertical and horizontal transmissions of culture the issue of whether omitted variable bias remains.

3 Data

3.1 Ancestry: immigration, ancestry and surnames in Brazil

Since ancestry data are not directly available, we link school records using family names, henceforth surnames, to surname-ancestry identification algorithm proposed by Monasterio (2017).

Brazilian censuses and administrative data (such as School Census and Annual Social Information Report - *Relação Anual de Informações Sociais*, RAIS) do not record information on the people's ancestry. There are just five categories of color/race, namely, white (*branco*), mixed (*pardo*), black (*preto*), far-eastern (*amarelo for the Chinese and Japanese*) and native-Brazilian (i.e. Amerindians, *índio*). [3]

INEP databases provide information on the color/race for students, but not for their parents. Furthermore, white (Iberian), mixed and black Brazilians do not have distinct surnames that would make it feasible to identify their ancestry based only by piece of information. Therefore we had to focus this paper on the students identified as white, mixed and far-eastern as a way to overcome this limitation of the dataset.

3.1.1 Immigration to Brazil

Non-Iberian immigration to Brazil was promoted by the government when it became clear that the days of legal slavery were counted. There were experiments of subsidized immigration starting on the first half of the 19^{th} century, but the main flows were concentrated after the emancipation of slaves in 1888. Between 1872 and 1920 more than 3,2 million foreigners arrived in Brazil (LEVY, 1974). In 1920, 5.1% of the population were foreigners or naturalized. From then on, Brazilian legislation and public policies closed the country to immigration. In 2010, just 0.3% of the population were born outside the Brazil.

Subsidized immigration was a positive human capital shock. Educational levels were really low in Brazil, even compared to other poor countries (CHAUDHARY et al., 2012). In 1920, only 23% of Brazilian of all ages could read of write. In general, immigrants were not high qualified in their home countries, but their human capital endowments were probably higher than the Brazilians: 52% of foreigners were literate (DGE, 1929).

Were there differences in human capital or culture between immigrant groups? If so, how to measure these differences? We resort to method proposed by A'Hearn, Baten & Crayen (2009). It is based on a simple idea: individuals with low quantitative skills tend to round their ages in numbers ending in 5 or 0. Therefore the distribution of the last digit of age will be skewed in favor of these numbers in groups with low numeracy skills. The ABCC index is a measure of the degree of the (inverse) of age heaping for groups. Its interpretation is straightforward: it varies from 0 to 100 and the higher the value, the higher the numeracy of the group (CRAYEN; BATEN, 2010).

Stolz, Baten & Botelho (2013) have estimated ABCC for immigrants using data from arrivals in Rio de Janeiro. Their estimates suggests that numeracy skills of non-Iberian European immigrants were much higher than the Brazilians. We follow the same method using data for individuals born between 1860 and 1920 registered by the *Hospedaria dos Imigrantes de São Paulo* (São Paulo Immigrant Hostel). This data includes internal migrants in Brazil. The results are shown in the first column of Table 1. They result suggest that migrant Brazilians had a numeracy level much lower than the foreign immigrants, even lower than the ones that came from Portugal

[3] See Piza & Rosemberg (1999) and Osório (2004) for discussions on color and race in the Brazilian official surveys.

and Spain. It is remarkable that the Japanese immigrants reached the highest scores.[4]

The second column of the table shows the average unweighted ABCC index for the ancestries on their country of origin, ie. non-migrants. We have averaged the ABCC indexes for countries available at Clio Infra between 1880 and 1920. The result sort of confirms our estimates: Brazilian residents still have the lower numeracy levels that the European and Japanese residents. The Syrian-Lebanese residents had astonishing low levels of numeracy.

Table 1: ABCC Index for ancestry groups for migrants and residents

	Ancestry	Migrants	Population
1	BRA	87.65	88.69
2	EAS	97.82	98.47
3	GER	97.18	99.93
4	IBR	96.24	95.78
5	ITA	98.09	99.94
6	JPN	100.00	99.46
7	SLB	NA	68.57

Note: "Migrants"= ABCC index for immigrants in the São Paulo Immigrant Hostel born between 1860 and 1920. "Population"= average of ABCC index for residents in selected years between 1880 and 1920.

3.1.2 Surname classification

Most children in Brazil have two surnames inherited from the mother and the father, in that order. Ideally the full names of both parents of the student would give us the ancestry of his four grandparents. But this is rarely the case. First of all, people may or may not adopt the surname of the other party but traditionally women drop the surname of the mother and add that of the husband in the surnames. As long as the maiden name of the mother is not available, we loose her matrilineal ancestry. In fact, we had to deal with multiple possibilities.

The classification of surnames by ancestry followed the same steps as Monasterio (2017) with some minor differences. The main idea is to build - from historical and contemporary sources - a reference table that links each unique surname to its place of origin. This reference table plays a role in the two main steps of surname classification:

1. Surnames on the target dataset were fuzzy matched to a database of contemporary surnames. We have used a quite conservative fuzzy matching criterion: Optimal String Alignment (OSA) which allows matching when the distance between the surnames are bellow 2. Distance refers to the number of changes that are needed for the strings to become identical.

2. Surnames that failed to be matched to a nationality in the first step were classified following the method suggested by Cavnar & Trenkle (1994). This is a very simple and fast method for word classification, especially when compared to other machine learning algorithms (MONASTERIO, 2017). The intuition is quite simple. Surnames in the reference table are broken into n-grams and frequency profiles of for each ancestry are created according to the frequency of n-grams.[5] Then each *new* surname is also broken in n-grams and matched to ancestry.

[4] ABCC index for SLB ancestry is non-available because our sample was too small

[5] N-grams are sequences of characters in a word. For instance, the surname SILVA has the following 3-grams: "SIL", "ILV" and "LVA".

Our reference table has 74,608 unique surnames linked to the the following ancestries: Iberian (i.e. Portuguese and Spanish; IBR; 10,204 surnames); German (GER; 22,295); Italian (ITA; 26,153); Japanese (JPN; 5,359); Eastern European (EAS; 7,717); and Syrian-Lebanese (SLB; 2,880)[6]. This covers the main ethnic groups that have arrived in Brazil during the age of migrations.

The 2013 Annual Social Information Report (Relação Anual de Informações Sociais), a restricted-access administrative file that contains 46.8 million observations of workers in the formal labor market, was our source of contemporary surnames. In RAIS there are 530,876 unique surnames. The ancestry algorithm of surname classification outlined above was then applied to this dataset. Finally, these pairs of surnames-ancestry were merged to the surnames of the students and their parents in the INEP dataset.

3.2 Education

INEP administers the Brazilian national basic education evaluation system (known by its Portuguese acronym, *SAEB*) that entails a biannual standardized proficiency test on Math and reading for students from the 3rd, 5th, and 9th grades[7]. All students from public schools that have at least 10 students in the 3rd grade and 20 students in the 5th grade do the exams. For the 3rd graders, exams (*ANA*) are taken in only one discipline, Math or reading. Pupils are randomly assigned to Math or reading test. Meanwhile 5th graders take *Prova Brasil* exams of both disciplines. This exam data also includes contextual information obtained from 5th graders' questionnaire, from where we get students socioeconomic characteristics and some of theirs family rearing practices. Although *ANA* and *Prova Brasil* grades are not in the same proficiency scale, it is arguable that they are comparable since their contents are quite similar.

The *ANA* exam is the first nationwide standardized test taken by Brazilian public schools students, when they are at the end of the 3rd grade. According to the Brazilian School Census, in 2013, there were 2,758,968 students in the 3rd grade enrolled in public schools, corresponding to 84.2% of the total of all 3rd graders. Of these students, 1,131,566 took the Reading test, and 1,127,076 the Math test, with 41.0% and 40.9% of coverage, respectively.

Despite the lack of a unified unique identification number for individuals in Brazil, since 2008 every student enrolled in basic education has a unique identifying code for School Census administered by INEP[8]. Every school system has to inform yearly their students enrollment status. Consequently we managed to follow the cohort of students who were enrolled in the 3rd grade in 2013 academic year back and forth, from daycare in 2008 or school entrance between 2008 and 2011 to the 2015 academic year. The students who had progressed from the 3rd grade without repetition or dropout reached the 5th grade in 2015.

Using the individual identifying number, we merged proficiency tests microdata with school census microdata at the individual level. From school censuses we get 3rd grade students' surnames,

---

6 Monasterio (2017) does not identify Syrian-Lebanese surnames.

7 The national basic education evaluation system, *SAEB* (*Sistema de Avaliação da Educação Básica*), is composed by three standardized proficiency tests: (i) the National School Performance Assessment (*Anresc*, also known as *Prova Brasil*), taken by public schools 5th and 9th graders; (ii) the National Evaluation of Basic Education (*Aneb*) that includes a representative sample of private schools students and smaller public schools students from the 5th, 9th and 12th grades; and, more recently, (iii) the National Assessment of Literacy (*Avaliação Nacional da Alfabetização, ANA*), taken by public schools 3rd graders. The exams are taken only in odd academic years. We get data from 2013 *Avaliação Nacional da Alfabetização - ANA*, the first edition of this exam, and 2015 *Prova Brasil.*

8 INEP databases containing identification of individuals, including their parents and own names can be obtained through the Research Support Service (SAP). However, researchers have to work with data inside INEP dependencies. Further information on how to request access to the databases can be obtained at http://portal.inep.gov.br/solicitacao-de-acesso

demographic characteristics, and schooling trajectory from their entrance into Brazilian school system, whether in daycare, kindergarten or 1^{st} grade, to their 5^{th} grade in the 2015 academic year. Since census data contains student's birth place and municipality of residence in each academic year, the longitudinal data allowed us to recover family geographical mobility history - it was possible to gather information about whether the student lives where he was born and whether one's family moved to another municipality in the period of analysis and when (2008-2015).

Considering the 3^{rd} graders in 2013 who get to 5^{th} grade in 2015, 2,115,542 students were enrolled in public schools and, therefore, would be potential participants of *ANA* and *Prova Brasil* exams, our sources of information on school proficiency. In fact, 1,592,674 of these students took the 5^{th} grade Math exam and 664,108 had also taken the 3^{rd} grade Math exam, corresponding to our sample for the analysis of the gap in terms of gains in proficiency from the 3^{rd} grade to 5^{th} grade between students from different ancestry. However, in order to avoid problems already mentioned in section 3.1 and to make the students of the sample more comparable among each other we restricted the analysis to individuals who self-declared as white, mixed or far-eastern. Then these students were labeled according to their ancestry. The Appendix presents descriptive statistics comparing four groups of students according to their ancestry status: (i) whether having both parents of Iberian ancestry; (ii) having just one of them of non-Iberian ancestry; (iii) having both parents of non-Iberian ancestry, but not the same ancestry; and (iv) having both parents of the same ancestry.

We study the following two different outcomes: 3^{rd} graders test scores in mathematics and 5^{th} graders test scores in Mathematics. Studying test score growth is especially important because test score levels might reflect some omitted variable correlated with country-of-ancestry, but it is very rare for students to dramatically change their relative position in the nationwide test score distribution between grades 3 and 5.

4 Empirical strategy

As discussed in the introduction, our empirical strategy is to isolate the effects of culture from those of markets and institutions by studying the educational outcomes of children who were mostly born in, and reside in, Brazil, but whose ancestors were born in another country.

To investigate this hypothesis, we face several challenges, for details see Fernández & Fogli (2009), Guiso, Sapienza & Zingales (2006). First, if parents share a culture of high educational attainment, they are likely to be highly educated and, thus, more likely to have high income and live in areas with better schools, therefore hindering our ability to distinguish between a transmission of cultural values and a direct effect of parental education or income. Therefore, controlling for school and classroom fixed effects, as well as individual characteristics and measures of family income might help us track the direct effect of culture on education. We also control for local geographic variation in markets and institutions by including controls for municipality characteristics and mobility controls - whether the student has moved from its place of origin in 2013 or 2015, and we cluster observations at the classroom level.

As pointed out in Figlio et al. (2016) one may object from the outset that there is no reason to believe immigrants have the preferences/beliefs that are representative of the average in their country of origin. This may or may not be true for different cultural attributes, but in any case, this factor will tend to bias the test of our hypothesis toward not finding any effect of culture on education. More importantly, for our analysis to be meaningful, culture should evolve relatively slowly over the time period in which we are interested. Otherwise, in general, the beliefs transmitted from parents to children would not be captured[9].

All our regressions contain a large set of controls, including demographics (gender and age-grade distortion and special education needs), socioeconomic status *SES* (measured by indicator variables regarding availability of household goods and amenities, and parental education). Because special education and family income are all potential consequences of parental cultural background, we investigate the degree to which our results are driven by the decision of whether or not to control for these variables. In our main specifications, we control for these variables.

4.1 Ancestry Dummies

Before starting our empirical analysis, we first examine whether there exist systematic differences between each educational outcome and culture proxy. We turn to the more traditional approach by using country-of-ancestry dummies rather than the quantitative home country variables as our cultural proxies. This has the benefit of not requiring the relationship between culture and outcomes to be linear in the cultural proxy. Furthermore, it may allow different features of culture to play a role in educational outcomes other than those captured past aggregate cultural proxies. It has, however, the drawback of not specifying how culture matters. We estimate the following model:

$$Y_{iam} = \alpha Z_a + \beta X_i + \gamma_g + \zeta_m + \delta_c + \mu_s + \omega_p + \epsilon_{ic} \quad (1)$$

where Y_{iam} is an outcome of interest for student i coming from ancestry a who resides in municipality m, and Z_a is our measure of proxy for culture measured at the country-of-ancestry level. X_i are time invariant individual controls. Our specification also includes, a dummy indicating whether the student is considered a *Mover*, a continuous variable γ_g reflecting past performance in standardized Math tests, in the outcomes for which this is relevant, a full set of fixed effects to control for specific differences in performance across different classrooms/schools. The standard errors are clustered at the classroom level.

4.2 Variation to Culture Exposure

A potential concern with the OLS estimates reported up to this point is that the culture proxy, assigned based on a student's father surname, could capture some omitted ancestry characteristics, in particular unobserved human capital. Following SØrensen et al. (2016) we exploit the fact that students with parents whose ancestors originated from the same country receive varying exposure to its culture. We compare students with one parent with non-Iberian ancestry, two parents with non-Iberian ancestries but not identical and two non-Iberian parents with identical ancestries, and assume that the cultural impulse is larger with two non-Iberian parents sharing identical ancestries[10]. The identifying assumption is that the assignment of Iberians and non-Iberians to one- and two non-Iberian couples is plausibly random, conditional on country of ancestry and parental characteristics. We then repeat the previous specification replacing Z_a by a set of dummies of culture exposure while keeping our controls unchanged.

[9] Some cultural traits may change faster (e.g eating habits) while other (e.g fertility decisions and living arrangements) may evolve at a much slower pace

[10] It is still unclear why and how having parents sharing same set of values,norms and beliefs would, hypothetically, render better outcomes. We aim to further investigate the interaction between ancestry and culture exposure in future research

4.3 Institutions and Horizontal Cultural Transmission

Transmission through family ties – ie. vertical transmission – is only one of the possibilities that values and beliefs of individuals are passed to each other. Social interactions outside the nuclear family and institutions (in the broad sense) transmit culture as well. Immigration may have impacted not only the ancestry of Brazilians but also its local institutions and therefore individual test scores.

There is a vast literature on the long term impact of historical events (GUISO; SAPIENZA; ZINGALES, 2006; ALESINA; GIULIANO; NUNN, 2013; NUNN, 2008; TABELLINI, 2008; ALGAN; CAHUC, 2010). According to this line of research, fortuitous or exogenous facts or decisions show their local impact in the very long term. In the Brazilian case, the initial location of the newly arrived may have caused changes in local institutions that persisted and show up in test scores.

Carvalho Filho & Colistete (2010), Carvalho Filho & Monasterio (2012), and Rocha, Ferraz & Soares (forthcoming) study the long term impact of immigration to Brazil on contemporary data. The first paper asserts that foreign-born immigrants caused an increase in the supply of local public instruction and this results in higher average test scores (and higher income per capita) at the municipal level. The second paper focus on the long-term impact of subsidized immigration to Rio Grande do Sul. Their results suggest that municipalities closer to official settlements have better socioeconomic outcomes in 2000. Finally, the third paper shows that in municipalities in São Paulo that received state sponsored settlements still have higher levels of schooling and income per capita.

Here we use the data provided by Carvalho Filho & Monasterio (2012), and Rocha, Ferraz & Soares (forthcoming) in order to disentangle the channels that relate immigration to test scores. Both papers suggest that locations of colonies were exogenous to contemporary output variables of interest. We will use the location of state sponsored settlements in Rio Grande do Sul and São Paulo as a way to control for the impact of immigrant institutions on individual test scores. Figure 1 shows the minimum distance of the centroids of municipalities of these States to any official colony.

More precisely, we will use the 'epidemiological' approach: we will look at students that were born in one municipality and go to school in another municipality as a way to identify vertical (family) and horizontal (institutional) transmission of culture.

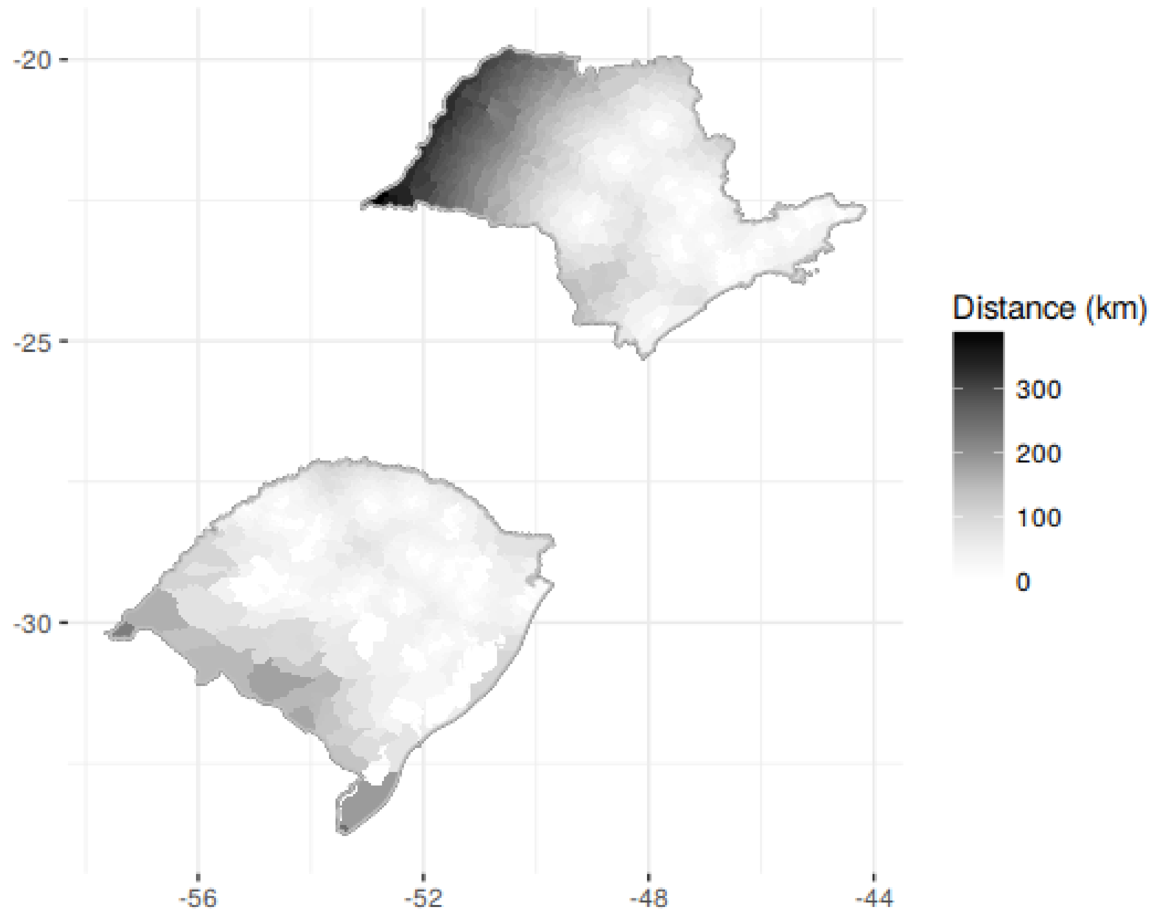
We test the following specification in order to estimate the effects of institutions and culture on student achievement:

$$Y_i = \alpha + \mu Movers_i + \sigma Colony_i + \gamma Ancestry_i + \beta_1 Movers_i.Colony_i + \beta_2 Movers_i.Ancestry_i + \beta_3 Colony_i.Ancestry_i + \beta_4 Movers_i.Colony_i.Ancestry_i + \epsilon_i \quad (2)$$

where Y_i is the score for student i; $Movers_c$ is a dummy for students that study in a municipality different from their place of birth; $Ancestry_i$ is our measure of surname ancestry; and $Colony_i$ is the shortest euclidean distance to an state-sponsored settlement and, alternatively, a dummy for these colonies. For the sake of simplicity, individual and municipality controls were omitted from equation (2).

Our main interest lies in the coefficient γ, because it captures the "pure" effect of family. σ identifies the effect of living in (or close to) a former colony, while μ the effect of moving. Coefficients associated with interaction terms are useful to identify the roles of foreign ancestry and institutions. $\gamma + \beta_2$ is the effect on scores for students with non-Iberian ancestry that were

Figure 1: Minimum distance to official settlements in Rio Grande do Sul and São Paulo



Source: Carvalho Filho & Monasterio (2012) for Rio Grande do Sul; Rocha, Ferraz & Soares (forthcoming) for São Paulo *Note*: axes indicate latitude and longitude.

not born in a Colony but have moved. $\gamma + \beta_3$ the impact for those who where born in a colony and not moved. Finally, $\gamma + \beta_1 + \beta_4$ shows the effect on scores for those students with non-Iberian ancestry that were born in a colony and have moved.

5 Results

5.1 Individual Results

Panel A of Table 2 contains the results from the estimation of (1) by OLS with clustered standard errors. The dependent variable is the individual Math score. Models 1 to 3 have fixed effects at the class level. Model 4 has controls for each school just to show that results are robust. Model 5 has fixed effects for each pair of place of birth and place of residence for the students.[11]

Model 1 shows that estimated coefficients for non-Iberian surname ancestry have statistical and substantive significance. Students with Japanese surname ancestry get 22 % standard deviations higher than their peers with Iberian surnames.

[11] The number of observations change because individual controls are available only for those students that took Prova Brasil and answered the socioeconomic questionnaire.

Table 2: 3rd and 5th graders Math scores and Student's ancestry

Panel (A)	Dependent Variable: 3rd Math test scores				
Student's Ancestry	(1)	(2)	(3)	(4)	(5)
Japanese	0.216***	0.171***	0.159***	0.160***	0.198***
	(0.0207)	(0.0225)	(0.0288)	(0.0261)	(0.0241)
Germanic	0.157***	0.102***	0.101***	0.102***	0.102***
	(0.00821)	(0.00889)	(0.0114)	(0.0101)	(0.00926)
Italian	0.121***	0.0791***	0.0780***	0.0791***	0.0869***
	(0.00541)	(0.00591)	(0.00750)	(0.00669)	(0.00616)
Eastern European	0.146***	0.101***	0.0915***	0.0902***	0.101***
	(0.0139)	(0.0152)	(0.0184)	(0.0164)	(0.0155)
Syrian-Lebanese	0.0494***	0.0160	0.0199	0.0158	0.00354
	(0.0165)	(0.0186)	(0.0231)	(0.0209)	(0.0195)
Constant	0.309***	0.545***	0.538***	0.480***	0.544***
	(0.000887)	(0.00250)	(0.00319)	(0.00260)	(0.00349)
Observations	274,176	197,527	196,973	318,231	197,527
R-squared	0.356	0.370	0.371	0.256	0.278
Panel (B)	Dependent Variable: 5th graders Math test scores				
ANA 2013 Math scores	0.563***	0.556***	0.556***	0.554***	0.576***
	(0.00244)	(0.00251)	(0.00251)	(0.00184)	(0.00276)
Japanese	0.129***	0.128***	0.139***	0.149***	0.151***
	(0.0164)	(0.0165)	(0.0222)	(0.0200)	(0.0176)
Germanic	0.0568***	0.0546***	0.0567***	0.0587***	0.0583***
	(0.00634)	(0.00636)	(0.00814)	(0.00742)	(0.00678)
Italian	0.0428***	0.0405***	0.0353***	0.0406***	0.0439***
	(0.00419)	(0.00422)	(0.00540)	(0.00490)	(0.00447)
Eastern European	0.0467***	0.0399***	0.0407***	0.0423***	0.0407***
	(0.0110)	(0.0111)	(0.0139)	(0.0128)	(0.0114)
Syrian-Lebanese	0.0320**	0.0278**	0.0258	0.0281*	0.0280**
	(0.0129)	(0.0132)	(0.0171)	(0.0156)	(0.0139)
Constant	-0.572***	-0.513***	-0.520***	-0.537***	-0.525***
	(0.00140)	(0.00223)	(0.00267)	(0.00211)	(0.00297)
Observations	229,314	222,792	222,161	345,711	222,670
R-squared	0.562	0.564	0.564	0.512	0.539
FE	Class	Class	Class	School	O/R
F test	0	0	0	0	0
Individual controls		Yes	Yes	Yes	Yes
Family SES		Yes	Yes	Yes	Yes
Municipality controls		Yes	Yes	Yes	
Mobility controls			Yes	Yes	

Robust standard errors clustered at classroom level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
O/R: Pair Place of Origin/ Place of Residence Fixed Effects

As expected, the inclusion of controls for individual characteristics, family socioeconomic status and municipality controls [12] reduces the size of the coefficients; Syrian-Lebanese surnames lose statistical significance. We must stress the result in Model 5. We are comparing students that not only share the same observable individual and family characteristics, but were born and study in the same pair of municipalities. Even in that full specification the coefficients for surname ancestry remain stable and significant.

Models 1 to 5 in Panel B of 2 have the same controls of Panel A with two major differences: dependent variable is the 5th grader Math score and we are controlling for the Math score that the same student got 2 year earlier (ANA 2013 Math score). Estimated coefficients associated with ancestry variables retain the general pattern. Interestingly, coefficients associated with Germanic, Italian, and Eastern European ancestry have dropped. However coefficients for Japanese have grown ever more and for Syrian-Lebanese have become statistically significant. The fact that ancestry dummies remain significant after the inclusion of past scores suggest that the effects of ancestry take a while to have full effect.

Table 3: 3^{rd} and 5^{th} graders Math scores and Student Variation in Culture Exposure

Panel (A)	Dependent Variable: 3^{rd} graders Math test scores				
Culture Exposure	(1)	(2)	(3)	(4)	(5)
One Ancestry	0.104***	0.0684***	0.0668***	0.0652***	0.0818***
	(0.00440)	(0.00484)	(0.00605)	(0.00488)	(0.00502)
Two Ancestries (not same)	0.182***	0.135***	0.131***	0.125***	0.151***
	(0.0100)	(0.0110)	(0.0136)	(0.0120)	(0.0111)
Two Ancestries (same)	0.254***	0.176***	0.176***	0.176***	0.208***
	(0.0114)	(0.0123)	(0.0157)	(0.0140)	(0.0125)
Constant	0.313***	0.545***	0.538***	0.484***	0.541***
	(0.00126)	(0.00278)	(0.00353)	(0.00277)	(0.00371)
Observations	259,694	188,578	188,042	302,018	188,578
R-squared	0.359	0.374	0.374	0.257	0.279
Panel (B)	Dependent Variable: 5^{th} graders Math test scores				
ANA 2013 Math test scores	0.564***	0.558***	0.557***	0.561***	0.577***
	(0.00250)	(0.00257)	(0.00258)	(0.00246)	(0.00282)
One Ancestry	0.0412***	0.0378***	0.0303***	0.0307***	0.0435***
	(0.00340)	(0.00343)	(0.00437)	(0.00414)	(0.00362)
Two Ancestries (not same)	0.0773***	0.0754***	0.0773***	0.0811***	0.0848***
	(0.00771)	(0.00773)	(0.00971)	(0.00905)	(0.00795)
Two Ancestries (same)	0.102***	0.0969***	0.104***	0.105***	0.115***
	(0.00871)	(0.00873)	(0.0111)	(0.0103)	(0.00895)
Constant	-0.574***	-0.515***	-0.522***	-0.524***	-0.528***
	(0.00160)	(0.00240)	(0.00288)	(0.00292)	(0.00312)
Observations	218,942	212,850	212,245	212,130	212,735
R-squared	0.566	0.568	0.568	0.533	0.539
FE	Class	Class	Class	School	O/R
F test	0	0	0	0	0
Individual controls		Yes	Yes	Yes	Yes
Family SES		Yes	Yes	Yes	Yes
Municipality controls		Yes	Yes	Yes	
Mobility controls			Yes	Yes	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
O/R: Pair Place of Origin/Place of Residence Fixed Effects

[12] Municipality controls include the following set of variables: urban population, workforce, share of workers in the formal sector, share of workers employed with and without regular employment, total population, income per capita, Theil and Gini indexes for income distribution.

Table 3 aims to test if the "dosage" of cultural ancestry impacts the test scores of students. So we repeat the same empirical strategy of 2 but now with three dummy variables: "One Ancestry" when the student has only one parent with non-Iberian surname; "Two Ancestries (same)" when the parents share the same ancestry; and "Two Ancestries (not same)" when the student parents have non-Iberian ancestries but they are different. The reference, obviously, are the students without non-Iberian surnames.

Results on dosage are quite robust. According to Panel A of 3 having one parent with non-Iberian surname is associated with higher scores, but two parents seem to increase the results even more. Besides, when both parents share the same non-Iberian ancestry there is an additional effect. We have no clear hypotheses for the reason of this effect.

All models in Panel B of Table 3 includes the previous test scores of student. Coefficients associated with ancestry dosage drop in value but remain significant. Again, this result indicates that the cultural effects of ancestry on tests scores grow with time.

5.2 Family and Institutions

In this section we check if the impact of family ancestry is robust to the inclusion of variables that control for local institutions. As we have said, we have opted for very simple proxies for non-Iberian institutions: a dummy if the municipality have a ever received a government sponsored settlement and the minimum distance to any of this colonies.[13]

The results in Panel A and B of Table 4 can not compared to the ones in Table 2 because our sample includes only Rio Grande do Sul and São Paulo. Column 1 is now our benchmark specification, ie without controls for institutions. In this restricted sample only the coefficients for Japanese and Italian surname ancestries remain significant. The variable of interest in Models 2-6 is γ of 2; which means the direct impact of the family when we control for all interactions between ancestry, movers and those who were born in a former settlement.

The comparison of the column 1 with 2 and 3; and 4 with 5 and 6 shows some increase in the coefficients associated with Japanese and Italian ancestry. The inclusion of our institutional proxies suggest that vertical transmission of culture remains for these groups even with full controls.

[13] We could not create proxies for distance to each type of colony because Rocha, Ferraz & Soares (forthcoming) do not provide information on the country of origin for all the settlements in their dataset.

Table 4: 3rd graders Math scores, student's ancestry, culture exposure and institutions. 5th graders Math scores, student's ancestry, culture exposure and institutions

Dependent Variable:	3rd graders Math scores			5th graders Math scores		
Panel (A) Student's Ancestry	(1)	(2)	(3)	(4)	(5)	(6)
ANA 2013 Math Scores				0.560***	0.560***	0.560***
				(0.010)	(0.008)	(0.008)
Japanese	0.159**	0.272***	0.241**	0.193***	0.150**	0.207***
	(0.0786)	(0.102)	(0.107)	(0.071)	(0.075)	(0.078)
Germanic	0.0619	0.0959*	0.0348	0.0486	0.0543	0.0740*
	(0.041)	(0.054)	(0.057)	(0.0361	(0.036)	(0.038)
Italian	0.0467**	0.0942***	0.0741**	0.0540***	0.0666***	0.0659***
	(0.0219	(0.029)	(0.030)	(0.020)	(0.020)	(0.021)
Eastern European	0.00849	-0.0382	0.0178	-0.0660	-0.0544	-0.117*
	(0.0759)	(0.105)	(0.113)	(0.0667)	(0.0718)	(0.067)
Syrian-Lebanese	-0.0581	-0.178*	0.0253	0.0010	0.0644	-0.0095
	(0.079)	(0.098)	(0.102)	(0.078)	(0.078)	(0.079)
Constant	-0.287	-0.0127	-0.0733	-0.819*	-0.897**	-0.828**
	(0.476)	(0.537)	(0.537)	(0.422)	(0.358)	(0.358)
Observations	22,332	22,332	22,332	25,930	25,930	25,930
R-squared	0.431	0.591	0.591	0.705	0.705	0.705
Panel (B) Culture Exposure						
ANA 2013 Math Score				0.559***	0.560***	0.560***
				(0.0087)	(0.0087)	(0.0087)
One Ancestry	0.0802***	0.114***	0.0652**	0.0405***	0.0482***	0.0481***
	(0.021)	(0.024)	(0.026)	(0.014)	(0.017)	(0.017)
Two Ancestries (not same)	0.0910*	0.120*	0.0812	0.0522	0.0891**	0.0788*
	(0.052)	(0.062)	(0.061)	(0.035)	(0.044)	(0.042)
Two Ancestries (same)	0.157***	0.184***	0.183***	0.129***	0.132***	0.157***
	(0.049)	(0.059)	(0.059)	(0.033)	(0.038)	(0.040)
Constant	-0.0908	-0.0240	-0.130	-0.876**	-0.954***	-0.936**
	(0.554)	(0.556)	(0.556)	(0.367)	(0.369)	(0.368)
Observations	21,657	21,657	21,657	25,096	25,096	25,096
R-squared	0.597	0.597	0.597	0.707	0.708	0.708
FE	Class	Class	Class	Class	Class	Class
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Family SES	Yes	Yes	Yes	Yes	Yes	Yes
Municipality controls	Yes	Yes	Yes	Yes	Yes	Yes
Mobility controls	Yes	Yes	Yes	Yes	Yes	Yes
F test	0	0	0	0	0	0
Settlement identification		Dummy	Distance		Dummy	Distance

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For the sake of completeness, Panel B of Table 4 repeats the "exposure" estimations of Table 3 including our institutional variables. The ordering of coefficients is the same as before. Controlling for institutions, students with parents with the same non-Iberian ancestry have the highest scores, followed by those with parents with different non-Iberian ancestries, and those with one non-Iberian parent.

6 Final remarks

This paper sheds light on the role of culture in explaining the diversity of educational outcomes among primary school students in Brazil. This is the first attempt, to our knowledge, of using individual-level data to study the impact of distinct cultural traits on academic achievement. Leveraging the capabilities of a surname-based matching algorithm we were able to identify the ancestry of major immigrant groups that arrived in Brazil in the turn of the XIX century and the beginning of the XX century.

In order to isolate the effect of culture from those of institutions and markets we follow two avenues. First, we follow Fernández & Fogli (2009), Figlio et al. (2016), SØrensen et al. (2016) and use the *epidemiological approach* to study higher than second-generation Brazilians. Second, even though the *epidemiological approach* allow us to control for the *Macro* institutional setting, we explore the historical episode of mass migration to disentangle potential persistent effects of culture from those of local institutions, in the lines of Tabellini (2008), Guiso, Sapienza & Zingales (2006) and control for colonial settlements and presence of movers in our data.

Our results suggests that there is an *Ancestry premium* in educational performance. Using Iberian Brazilians as our reference group, we provide evidence that students bearing either non-Iberian europeans (Germanic, Italians, Eastern Europeans) or japanese ancestries achieve substantially higher scores in Mathematics standardized test conducted at 3^{rd} and 5^{rd} grades. Not only it seems to exist a gap, but as the students move up from 3^{rd} and 5^{rd} the gap gets wider. These results are robust to several specifications, including individual and family controls, as well as municipality and mobility controls. It is noteworthy, that the results stands while using classroom fixed effects (i.e, controlling for teacher quality, classroom size and so forth). Implications in terms of social inequality is therefore paramount.

To test the strength of our argument, we also proceed to control for potential interaction between ancestry and local institutions. Performing this analysis for the States which we have data on exact location of colonial settlements we show that, students bearing either japanese or italian ancestries seems to hold the effect of culture, implying a stronger vertical transmission of preferences and beliefs(ie. within family). On the other hand, Germanic and Eastern Europeans descendents exhibit statistically insignificant coefficients for our culture proxy, which might suggest that horizontal rather than vertical transmission is occurring.

Additionally, we explored the role of variation in culture exposure, regardless from which ancestry group the student belongs, and we found statistically significant results pointing to a monotonically increasing differential between those students with one foreign ancestry, two non-identical foreign and two identical ancestries.

We have been able to investigate whether culture matters for educational performance. Future research, might indicate *How* and *Why* it matters.It might be that parenting skills are important, discrimination or unobserved human capital are actually playing along with our culture proxy. Moreover if there is gap between students, there is also a room for policy aiming to mitigate this educational differential, that may impact income distribution and social mobility in the long-run.

Bibliography

ABRAMITZKY, R.; BOUSTAN, L. P.; ERIKSSON, K. A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. *Journal of Political Economy*, v. 122, n. 3, p. 467–506, 2014. Disponível em: <https://doi.org/10.1086/675805>.

ACEMOGLU, D.; JOHNSON, S.; ROBINSON, J. A. Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *The Quarterly Journal of Economics*, v. 117, n. 4, p. 1231–1294, nov. 2002. ISSN 0033-5533, 1531-4650. Disponível em: <http://qje.oxfordjournals.org/content/117/4/1231>.

A'HEARN, B.; BATEN, J.; CRAYEN, D. Quantifying quantitative literacy: age heaping and the history of human capital. *The Journal of Economic History*, Cambridge University Press, v. 69, n. 3, p. 783–808, 2009.

ALESINA, A.; GIULIANO, P. Culture and institutions. *Journal of Economic Literature*, v. 53, n. 4, p. 898–944, December 2015. Disponível em: <http://www.aeaweb.org/articles?id=10.1257-/jel.53.4.898>.

ALESINA, A.; GIULIANO, P.; NUNN, N. On the origins of gender roles: Women and the plough *. *The Quarterly Journal of Economics*, v. 128, n. 2, p. 469–530, 2013. Disponível em: <+ http://dx.doi.org/10.1093/qje/qjt005>.

ALGAN, Y.; CAHUC, P. Inherited trust and growth. *American Economic Review*, v. 100, n. 5, p. 2060–92, December 2010. Disponível em: <http://www.aeaweb.org/articles?id=10.1257/aer-.100.5.2060>.

BISIN, A.; VERDIER, T. "beyond the melting pot": Cultural transmission, marriage, and the evolution of ethnic and religious traits*. *The Quarterly Journal of Economics*, v. 115, n. 3, p. 955–988, 2000. Disponível em: <+ http://dx.doi.org/10.1162/003355300554953>.

BLACK, S. E.; DEVEREUX, P. J.; SALVANES, K. G. The more the merrier? the effect of family size and birth order on children's education*. *The Quarterly Journal of Economics*, v. 120, n. 2, p. 669–700, 2005. Disponível em: <+ http://dx.doi.org/10.1093/qje/120.2.669>.

BORJAS, G. J. Ethnicity, neighborhoods, and human-capital externalities. *The American Economic Review*, American Economic Association, v. 85, n. 3, p. 365–390, 1995. Disponível em: <http://www.jstor.org/stable/2118179>.

BORJAS, G. J.; BRONARS, S. G.; TREJO, S. J. Self-selection and internal migration in the united states. *Journal of Urban Economics*, v. 32, n. 2, p. 159 – 185, 1992. ISSN 0094-1190. Disponível em: <http://www.sciencedirect.com/science/article/pii/0094119092900034>.

CARD, D.; DINARDO, J.; ESTES, E. *The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s*. [S.l.], 1998. (Working Paper Series, 6519). Disponível em: <http://www.nber.org/papers/w6519>.

CARVALHO FILHO, I. de; COLISTETE, R. P. MPRA Paper, *Education Performance: Was It All Determined 100 Years Ago? Evidence From São Paulo, Brazil*. 2010. Disponível em: <https://mpra.ub.uni-muenchen.de/24494/>.

CARVALHO FILHO, I. de; MONASTERIO, L. Immigration and the origins of regional inequality: Government-sponsored European migration to southern Brazil before World War I. *Regional Science and Urban Economics*, v. 42, n. 5, p. 794–807, 2012. ISSN 0166-0462. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0166046211000962>.

CAVNAR, W. B.; TRENKLE, J. M. N-gram-based text categorization. *Ann Arbor MI*, v. 48113, n. 2, p. 161–175, 1994. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10-.1.1.21.324rep=rep1type=pdf>.

CHAUDHARY, L. et al. Big BRICs, weak foundations: The beginning of public elementary education in Brazil, Russia, India, and China. *Explorations in Economic History*, Elsevier, v. 49, n. 2, p. 221–240, 2012.

CHETTY, R.; HENDREN, N.; KATZ, L. F. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, v. 106, n. 4, p. 855–902, April 2016. Disponível em: <http://www.aeaweb.org/articles?id=10-.1257/aer.20150572>.

CRAYEN, D.; BATEN, J. Global trends in numeracy 1820–1949 and its implications for long-term growth. *Explorations in Economic History*, Elsevier, v. 47, n. 1, p. 82–99, 2010.

DGE. *Recenseamento do Brazil*. Typ. da Estatística, 1929. Disponível em: <http://servicodados-.ibge.gov.br/Download/Download.ashx?http=1u=biblioteca.ibge.gov.br/visualizacao/livros-/liv31687.pdf>.

ENGERMAN, S.; SOKOLOFF, K. Factor endowments, institutions, and differential paths of growth among new world economies: a view from economic historians of the United States. In: *How Latin America Fell Behind*. Stanford CA: Stanford University Press, 1997.

FERNÁNDEZ, R.; FOGLI, A. Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics*, v. 1, n. 1, p. 146–77, January 2009. Disponível em: <http://www.aeaweb.org/articles?id=10.1257/mac.1.1.146>.

FIGLIO, D. et al. *Long-Term Orientation and Educational Performance*. [S.l.], 2016. (Working Paper Series, 22541). Disponível em: <http://www.nber.org/papers/w22541>.

GALOR, O.; MICHALOPOULOS, S. Evolution and the growth process: Natural selection of entrepreneurial traits. *Journal of Economic Theory*, v. 147, n. 2, p. 759 – 780, 2012. ISSN 0022-0531. Issue in honor of David Cass. Disponível em: <http://www.sciencedirect.com/science-/article/pii/S0022053111000573>.

GALOR, O.; MOAV, O. Natural selection and the origin of economic growth*. *The Quarterly Journal of Economics*, v. 117, n. 4, p. 1133–1191, 2002. Disponível em: <+ http://dx.doi.org/10.1162/003355302320935007>.

GUISO, L.; SAPIENZA, P.; ZINGALES, L. Does culture affect economic outcomes? *Journal of Economic Perspectives*, v. 20, n. 2, p. 23–48, June 2006. Disponível em: <http://www.aeaweb-.org/articles?id=10.1257/jep.20.2.23>.

HERTZ, T. et al. The inheritance of educational inequality: International comparisons and fifty-year trends. *The B.E. Journal of Economic Analysis & Policy*, v. 7, n. 2, p. 1–48, 2008. Disponível em: <http://EconPapers.repec.org/RePEc:bpj:bejeap:v:7:y:2008:i:2:n:10>.

JR., R. G. F.; LEVITT, S. D. Understanding the black-white test score gap in the first two years of school. *The Review of Economics and Statistics*, v. 86, n. 2, p. 447–464, 2004. Disponível em: <http://dx.doi.org/10.1162/003465304323031049>.

KRAPOHL, E. et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences*, v. 111, n. 42, p. 15273–15278, 2014. Disponível em: <http://www.pnas.org/content/111/42/15273-.abstract>.

KREUTZ, L. A. Escolas comunitÃ¡rias de imigrantes no Brasil: instÃde coordenaÃ§Ãe estruturas de apoio. *Revista Brasileira de EducaÃ§Ã*, scielo, p. 159 – 176, 12 2000. ISSN 1413-2478. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttextpid=S1413-24782000000300010nrm=iso>.

LEVY, M. S. F. O papel da migração internacional na evolução da população brasileira (1872 a 1972). *Revista de Saúde Pública*, SciELO Brasil, v. 8, n. supl., p. 49–90, 1974.

MONASTERIO, L. Surnames and ancestry in brazil. *PloS ONE*, Public Library of Science, v. 12, n. 5, p. e0176890, 2017.

MUSACCHIO, A.; MARTINEZ, A.; VIARENGO, M. *Colonial Institutions, Commodity Booms, and the Diffusion of Elementary Education in Brazil, 1889-1930*. [S.l.], 2014. (Working Paper Series, 20029). Disponível em: <http://www.nber.org/papers/w20029>.

NARITOMI, J.; SOARES, R. R.; ASSUNçãO, J. J. Institutional Development and Colonial Heritage within Brazil. *The Journal of Economic History*, v. 72, n. 02, p. 393–422, 2012.

NUNN, N. The Long-Term Effects of Africa's Slave Trades. *Quarterly Journal of Economics*, v. 123, n. 1, p. 139–176, 2008. Disponível em: <http://dash.harvard.edu/handle/1/3710252>.

NUNN, N.; WANTCHEKON, L. The slave trade and the origins of mistrust in africa. *American Economic Review*, v. 101, n. 7, p. 3221–52, December 2011. Disponível em: <http://www.aeaweb.org/articles?id=10.1257/aer.101.7.3221>.

OSÓRIO, R. G. O sistema classificatório de cor ou raça do IBGE. In: Joaze Bernardino; Daniela Galdino (Ed.). *Levando a raça a sério: ação afirmativa e a universidade*. Rio de Janeiro: LPP, UERJ,, 2004. p. 85–135.

PIZA, E.; ROSEMBERG, F. Cor nos censos brasileiros. *Revista USP*, v. 40, p. 122–137, 1999.

ROCHA, R.; FERRAZ, C.; SOARES, R. Human Capital Persistence and Development. *American Economic Review: Applied Economics*, forthcoming. Disponível em: <http://bibliotecadigital-.fgv.br/ocs/index.php/sbe/EBE10/paper/download/2328/1133>.

ROUSE, C. E.; BARROW, L. Us elementary and secondary schools: equalizing opportunity or replicating the status quo? *The Future of Children*, Woodrow Wilson School of Public and International Affairs at Princeton University and the Brookings Institution, p. 99–123, 2006.

SACERDOTE, B. Slavery and the intergenerational transmission of human capital. *The Review of Economics and Statistics*, v. 87, n. 2, p. 217–234, 2005. Disponível em: <http://dx.doi.org/10-.1162/0034653053970230>.

SØRENSEN, R. J. et al. *Culture and school performance:Evidence from second generation immigrants to Norway*. [S.l.], 2016. Disponível em: <http://home.bi.no/fag89001/Culture.pdf>.

STOLZ, Y.; BATEN, J.; BOTELHO, T. Growth effects of nineteenth-century mass migrations: "Fome Zero" for Brazil? *European Review of Economic History*, Oxford University Press, v. 17, n. 1, p. 95–121, 2013.

TABELLINI, G. Presidential address institutions and culture. *Journal of the European Economic Association*, Blackwell Publishing Ltd, v. 6, n. 2-3, p. 255–294, 2008. ISSN 1542-4774. Disponível em: <http://dx.doi.org/10.1162/JEEA.2008.6.2-3.255>.

VOIGTLÄNDER, N.; VOTH, H.-J. Persecution perpetuated: The medieval origins of anti-semitic violence in nazi germany*. *The Quarterly Journal of Economics*, v. 127, n. 3, p. 1339–1392, 2012. Disponível em: <+ http://dx.doi.org/10.1093/qje/qjs019>.

Table A1: Appendix: Descriptive Statistics

	Iberian Ancestry Only			One non-Iberian Ancestry			Two non-Iberian Ancestries			Two non-Iberian Ancestries (same)		
Variables	mean	sd	n.obs	mean	sd	n.obs	mean	sd	n.obs	mean	sd	n.obs
Individual characteristics:												
Female	0,465	0,499	2.080.736	0,478	0,5	236.126	0,488	0,5	18.676	0,477	0,499	26.677
Age-grade distortion at 5th grade (years)	-0,31	1	1.332.717	-0,134	0,867	187.114	0,019	0,665	17.230	-0,072	0,815	22.650
Geographic distribution:												
State PR	0,032	0,176	2.080.736	0,122	0,328	236.126	0,156	0,363	18.676	0,195	0,396	26.677
State RS	0,025	0,157	2.080.736	0,105	0,306	236.126	0,240	0,427	18.676	0,187	0,390	26.677
State SC	0,013	0,113	2.080.736	0,081	0,273	236.126	0,231	0,422	18.676	0,197	0,398	26.677
State SP	0,154	0,361	2.080.736	0,240	0,427	236.126	0,220	0,415	18.676	0,154	0,361	26.677
Brazil (Excluding, PR,RS,SC and SP)	0,775	0,417	2.080.736	0,452	0,498	236.126	0,153	0,360	18.676	0,268	0,443	26.677
Test scores:												
3rd grade Math (sd)	0,08	0,974	618.704	0,444	0,912	80.482	0,841	0,803	7.192	0,644	0,861	9.431
3rd grade Read (sd)	0,09	0,974	622.053	0,412	0,9	80.771	0,714	0,781	7.261	0,524	0,865	9.497
5th grade Math	216,6	45,7	1.375.636	232,6	46,5	190.189	250,2	43,8	17.312	240,6	45,7	22.911
5th grade Portuguese	204,9	47,5	1.375.636	219,1	47,1	190.189	232,2	44,4	17.312	224,5	46,2	22.911
Mother's education:												
College	0,196	0,397	825.219	0,26	0,439	113.854	0,313	0,464	10.649	0,286	0,452	13.923
High School	0,236	0,424	825.219	0,272	0,445	113.854	0,265	0,441	10.649	0,254	0,435	13.923
Elementary School	0,144	0,351	825.219	0,135	0,341	113.854	0,113	0,317	10.649	0,123	0,329	13.923
Father's education:												
College	0,193	0,395	714.030	0,251	0,434	97.792	0,259	0,438	9.256	0,261	0,439	11.868
High School	0,186	0,389	714.030	0,229	0,42	97.792	0,228	0,42	9.256	0,21	0,407	11.868
Elementary School	0,146	0,353	714.030	0,147	0,354	97.792	0,138	0,344	9.256	0,14	0,347	11.868
Household characteristics:												
Cars	0,705	0,847	1.339.338	1,02	0,904	187.366	1,36	0,874	17.200	1,18	0,905	22.647
Bedrooms	2,40	0,90	1.324.404	2,55	0,852	185.894	2,83	0,768	17.079	2,75	0,823	22.470
Bathrooms	1,36	0,68	1.321.953	1,49	0,755	185.389	1,70	0,835	17.061	1,60	0,811	22.412
Computers	0,81	0,86	1.338.035	1,09	0,94	187.157	1,34	0,96	17.171	1,21	0,97	22.598
Child rearing practices:												
Live with both parents	0,596	0,491	1.319.703	0,645	0,478	185.201	0,776	0,417	17.062	0,66	0,474	22.455
Parents always go to school meetings	0,586	0,493	1.323.831	0,609	0,488	185.543	0,652	0,476	17.053	0,609	0,488	22.441
Work	0,127	0,333	1.313.485	0,095	0,293	184.534	0,072	0,258	16.998	0,089	0,284	22.359
Household chores (h)	1,37	1,09	1.309.133	1,34	1,04	184.420	1,36	0,98	16.998	1,38	1,02	22.329
Hours in internet, games and tv	2,25	1,29	1.309.004	2,41	1,27	184.168	2,44	1,21	16.960	2,39	1,26	22.318

Notes: All individuals in our sample of 5th graders in 2015 and/or 3rd graders in 2013 classified as white, mixed (pardo) or far-eastener (amarelo).
Age-grade distortion at 5th grade (years): if negative, it means that student at the 5th grade has an age-grade distortion, i.e., is more than 10 years of age.