

Baronchelli, Adelaide; Foresta, Alessandra; Ricciuti, Roberto

**Working Paper**

## The Words That Keep People Apart. Official Language, Accountability and Fiscal Capacity

CESifo Working Paper, No. 8437

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Baronchelli, Adelaide; Foresta, Alessandra; Ricciuti, Roberto (2020) : The Words That Keep People Apart. Official Language, Accountability and Fiscal Capacity, CESifo Working Paper, No. 8437, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/223509>

**Standard-Nutzungsbedingungen:**

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

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

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

**Terms of use:**

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

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

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

**The Words That Keep People  
Apart. Official Language,  
Accountability and Fiscal  
Capacity**

*Adelaide Baronchelli, Alessandra Foresta, Roberto Ricciuti*

## **Impressum:**

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email [office@cesifo.de](mailto:office@cesifo.de)

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: <https://www.cesifo.org/en/wp>

# The Words That Keep People Apart. Official Language, Accountability and Fiscal Capacity

## Abstract

This paper empirically evaluates the impact of accountability on fiscal capacity. It adopts an instrumental variable approach using, as an instrument, the measurement of how far, on average, official language differs from ordinary language. The main hypothesis is that if the average citizen cannot understand the central government and the elite, she will find it difficult/impossible to hold the government to account. The first stage results suggest that this instrument is strong and reliable and is negatively correlated with our measure of accountability in line with the hypothesis. The results in the second stage support the main hypothesis. Our results are robust to plausible exogeneity tests and different specifications.

JEL-Codes: H200, D020, D720, C260.

Keywords: language, accountability, fiscal capacity, insulation.

*Adelaide Baronchelli*  
*University of Verona / Italy*  
*adelaide.baronchelli@univr.it*

*Alessandra Foresta*  
*University of Verona / Italy*  
*alessandra.foresta@univr.it*

*Roberto Ricciuti*  
*University of Verona / Italy*  
*roberto.ricciuti@univr.it*

## 1. Introduction

Taxes are crucial for the proper functioning of states because they enable fundamental services, such as education, healthcare, and military defense to be provided. According to the theoretical literature, fiscal capacity (the ability of the government to raise revenues from a broad base) depends on government accountability. In turn, accountable institutions are likely to be found in rich countries, which are able to raise higher taxes. To solve this endogeneity issue, we propose an instrumental variable based on a channel relating institutions and accountability running through the official language of the country.<sup>1</sup> We suggest that if the official language differs markedly from the language spoken every day by common folk, this insulates<sup>2</sup> the central government from the population. Consequently, the accountability of the political elite is reduced, which in turn leads to lower fiscal capacity.

Fiscal capacity (as analyzed by Acemoglu, 2005; Acemoglu et al., 2011; Besley & Persson 2009, 2011 and Glaeser et al., 2004) is related to economic development for two main reasons. Firstly, as suggested by Besley & Persson (2014), a higher level of fiscal capability increases the amount the resources available to the state for public goods provision. Secondly, a higher level of fiscal capacity is usually associated with a larger bureaucracy which can provide an environment for more capable states (Moore, 2004).

The importance of accountability for fiscal capacity is dealt with in previous work such as Besley & Persson (2014) and Ricciuti et al. (2019). These authors suggest that accountability and transparency are fundamental elements for the proper functioning of the tax system. As proposed by Besley & Persson (2009), the existence of checks and balances forces the incumbent to promote common interests, rather than personal or group gains. Additionally, the presence of mechanisms of accountability facilitates the “fiscal contract” between citizens and the state, reducing taxation transaction costs (Levi, 1988) and generating a “tax morale” (e.g., Doerrenberg & Peichl, 2013 and Luttmer & Singhal, 2014).

In this framework, we investigate a new causal channel linking accountability with fiscal capacity. We argue that when the official language, used by the elite, is distant from the

---

<sup>1</sup> The *Concise Oxford Companion to the English Language* defines the official language, or state language, as a language given a special legal status in a particular country, state, or other jurisdiction. Typically, a country's official language refers to the language used in government dealings (by the judiciary, legislative bodies and administration).

<sup>2</sup> Aghion et al. (2004) model the trade-off between ex ante insulation (the share of votes that can obstruct a leader trying to implement legislation) written into the constitution and ex post control. In their analysis insulation is a feature of the delegation of power.

language spoken by the people, a gap is created between these two groups. The average citizen, unable to relate to the government, loses interest and trust in its functioning while the elite isolates itself more and more from the population, resulting in the selection of bad policies. To implement our analysis, we construct a cross country database using information from different sources, such as the World Bank and IMF. We first follow an OLS strategy, which finds a strong and positive correlation between fiscal capacity and accountability. This is followed by an IV approach using an instrument based on a measure of linguistic distance. Due to its construction, we consider the measure as exogenous and, due to the importance of language in communication and bureaucracy, we consider the instrument strongly correlated with our main measure of accountability. The results are in line with our main hypothesis. To reinforce our findings, we implement a series of sensitivity checks, based on Conley et al. (2012), and a series of robustness checks. These findings are mostly in line with our main results.

Economists have addressed language-related topics in a number of ways (Ginsburgh and Weber, 2020). First, in international trade, language similarity is considered a facilitating factor that, other things being equal, increases exchanges (Melitz, 2008). Second, for policy outcomes such as in healthcare, if government guidance is provided in the language normally used by laypeople, it is much more effective (Djité, 2008; Gomes, 2014). Third, assimilation policies, in which education is compulsory given in a language that is different from the one of the immigrant groups, may lead to more in-ward behavior by the member of the same groups (Bisin et al., 2011; Fouka, 2020). Fourth, ethnolinguistic fractionalization, dealing with the number, size, and geographical location of distinct cultural groups in a state, has been studied as a source of civil conflict, underprovision of public goods and low economic growth (Alesina et al., 2003; Easterly & Levine, 1997). Fifth, a strand of studies analyzing the effect of culture on economic outcomes uses linguistic variables as a proxy for cultural values and beliefs (Licht et al., 2007; Tabellini, 2008, 2010; Davis & Abdurazokzoda, 2016; Galor et al., 2018).

Over the last two decades “persistence studies” have addressed several issues claiming that the legacy of some events can be seen in current outcomes. Colonization is the most important event considered in this literature, which analyzes colonial institutions, policies, and economic outcomes (Acemoglu et al., 2001; Nunn, 2008). Other studies address permanent features such as genetic diversity (Ashraf & Galor, 2013) and long-term climatic fluctuations (Ashraf & Michalopoulos, 2015). In our case, colonization transferred a language from Europe to other territories, creating the conditions for keeping it even long after colonization ended. Moreover,

not only the transfer of institutions from Europe to Africa and the Americas, is in place here, but a more subtle interplay exists between foreign and national elites, determining linguistic differences, as in parts of Asia.

The rest of the paper is structured as follows: section two reviews the literature linking language to several economic and political outcomes, highlighting the mechanisms by which language impacts on accountability. Section three provides the historical background. Section four describes the measurement of linguistic distance. Section five presents the data. Sections six and seven set out the results of our analysis, and section eight deals with robustness checks. Section nine concludes.

## **2. Official language and accountability**

The key element in this study is the difference between the official language and the languages spoken by common people and how this relates to accountability. This section explains how this crucial relationship unfolds. Specifically, we highlight two channels through which linguistic distance impacts on accountability, i.e. via identity and communication. These channels convey two different concepts of the role of language. On the one hand, language shapes the identity of the elite and, on the other, it acts as the mean of communication between the elite and the people. These two channels isolate the elite from laypeople, resulting in low accountability.

### *2.1 Language as an identity maker*

Linking linguistic distance with accountability, we investigate the relationship between official language and social identity, defined as the part of an individual's self-image that comes from membership of a social group (Landa & Duell, 2015). Social identity is crucial in determining the way an individual thinks of him or herself in relation to the world. We argue that official language plays a key role in establishing the social identity of elites. The further the official language from the language of common people, the greater the difference in the social identities of the two groups. Different social identities, in turn, create a rift between the elites in a state and its citizens, which results in low accountability. The literature includes two interpretations of the link between language and social identity: the first identifies language as an *identity marker*; the latter as an *identity maker*.

According to the first interpretation (*identity marker*), language is one of the most distinguishing features of a social group, along with behavior, values and lifestyle. It is also the easiest to identify. This interpretation can be found in the literature on ethnolinguistic fractionalization, which suggests that diversification along linguistic and ethnic lines within a state severely impairs public good provision, social comity and economic growth (e.g., Easterly & Levine, 1997; Fearon, 2003; Alesina et al., 2003; Alesina & La Ferrara, 2005, Desmet et al., 2012). In these studies, ethnicity is marked by different traits, including a common language, religion and customs (Fearon, 2003). We agree with the literature on ethnolinguistic fragmentation in arguing that the presence of different social groups within a state negatively impacts on its functioning. However, we differ from this strand of literature on two essential issues. In our explanation of how governments are held to account, we are not concerned with the impact of horizontal diversity (i.e. among groups in the population) but focus on vertical diversity. We believe that low accountability levels may be explained by the existence of a ruling class which, over the decades, has created its own specific social identity. In addition, we interpret language not as an identity marker but as an identity maker. We argue that official language has been a key determinant in the formation of social identity.

According to the interpretation of language as an *identity maker*, language does not simply identify values and beliefs but shapes them. Thus, the use of the official language has progressively influenced the beliefs of the elite, its values, lifestyles, and, ultimately, social identity. The linguistic literature supports this interpretation. Whorf (1956) and Sapir (1970) advanced the hypothesis that language functions as a constraint on the development of cultural norms. Hill & Mannheim (1992) pointed out that grammatical categories implicitly reinforce specific cognitive or social categories. By conducting experiments on multilingual individuals, Nisbet (2003) shows that the same question in different languages produces different outcomes, arguing that each language reflects a specific interpretation of the world. Kashima & Kashima (1995) empirically test the correlation between the global characteristics of cultures and the languages used in those cultures. They found that the rules correlated to pronoun dropping are correlated with a high level of individualism, suggesting that some grammatical rules act as a constraint on self-other relations. According to these studies, the use of a language reproduces the values and beliefs of the culture conveyed by that language. Specifically, language contributes to the formation of cognitive and social categories that define a person's sense of self and how people should behave. We argue, therefore, that language is a key determinant of a person's social identity.



By linking accountability with linguistic distance, our argument is twofold. First, in line with the linguistic literature, we argue that language influences the social and cognitive categories that form individual social identity. As a consequence, when the official language differs from the language spoken by ordinary people, the elite behaves in a way that is foreign to the background of the territory it governs. This gradually creates a vertical rift in the population. Two separate groups are created, with their own set of values, beliefs and lifestyles: the elite at the top, citizens at the bottom.

Second, as in the literature on ethno-linguistic fragmentation, we show the consequences of fragmentation on accountability, identifying two types of mechanism that may be at play in linking vertical diversity with low accountability. On the one hand, membership of a social group shapes individual preferences and members of the group are also more likely to “attribute positive utility to the well-being of members of their own group and negative utility to that of members of other groups” (Alesina & La Ferrara, 2005: 765) On the other hand, within a community, norms of reciprocity are effective since cooperation is expected and noncooperators are sanctioned (Miguel & Gugerty, 2005). Conversely, non-cooperation is socially acceptable outside the group.

These two mechanisms combined contribute to the insulation of the elite from their citizens, which in turn renders the elite unaccountable for their decisions, reducing the quality of governance.

## *2.2 Language as a tool of communication*

The official language is a key element in the functioning of a state since it allows communication between the state and its citizens. Countless activities are normed in this language, from registering as a business to drafting an employment contract, from patenting an idea to paying taxes, accessing health services and settling a lawsuit in court, to name only a few. All government bodies and administrative matters require a homogenous tool of communication between people. Translation is a possibility but is extremely costly compared to direct communication (Melitz, 2008). Therefore, the inability to speak the official language of the state can exclude people from full citizenship in their own country. Without a viable channel of communication, many people simply do not understand what they are being asked to do and what they can rightly demand.

The communication gap between the state and its citizens is widened when the official language is completely foreign to the users. As argued by Laitin & Ramachandran (2016), in these cases the process of understanding what is expressed in the state language is much more difficult than it would be if the chosen channel of communication was less remote. This is evident in some former colonies which adopted the colonizer's language as the state language. Here, the majority of the population is not able to properly speak the official language of their own country. Albaugh (2014), for instance, reports that in Sub-Saharan African countries, on average, only 18.7% of the population can speak the official language of the state, with a minimum of 4.5% in Niger and 5% in Guinea.

Here, we argue that the linguistic distance between the people and the elite may severely impact on accountability. The inability to communicate is crucial. When citizens do not understand the official language, they are prevented from accessing many services. In other words, their inability to speak the state language acts as a wall in their communications with the state. Hence, they have no instruments to monitor the actions of the ruling elite and they are not able to hold the government and the elite to account.

Several studies have already highlighted that language acts as a barrier to the comprehension of key information and access to many public services. First, language is crucial to the success of many health programs and individual health outcomes (Djité, 2008). In a field experiment conducted in Kenya, evidence from Translators Without Borders (2015) shows that providing medical information in Swahili in Tanzania instead of English enormously increases the awareness of how Ebola is transmitted. Gomes (2014) shows that in Africa the higher the individual linguistic distance from neighbors, the higher the child mortality. In all these studies, the key argument is that if people do not understand the language of the medical professionals, they lack the means to properly follow medical instructions.<sup>3</sup>

In addition to health, education outcomes are also influenced by the choice of the language of instruction. In schools and in the other public offices, the official language of the state is used. Evidence from multiple studies, however, shows that when lessons are in the students' mother tongue both grades and school attendance improve. Field experiments in Cameroon (Laitin et al., 2019) and Uganda (Kerwin & Thornton, 2018) reveal that students learning in the indigenous language performed significantly better than those taught exclusively in English.

---

<sup>3</sup> This also holds for migrants in developed countries, who have acquired little familiarity with the language of the host county (Bleakley & Chin, 2004, 2008, 2010; Ramirez et al., 2008).

The key mechanism that links education in a foreign language to these various aspects is the lack of comprehension of the relevant concepts. When children are taught in a language they do not frequently use at home or with their relatives and friends, they are less able to understand what they study.

[Figure 1 approximately here]

Figure 1 summarizes how linguistic distance influences accountability through two main channels. On the one hand, the adoption of a foreign official language contributes to the formation of elites with a social identity that is profoundly different from that of the population. On the other hand, the inability to understand the official language is an obstacle to effective communication between the elite and lay people.

### **3. Historical background**

We cannot talk about language differences and the distance between common citizens and the elite, without mentioning colonization. This phenomenon was one of the main causes of the linguistic separation between the people and the elite and still operates in many countries. Colonial governments coopted educated indigenous people to translate and help run the local government, and chose small, educated elites they were able to control and tie to the interests of the colonial state. In addition, colonial governments had little interest in mass education since it was expensive, reduced revenues and potentially fostered rebellion. If education was encouraged, government officials primarily backed a practical variety, such as carpentry, masonry, and horticulture (Kelly, 2000a; 2000c; Sundkler & Steed, 2000; Manning, 1998; Furley and Watson, 1978).

For example, in Vietnam, the French colonial governments closed down indigenous schools, pressured the Japanese government to prevent the Vietnamese from acquiring education in Japan, and educated only as many Vietnamese after primary school as the colonial government could hire and control (Kelly, 2000a; 2000c). The French focused education on practical skills (particularly farming) and did not provide the Vietnamese with skills that would let them compete with French settlers for senior positions in the colonial administration. Throughout Africa, the French educated only a small elite, which was purposely trained to be separate from other Africans in language and culture (Kelly, 2000b; 2000d; Grier, 1999). Similarly, the

Italians, Portuguese, and Spanish also educated only a small portion of the non-European population in their colonies (Isichei, 1995).

A similar approach was adopted by the United Kingdom. The British made little effort to educate colonial subjects since they tried to run their colonies as cheaply as possible (Ferguson, 2002). Slave owners and those who used unskilled and forced labor were especially averse to education, often even refusing to teach slaves how to read (Blouet, 1990). For example, before 1813, the British East India Company schools in India trained a total of only a few hundred students, almost exclusively elite Muslims and Hindus from the highest castes, and lower-caste Hindus were explicitly excluded from company schools (Ingham, 1956).

Later, however, Protestant missionaries allowed the British to service their colonies cheaply. In areas with high missionary influence, the British government tried to shape education to its own interests, establishing financial incentives for missionaries who educated a few students more intensely, using the English language and adopting a utilitarian, government-imposed curriculum. Additionally, the colonial government encouraged missionaries involved in mass education to focus on practical skills (Furley and Watson, 1978; Ingham, 1956). The main consequence of this system was the creation of a small elite and a broader pool of English-speaking laborers who had the skills required by British companies and planters, carpenters, bricklayers, etc.

The legacy of colonialism can be seen in today's institutions. For instance, the relationship between the elite and the people in Africa can be described through the lens of "neo-patrimonial politics" (Chabal & Daloz, 1999). African politics works through the distribution of the resources of the state (power, status, wealth, access to markets, etc.) via informal, personal patron-client networks, rather than the formal, impersonal channels of the Weberian legal-rational state. For political elites, power thus becomes a zero-sum, winner-takes-all struggle for control of the legally recognized state and its resources. This power struggle permeates all levels of society and often leads to short-termism, rather than long-term developmental planning.

#### **4. Measuring linguistic distance**

To measure the distance between the language spoken by the different linguistic groups within a country and the official language, we use the Average Linguistic distance from the Official Language (ADOL) indicator put forward by Laitin & Ramachandran (2016). This measure is

based on a cladistic definition of linguistic distance as shown by the Ethnologue linguist tree (Eberhard et al., 2019). Tree diagrams classify the structural relationship between languages using nodes, which represent the evolution points of the development and differentiation of languages. Previous literature, like Fearon & Laitin (1999), Fearon & Laitin (2000), Laitin (2000), and Fearon (2003), uses the linguistic tree to produce noisy measures of the distance between cultural groups that speak different languages. To better understand the idea behind linguistic trees, Figure 2 shows the (simplified) Proto-Indo-European Language Tree.<sup>4</sup>

[Figure 2 approximately here]

Consider Bengali, a language spoken in India and Bangladesh, as our languages of reference. Bengali has only one node in common with English, because both are Indo-European languages, although they are quite different. In contrast, Bengali has five nodes in common with Hindi. Both languages belong to the Indic area of the Indo-Iranian segment of Indo-European languages. Finally, Bengali has two nodes in common with Kurdish, for example, being both Indo-Iranian languages. Finally, it has no nodes in common with non-Indo-European languages, like Hungarian, which is a Uralic language.

Similar reasoning can be applied to non-Indo-European languages, like Amharic, a language spoken in Ethiopia. Amharic is a Semitic language, part of the Afro-Asian and not the Euro-Asian linguistic platform, from which it began to differentiate a long time ago. Thus, this language is utterly unlike Indo-European languages like English or Italian. It has similarities with Hebrew and Arabic with which it has several nodes in common. To conclude, the linguistic tree shows that closer languages, with more nodes in common, have a similar evolution and languages without nodes in common developed separately.

To construct their measure of the linguistic distance, Laitin and Ramadrachan (2016) used the procedure developed by Fearon (2003). Formally, the linguistic distance between two languages is calculated as:

$$d_{ij} = 1 - \left( \frac{\text{no. of common nodes between } i \text{ and } j}{\frac{1}{2}(\text{no. of nodes for language } i + \text{no. of nodes for language } j)} \right)^\lambda \quad (1)$$

Here,  $d_{ij}$  is equal to one when the difference between two languages is maximal i.e. there are no common nodes between the two languages. Conversely, the difference is minimal the lower

---

<sup>4</sup> Due to space constraints, the linguistic tree is a simplified version of the tree used in this paper and of the Ethnologue tree.

the measure and the higher the number of common nodes. Additionally, following Fearon (2003) and others, we select a value for  $\lambda$  equal to 0.5.

Laitin & Ramachandran (2016) calculated the measure in Eq. (1) for every language existing in a country with respect to the official language. They then combined the distances with the different shares in population in each country. The ADOL for any country  $i$  is calculated as follows:

$$ADOL_i = \sum_{j=1}^n P_{ij} d_{jo} \quad (2)$$

where  $n$  represents the number of linguistic groups in the country,  $P_{ij}$  is the population share of the group  $j$  in country  $i$ . Finally,  $d_{jo}$  measures the distance of the language of group  $j$  from the official language  $o$ .

This measure is different from the indices of linguistic fragmentation in the literature (e.g. Alesina et al., 2003; Desmet et al., 2009; Greenberg, 1956), focused more on linguistic heterogeneity within the total population. For example, the index proposed by Alesina et al. (2003) focuses on the weight of the different shares of the populations with respect to the overall population of the country. This index stresses the differences over the population (“horizontal”) and not the differences between the different segments of the population and the elite. In contrast, the peripheral heterogeneity index proposed by Desmet et al. (2009)<sup>5</sup> is more similar to ADOL, describing the distance between the languages spoken by the lay people (as ADOL) and the language of the largest linguistic group in the country, while ADOL uses the official language. Haiti is a good example of the distinction between the language spoken by the majority of the population and the official language (Liu & Pizzi, 2016). In this country, the official language is French, although the French-speaking group is only 4% of the population. Simply referring to the largest group can help to understand the “horizontal” dynamics but not “vertical fragmentation”.<sup>6</sup>

---

<sup>5</sup> In this paper, the authors analyze the relevance of including linguistic distance in a diversity index and the different impact of these measures on redistribution. The indices other than the peripheral heterogeneity index are: a diversity index without distances (ELF) (Atlas Narodov Mira, 1964); a diversity index with distances (Greenberg, 1956); a polarization index without distances (Reynal-Querol, 2002); a polarization index with distances (Esteban and Ray, 1994).

<sup>6</sup> The concept of vertical fragmentation is close to that of linguistic disenfranchisement elaborated by Ginsburgh and Weber (2005; 2011). This notion refers to the introduction of “ideologies and structures which are used to legitimate, effectuate, and reproduce unequal division of power and resources (both material and non-material) between groups defined on the basis of language” (Skutnabb-Kangas and Phillipson, 1989: 455). Indices of linguistic disenfranchisement account for multilingualism measuring the distance between the languages spoken

Figure 3 plots the values of ADOL across the world. The Figure shows a clear concentration of high values in Africa and in most former colonies, such as India, with lower values in Asia and still lower in Europe, the home of the colonial powers. The level is also low in Latin and North America, which were also colonized, but where native languages were almost wiped out by colonization, unlike in Africa, for example.

[Figure 3 approximately here]

## 5. Data

This paper uses a cross-country cross-section sample. Data are collected at the country level from multiple sources, generating a novel dataset.

### 5.1 Fiscal Capacity

Following the previous literature, including Besley & Persson (2009) and Dincecco & Prado (2012), we use a classical measure of fiscal capacity, which is the overall amount of taxes divided by the GDP at the country level (per 100). The source of these data is the UN-WIDER *Government revenue dataset*.<sup>7</sup> We use an average measure between 1995 and 2017 at the country level. To understand the distribution of the index across different countries, the values of the variable across the world are shown in Figure 4.

[Figure 4 approximately here]

Figure 4 matches the theoretical results in the literature, for example in Besley & Persson (2009). On average, the richest states are better able to collect taxes than poor and developing countries, especially in Africa.

### 5.2 Accountability

The second fundamental variable in our analysis is accountability. This measure is not straightforward and cannot be easily measured, like taxes. We need to rely on an overall index based on expert evaluations. Following the literature, we use a variable of institutional quality

---

by an agent and the core languages of a state, while ADOL does not. However, ADOL shares with the notion of linguistic disenfranchisement the idea that denying an individual's linguistic rights negatively affects its political and economic opportunities within a society.

<sup>7</sup> <https://www.wider.unu.edu/project/government-revenue-dataset>

at the country level put forward by the World Bank. This index belongs to the World Governance Indicators and it is called *Voice and Accountability*.<sup>8</sup> The variable measures: “[...] *perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media*”.<sup>9</sup> We averaged this index over 1995-2017 at the country level. This measure varies from -2.5 to +2.5. The minimum values are for countries with low or no level of accountability, such as dictatorships, and the higher values are for more accountable countries, such as democracies. In this case too, the index is plotted onto a World map, shown in Figure 5.

[Figure 5 approximately here]

Figure 5 shows that our variable works exactly as predicted by the previous literature. Higher values of the index are concentrated in more democratic countries, especially in Europe and other democracies, with maximum values in Scandinavian countries. Lower values are concentrated in Latin America, Asia and Africa.

### 5.3 Other variables

Our models include a set of controls that could influence our results. To start with, we generate two dummies for the *British legal origin* and *Socialist legal origin* of the countries, because legal origins can influence a broad range of economic outcomes, as demonstrated for example by La Porta et al. (2008). We also include a variable, in logarithmic form, for the size of the population in different countries, averaged between 1995 and 2017. Population is relevant because it influences important national characteristics such as government size or the technical ability of the central government to collect taxes, as suggested for example by Besley & Persson (2009) and Desmet et al. (2020).

Moreover, we include in our models a set of variables connected with economic outcomes. First, we control for the average GDP per capita (in log) over the years 1995-2017 (like Besley & Persson, 2009; Desmet et al., 2020), using the information provided by the IMF at the country level.<sup>10</sup> Second, we insert a series of variables for the type of economy. Following Desmet et al. (2020), we include a variable for the “potential role” of agriculture, the amount of arable land (in log), averaged between 1995 and 2017. Additionally, we compute a variable for the

---

<sup>8</sup> <https://info.worldbank.org/governance/wgi/>

<sup>9</sup> Source: World Bank (<https://info.worldbank.org/governance/wgi/Home/Documents>)

<sup>10</sup> <https://data.imf.org/?sk=388DFA60-1D26-4ADE-B505-05A558D9A42&sId=1479329132316>



percentage of world oil reserves in the country, following Laitin & Ramachandran (2016). In line with the previous variable, we compute this measure as an average between 1995 and 2017. In addition, in line with the work of Besley & Persson (2009), we also include a variable for the average number of years spent fighting external wars. Finally, following the previous literature, we add a measure for linguistic fractionalization, as proposed by Desmet et al. (2009). The descriptive statistics for these variables are in Table 1.

[Table 1 approximately here]

## 6. Empirical analysis: OLS approach

To explore the relationship between fiscal capacity and accountability, we first use an OLS regression as presented in equation (1):

$$FC_i = \beta_0 + \beta_1 A_i + \beta_2 X_i + \varepsilon_i \quad (3)$$

where  $FC_i$  is the dependent variable, the average fiscal capacity,  $A_i$  the proxy for accountability and  $X_i$  represents a set of controls as indicated in Section 4. The variables are measured at the country level. Table 2 presents the results.

[Table 2 approximately here]

The results show that the accountability index is positive and highly significant. In terms of magnitude, in specification (1) the increase of one standard deviation is equivalent to an increase of 0.66 standard deviations in our dependent variable. However, with an increase in the number of control variables, our coefficient of interest slowly decreases. Overall, it remains more or less stable. For example, in specification (8), which contains all the control variables, the increase of one standard deviation is equivalent to an increase of 0.611 standard deviations in our proxy for fiscal capacity, quite similar to the results of specification (1).

Table 2 also provides other interesting findings. For example, there is a negative and non-significant effect for British legal origins, while the results for Socialist legal origins are positive and significant for all specifications. Other significant coefficients are those for arable land (in log) and the measure of linguistic fractionalization. The results for arable land are positive and statistically significant, while those for the measure of linguistic fractionalization are negative

and highly statistically significant. The remaining variables are non-significant in all specifications. To sum up, the OLS strategy proves that accountability is positively related to fiscal capacity, even when other variables are included in the model.

Although these findings are promising, the OLS strategy raises significant concerns about the validity of the causal relationship. First, the possible influence of omitted variables affecting both the main regressor and the dependent variable can seriously undermine the causal interpretation of our results. Examples are cultural variables, the education system, or the type of media communication. Although we try to avoid this problem using control variables, such as British legacy for the education system, it is not always possible to control for these variables.

Second, reverse causality, i.e. the impact of the dependent variable on the main regressor, may undermine the reliability of our conclusions. This is more unlikely, especially because we use the average measure, but a higher level of fiscal capacity could, for example, increase the infrastructure, the reputation, and quality of politicians and bureaucrats, raising the level of transparency and accountability.

## 7. Empirical analysis: Instrumental variable approach

### 7.1 IV results

To solve the possible endogeneity problem and to ensure the causal interpretation of the results, an instrumental variable approach is commonly used. Here we consider accountability correlated with linguistic distance and use the measure proposed in Laitin & Ramachandran (2016), as presented in Section 4 as an instrument for accountability. Formally our approach is presented in equation 2 and 3 below

$$A_i = \beta_0 + \beta_1 ADOL_i + \beta_2 X_i + \mu_i \quad (2)$$

$$FC_i = \delta_0 + \delta_1 \hat{A}_i + \delta_2 X_i + \varepsilon_i \quad (3)$$

where  $A_i$  is the proxy for accountability, instrumented by  $ADOL_i$  the instrument.  $FC_i$  is the dependent variable, the average fiscal capacity and  $X_i$  represents a set of controls, as presented in section 5.3.

To be valid, an instrument must meet two requirements: a) correlation with the endogenous variable (relevance condition); b) exogeneity, i.e. no correlation with the error term (exclusion restriction). We believe that accountability and language are strongly intertwined. As argued in

section 2, language is a fundamental tool for the understanding and interaction of citizens, in order to hold the central government and the elite to account.

However, relevance alone is not enough for an instrument to be valid. A twofold argument supports the exogeneity of our instrument. First, ADOL is computed and measured before the other variables, at the beginning of the '90s. This avoids concerns regarding reverse causality. Second, we consider our variable as a historical measure because languages and linguistic differences today are the result of lengthy historic processes, such as colonization, as noted in section 3. Using a variable that is highly correlated with our instrumented variable and related to past experience ensures the exogeneity of the instrument. Additionally, we carry out the *plausible exogeneity test*, based on Conley et al. (2012), which evaluates our results in the case of a small violation of the exogeneity assumption (for more details see section 7.2).

We implement an IV regression following the 2SLS methodology. The results from the first stage (Panel A) and the reduced form (Panel B) are presented in Table 3.

[Table 3 approximately here]

The findings from the first-stage regressions (Panel A) indicate a negative and significant relationship between our instrument (ADOL) and the instrumented variable for all specifications.<sup>11</sup> Additionally, the coefficients are relatively stable across different specifications. The coefficients for specification (1) of the variable of interest, with no controls, is -0.937, while the coefficient for specification (8), which includes all the controls, is -1.297. The F-statistics are all above 10, and above the critical values for weak test ID in the case of the 10% maximal IV size for all specifications put forward by Stock & Yogo (2002).<sup>12</sup> These results are in line with our main hypothesis for the instrument. An increase in the average linguistic distance leads to a decrease in the general level of accountability.

Other significant coefficients in our first stage include socialist legal origin, population (log), arable land (log), and the average percentage of oil reserves. To conclude, the results in Table 3 show that ADOL is highly correlated with our instrumented variable.

In Table 3, Panel B, the findings indicate that coefficients for our instrumented variable remain positive and statistically significant for all specifications. The size of coefficients varies for the

---

<sup>11</sup> We also implement our analysis in controls of the level of primary and secondary education, although Laitin & Ramachandran (2016) point out that this may be a bad control because the independent variables directly influence the level of education. The coefficients are in line with the findings presented in Table 3.

<sup>12</sup> In this case, the critical value is 16.38.

different models, but not the sign. For example, focusing on the specification (1) with no controls, the coefficient of interest is 13.19 and one standard deviation increase in the instrumented variable is equivalent to an increase of 1.17 standard deviation for fiscal capacity. On the contrary, specification (8) shows a coefficient of 7.071 and one standard deviation increase in accountability is equivalent to an increase of 0.62 standard deviation for fiscal capacity. As the number of controls increases, the smaller the coefficient of interest decreases, but it remains positive and significant. Moreover, it is interesting to note that the standardized coefficients for OLS and IV approaches are quite different for specification (1) while they are almost identical for specification (8). These findings show that an increase in accountability corresponds to an increase in fiscal capacity, as suggested by the theoretical models and in line with our hypothesis. Other significant variables in this framework include socialist legal origin, which is always positive and significant, arable land (log), positive and only slightly significant in specification (8), and the measure of linguistic fractionalization, which is negative and significant.

A possible concern over our instrument is the presence of countries that changed their official languages at some point. In our sample, only four countries changed their official language: Laos, Somalia, Sri Lanka and Sudan. Replicating the previous analysis excluding these countries from our sample, the results do not change.<sup>13</sup>

### 7.2 Sensitivity test: Conley test

As suggested in the previous sections, the findings depend on the reliability of our instrument. While in the previous section we argued that our identification strategy is likely to hold, in this Section, following Desmet et al. (2020), we challenge our findings allowing small violations of the exclusion restriction. Following the methodology of Conley et al. (2012), known as “plausible exogeneity”, allows the instrument to have a direct impact on our outcome variable  $Y_i$ . Consider the following equation:

$$Y_i = \gamma_0 X_i + \gamma_1 WGI_i + \gamma_2 ADOL_i + \epsilon_i \quad (4)$$

where  $ADOL_i$  is the instrument and  $WGI_i$  the instrumented variable. The exclusion restriction implies that  $\gamma_2 = 0$ . The Conley test<sup>14</sup> provides a procedure that allows inference even if  $\gamma_2$  is not exactly zero. We follow the *local to zero approach* which also assumes that

---

<sup>13</sup> Findings available upon request.

<sup>14</sup> We implement the test using *plausexog* command on STATA by Clarke (2017).

$\gamma_2 \sim \mathcal{N}(0, \delta^2)$ .<sup>15</sup> Using this methodology, we obtain a 95% confidence interval for  $\gamma_1$ , our coefficient of interest. The results of the test are presented in Table 4.

[Table 4 approximately here]

Table 4 presents the variation in confidence intervals for our instrumented variable according to different levels of violation of the exclusion restrictions for the different specifications of our model, based on Conley et al. (2012). The size of the allowed violation permitted depends on the chosen  $\delta$ . While the significance of our coefficient decreases, Table 4 suggests that our instrument is robust to violation of the exclusion restrictions up to 70%. This reinforces the findings of Table 3, which seem to be robust even in the presence of significant violations of the exclusion restriction.

## 8. Robustness checks

### 8.1 Using Executive constraints

Accountability is very difficult and complex to measure. The literature includes different indices based on differing definitions. To check if our results depend on the chosen definition of accountability, we implement our analysis using the measure of institutional quality of the Polity-IV dataset, which quantifies the extent of institutionalized constraints on the decision-making powers of chief executives, averaged over 1995-2017. This section presents the findings for our instrumental variable approach.<sup>16</sup> The results for the first stage (Panel A) and the reduced form (Panel B) of the IV model are presented in Table 5.

[Table 5 approximately here]

The results for the first stage are negative and statistically significant for all specifications, in line with the results presented in Table 3. The F-stats are above the Stock-Yogo 10% critical value.<sup>17</sup> Moreover, the findings in Table 5, Panel B, are in line with Table 3. In specification (8) the increase of one standard deviation in our instrumented variable is equivalent to an increase of 0.84 standard deviation for fiscal capacity. In terms of magnitude, the coefficients

---

<sup>15</sup> We implement plausible exogeneity also using the *Union of confidence intervals* methodology. The results are the same. Findings available upon request.

<sup>16</sup> We also perform the analysis for the OLS model. The results are analogous and available upon request.

<sup>17</sup> In this case, the critical value of Stock-Yogo is 16.38.

appear to be higher using this measure of accountability. These findings suggest that our results also hold for a different definition and measure of accountability.

## 8.2 *Federalism*

As presented in Section 5, our dependent variable is the total amount of taxes over GDP. However, federal and non-federal states collect and manage taxes at different levels and may have some differences in language policy, creating room for heterogeneity. This unaccounted heterogeneity between different types of states can influence our results and undermine our interpretation. We replicate our baseline analysis including a variable controlling for the presence of a Federal state, called *Federal*. The results are presented in Table 6.

[Table 6 approximately here]

The coefficients for Federal are never significant in both panels and in all specifications presented in Table 6. The signs and the magnitude of our instrument and instrumented variables in both specifications (1) and (8) are similar to those in Table 3. The standardized  $\beta$ s for the instrumented variable are 1.19 and 0.59 in specifications (1) and (8) respectively. In this case too, the results for the F-stat are above the 10% Stock-Yogo threshold.

## 8.3 *Colonialism*

As suggested in Section 3, this analysis cannot avoid the subject and the influence of colonialism. Colonial history had a massive impact on the countries in our samples, on their state organization and their institutions. However, linguistic distance is not homogeneous in colonized countries and the effect of colonization does not overlap with our instrument. To further explore this idea and to control for possible noise produced by non-colonies, the baseline analysis is carried out excluding these countries from the sample. The results are presented in Table 7.

[Table 7 approximately here]

The coefficients presented in Table 7, Panel A (first-stage results) and Panel B (reduced-form results) are similar in sign and size to the coefficients presented in Table 3, Panels A and B. The results of the first stage for ADOL are negative and significant and the F-statistics are generally above the 10% Stock-Yogo thresholds, which in this case is 16.38. Moreover, the

results for our instrumented variable in the second stage are positive and highly significant, in line with the findings shown in Table 3, Panel A. In specification (8) the increase of one standard deviation of our instrumented variable is equivalent to an increase of 0.44 standard deviation for fiscal capacity. Finally, the reduction in the sample is relatively small, due to the small number of non-former colonies in our sample. To sum up, Table 7 indicates that our results are not driven by non-colonies and our findings are in line with Table 3.

## **9. Conclusions**

This paper analyzes a new transmission channel in the relationship between accountability and fiscal capacity in a cross-country sample. We use an average measure of linguistic distance from the official language called ADOL (Laitin & Ramachandran, 2016). The intuition behind this instrument is that if the official language is different from the language used every day by people, this creates a distance between the ruling elite and the people, which insulates the former from the latter, reducing accountability. The existence of an official language different from the everyday language is a lasting consequence of colonialism. We instrument our measure of accountability with ADOL and implement a 2SLS strategy. The results show a negative association between linguistic distance and accountability, and a positive relationship between accountability and fiscal capacity, which is also robust to a different definition of accountability, changes in the sample, and the plausibility of the instrument.

This complements previous channels described in the literature, in particular, ethnolinguistic fragmentation. This variable measures ‘horizontal fragmentation’ across groups in a given polity. Our instrument appraises ‘vertical fragmentation’, which represents the linguistic distance between a ruling group and the society.

Further work may explore the relationship between language diversity and trust in the government at a subnational level, for example in Africa and India.

## References

- Acemoglu, D. (2005). Politics and economics in weak and strong states. *Journal of Monetary Economics*, 52(7), 1199-1226.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369-1401.
- Acemoglu, D., Ticchi, D., & Vindigni, A. (2011). Emergence and persistence of inefficient states. *Journal of the European Economic Association*, 9(2), 177-208.
- Aghion, P., Alesina, A., & Trebbi, F. (2004). Endogenous political institutions. *Quarterly Journal of Economics*, 119(2), 565–611.
- Albaugh, E. A. (2014). *State-building and multilingual education in Africa*. Cambridge University Press.
- Alesina, A., & Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of economic literature*, 43(3), 762-800.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2), 155-194.
- Alidou, H., Boly, A., Brock-Utne, B., Diallo, Y. S., Heugh, K., & Wolff, H. E. (2006). *Optimizing learning and education in Africa—the language factor*. Paris: ADEA.
- Ashraf, Q.H. & Galor, O. (2013). The 'out of Africa' hypothesis, human genetic diversity, and comparative economic development. *American Economic Review* 103(1), 1-46.
- Ashraf, Q.H. & Michalopoulos, S. (2015). Climatic fluctuations and the diffusion of agriculture. *Review of Economics and Statistics* 97(3), 589-609.
- Atlas Narodov Mira (1964). Miklukho-Maklai Ethnological Institute at the Department of Geodesy and Cartography of the State Geological Committee of the Soviet Union. Moscow.
- Besley, T., & Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*, 99(4), 1218-44.
- Besley, T., & Persson, T. (2011). *Pillars of prosperity: The political economics of development clusters*. Princeton University Press.
- Besley, T., & Persson, T. (2014). Why do developing countries tax so little? *Journal of Economic Perspectives* 28(4), 99-120.



- Bisin, A., Patacchini, E., Verdier, T. (2011). Formation and persistence of oppositional identities. *European Economic Review*, 55, 1046–1071.
- Bleakley, H., & Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants. *Review of Economics and Statistics*, 86(2), 481-496.
- Bleakley, H., & Chin, A. (2008). What holds back the second generation? The intergenerational transmission of language human capital among immigrants. *Journal of human resources*, 43(2), 267-298.
- Bleakley, H., & Chin, A. (2010). Age at arrival, English proficiency, and social assimilation among US immigrants. *American Economic Journal: Applied Economics*, 2(1), 165-92.
- Blouet, O. M. (1990). Slavery and Freedom in the British West Indies, 1823-33: The Role of Education. *History of Education Quarterly*, 30(4): 625-43.
- Chabal, P., & Daloz, J.P. (1999). *Africa Works: Disorder as Political Instrument*, International African Institute.
- Clarke, D. (2017). PLAUSEXOG: Stata module to implement Conley et al's plausibly exogenous bounds.
- Conley, T. G., Hansen, C. B., & Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1), 260-272.
- Desmet, K., Gomes, J. F., & Ortuño-Ortín, I. (2020). The geography of linguistic diversity and the provision of public goods. *Journal of Development Economics*, 143, 102384.
- Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of Development Economics*, 97(2), 322-338.
- Desmet, K., Ortuño-Ortín, I., & Weber, S. (2009). Linguistic diversity and redistribution. *Journal of the European Economic Association*, 7(6), 1291-1318.
- Dincecco, M., & Prado, M. (2012). Warfare, fiscal capacity, and performance. *Journal of Economic Growth*, 17(3), 171–203.
- Djité, P. G. (2008). *The Sociolinguistics of Development in Africa* (Vol. 139). Bristol: Multilingual Matters.
- Doerrenberg, P., & Peichl, A. (2013). Progressive taxation and tax morale. *Public Choice*, 155(3-4), 293-316.

- Easterly, W., & Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions. *The Quarterly Journal of Economics*, 112(4), 1203-1250.
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2019). Ethnologue: Languages of the world.
- Eriksson, K. (2014). Does the language of instruction in primary school affect later labour market outcomes? Evidence from South Africa. *Economic History of Developing Regions*, 29(2), 311-335.
- Esteban, J. M., & Ray, D. (1994). On the measurement of polarization. *Econometrica*, 62(4), 819-851.
- Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8(2), 195-222.
- Fearon, J. D., & Laitin, D. D. (1999). Weak states, rough terrain, and large-scale ethnic violence since 1945. In Annual meetings of the *American Political Science Association*, Atlanta, GA (pp. 2-5).
- Fearon, J., & Laitin, D. D. (2000). Ordinary language and external validity: Specifying concepts in the study of ethnicity. In Annual meeting of the American Political Science Association, Washington, DC.
- Ferguson, N. (2002). *Empire: The Rise and Demise of the British World Order and the Lessons for Global Power*. New York: Basic Books.
- Fouka, V. (2020). Backlash: the unintended effects of language prohibition in U.S. schools after World War I. *Review of Economic Studies*, 87(1), 204–239.
- Furley, O.W. & Watson, T. (1978). *A History of Education in East Africa*. New York: NOK Publishers.
- Galor, O., Özak, Ö., & Sarid, A. (2018). Geographical roots of the coevolution of cultural and linguistic traits. *Available at SSRN 3284239*.
- Gazzola, M., & Wickström, B. A. (Eds.). (2016). *The Economics of Language Policy*. Cambridge MA, MIT Press.
- Ginsburgh, V., & Weber, S. (2011). *How Many Languages Do We Need? The Economics of Linguistic Diversity*. Princeton: Princeton University Press.

- Ginsburgh, V., & Weber, S. (2020). The economics of language. *Journal of Economic Literature*, 58(2), 348-404.
- Ginsburgh, V., & Weber, S. (2005). Language disenfranchisement in the European Union. *JCMS: Journal of Common Market Studies*, 43(2), 273-286.
- Glaeser, E. L., La Porta, R., Lopez-de Silanes, F., & Shleifer, A. (2004). Do institutions cause growth? *Journal of Economic Growth*, 9(3), 271-303.
- Gomes, J. F. (2014). *The health costs of ethnic distance: evidence from Sub-Saharan Africa*. ISER Working Paper Series No. 2014-33.
- Greenberg, J. H. (1956). The Measurement of Linguistic Diversity. *Language*, 32(1), 109-115.
- Grier, R. (1999). Colonial legacies and economic growth. *Public Choice* 98(3-4): 317-335.
- Guiso, L., Sapienza, P., & Zingales, L. (2006). Does culture affect economic outcomes?. *Journal of Economic Perspectives*, 20(2), 23-48.
- Ingham, K. (1956). *Reformers in India 1793-1833: An Account of the Work of Christian Missionaries on Behalf of Social Reform*. Cambridge: Cambridge University Press.
- Isichei, E. (1995). *A History of Christianity in Africa*. Grand Rapids, MI: William B. Eerdmans.
- Kelly, G.P. (2000a). Colonial Schools in Vietnam; Policy and Practice. in D. H. Kelly (ed.) *French Colonial Education: Essays on Vietnam and West Africa*. pp. 3-25 New York: AMS Press.
- Kelly, G.P. (2000b). Learning to be Marginal: Schooling in Interwar French West Africa. in D.H. Kelly (ed.) *French Colonial Education: Essays on Vietnam and West Africa*. pp. 189-208 New York: AMS Press.
- Kelly, G.P. (2000c). The Myth of Educational Planning: The Case of the Indochinese University, 1906-1938. in D.H. Kelly (ed.) *French Colonial Education: Essays on Vietnam and West Africa*. pp. 27-43. New York: AMS Press.
- Kelly, G.P. (2000d). The Presentation of Indigenous Society in the Schools of French West Africa and Indochina, 1918 to 1938. in D. H. Kelly (ed.) *French Colonial Education: Essays on Vietnam and West Africa*. pp. 235-262 New York: AMS Press.
- Kerwin, J., & Thornton, R. L. (2018). Making the Grade: The sensitivity of education program effectiveness to input choices and outcome measures. *Available at SSRN 3002723*.

- La Porta, R., Lopez-de Silanes, F., & Shleifer, A. (2008). The economic consequences of legal origins. *Journal of Economic Literature*, 46(2), 285-332.
- Laitin, D. D. (2000). What is a language community?. *American Journal of Political Science*, 142-155.
- Laitin, D. D., & Ramachandran, R. (2016). Language policy and human development. *American Political Science Review*, 110(3), 457-480.
- Laitin, D. D., Ramachandran, R., & Walter, S. L. (2019). The Legacy of Colonial Language Policies and Their Impact on Student Learning: Evidence from an Experimental Program in Cameroon. *Economic Development and Cultural Change*, 68(1), 239-272.
- Landa, D., & Duell, D. (2015). Social identity and electoral accountability. *American Journal of Political Science*, 59(3), 671-689.
- Levi, M., (1988). *Of Rule and Revenue*. California University Press, Berkeley.
- Licht, A. N., Goldschmidt, C., & Schwartz, S. H. (2007). Culture rules: The foundations of the rule of law and other norms of governance. *Journal of Comparative Economics*, 35(4), 659-688.
- Liu, A. H., & Pizzi, E. (2018). The language of economic growth: A new measure of linguistic heterogeneity. *British Journal of Political Science*, 48(4), 953-980.
- Luttmer, E. F., & Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives*, 28(4), 149-68.
- Manning, P. (1998). *Francophone Sub-Saharan Africa: 1880-1995*. Cambridge: Cambridge University Press.
- Melitz, J. (2008). Language and foreign trade. *European Economic Review*, 52(4), 667-699.
- Miguel, E., & Gugerty, M. K. (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics*, 89(11-12), 2325-2368.
- Moore, M. (2004). Revenues, state formation, and the quality of governance in developing countries. *International Political Science Review*, 25(3), 297-319.
- Moore, M. (2007). How does taxation affect the quality of government. *Institute of Development Studies Working paper series*, 280.

- Nunn, N. (2008). The long-term effects of Africa's slave trades. *Quarterly Journal of Economics* 123(1), 139-176.
- Ramirez, D., Engel, K. G., & Tang, T. S. (2008). Language interpreter utilization in the emergency department setting: a clinical review. *Journal of Health Care for the Poor and Underserved*, 19(2), 352-362.
- Reynal-Querol, M. (2002). Ethnicity, political systems, and civil wars. *Journal of Conflict Resolution*, 46(1), 29-54.
- Ricciuti, R., Savoia, A., & Sen, K. (2019). How do political institutions affect fiscal capacity? Explaining taxation in developing economies. *Journal of Institutional Economics*, 15(2), 351-380.
- Smith, T. (1978). A Comparative Study of French and British Decolonization. *Comparative Studies in Society and History*. 20: 70-102.
- Stock, J. H., & Yogo, M. (2002). Testing for weak instruments in linear iv regression. National Bureau of Economic Research Cambridge, Mass., USA.
- Skutnabb-Kangas, T., & Phillipson, R . (1989). “Mother Tongue’: The Theoretical and Sociopolitical Construction of a Concept.” In *Status and Function of Languages and Language Varieties*, edited by Ulrich A., 450–77. Berlin: De Gruyter.
- Sundkler, B., & Steed, C. (2000). *A History of the Church in Africa*. New York: Cambridge University Press.
- Tabellini, G. (2008). Institutions and culture. *Journal of the European Economic Association*, 6(2-3), 255-294.
- Tabellini, G. (2010). Culture and institutions: economic development in the regions of Europe. *Journal of the European Economic Association*, 8(4), 677-716.
- Taylor, S., & von Fintel, M. (2016). Estimating the impact of language of instruction in South African primary schools: A fixed effects approach. *Economics of Education Review*, 50, 75-89.
- Translators Without Borders (2015). Does translated health-related information lead to higher comprehension? A study of rural and urban Kenyans.” Accessed at <http://translatorswithoutborders.org/sites/all/themes/twb/pdfs/twbworimpactstudyfinal.pdf>.

## Table and figures

Table 1: Descriptive statistics

	count	mean	s.d.	min	max
Avg. Tax/GDP	146	19.96	11.00	0.80	46.18
Avg. executive constraints	146	4.92	1.89	1	7
Avg. WGI (vae)	146	-0.18	0.98	-2.04	1.61
ADOL (delta 0.5)	146	0.37	0.37	0	1
Ln (Population)	146	16.33	1.44	13.44	21.01
Ln (Avg. GDP)	146	14.03	2.98	7.31	22.36
Ln (Avg. Arable land area)	145	14.45	2.03	6.46	18.87
Avg. perc. oil reserves	146	4.27	9.84	0	47.61
Avg. external conflict	146	0.01	0.04	0	0.35
Linguistic fractionalization	146	0.32	0.21	0	0.73
British legal or.	146	0.29	0.45	0	1
Socialist legal or.	146	0.21	0.41	0	1
Avg. external conflict	146	0.01	0.04	0	0.35
Africa	146	0.30	0.46	0	1
America	146	0.17	0.38	0	1
Asia	146	0.28	0.45	0	1
Oceania	146	0.02	0.14	0	1
Europe	146	0.23	0.42	0	1
OECD	146	0.18	0.39	0	1

Table 2: OLS regressions

	Avg. Tax/GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Avg. WGI (vae)	7.437*** (0.785)	7.882*** (0.691)	7.879*** (0.696)	7.757*** (0.703)	7.591*** (0.714)	7.420*** (0.783)	7.438*** (0.799)	6.927*** (0.826)
British legal or.		-1.859 (1.307)	-1.866 (1.301)	-2.074 (1.315)	-1.841 (1.340)	-2.010 (1.316)	-1.957 (1.337)	-1.060 (1.390)
Socialist legal or.		8.726*** (1.645)	8.708*** (1.687)	8.569*** (1.701)	7.857*** (1.669)	7.678*** (1.625)	7.684*** (1.635)	7.249*** (1.619)
Ln (Population)			-0.0399 (0.408)	0.395 (0.515)	-0.618 (0.741)	-0.578 (0.730)	-0.574 (0.733)	-0.666 (0.719)
Ln (Avg. GDP)				-0.351 (0.276)	-0.376 (0.276)	-0.381 (0.275)	-0.380 (0.276)	-0.461* (0.270)
Ln (Avg. Arable land area)					0.931* (0.505)	0.896* (0.486)	0.916* (0.493)	1.064** (0.529)
Avg. perc. oil reserves						-0.0433 (0.0798)	-0.0357 (0.0844)	-0.0276 (0.0824)
Avg. external conflict							-6.675 (9.904)	-7.540 (9.714)
Linguistic fractionalization a/c for distance								-8.539*** (3.193)
Constant	21.37*** (0.669)	20.08*** (0.917)	20.73*** (6.852)	18.63*** (6.909)	22.12*** (7.060)	22.30*** (7.101)	21.90*** (7.290)	24.85*** (7.072)
Observations	147	147	147	147	146	146	146	146
R2	0.433	0.561	0.561	0.566	0.571	0.572	0.573	0.594

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 3: IV Regressions

PANEL A: First stage reg.	Avg. WGI (vae)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ADOL	-0.937*** (0.176)	-1.211*** (0.183)	-1.252*** (0.188)	-1.220*** (0.191)	-1.364*** (0.186)	-1.274*** (0.175)	-1.271*** (0.178)	-1.297*** (0.227)
Socialist legal origin		-0.577*** (0.221)	-0.620*** (0.217)	-0.622*** (0.216)	-0.801*** (0.214)	-0.865*** (0.197)	-0.863*** (0.197)	-0.870*** (0.202)
British legal origin		0.168 (0.165)	0.159 (0.168)	0.134 (0.169)	0.230 (0.159)	0.0773 (0.152)	0.0691 (0.156)	0.0645 (0.158)
Ln (Population)			-0.0746 (0.0482)	-0.0277 (0.0595)	-0.238*** (0.0790)	-0.186** (0.0810)	-0.186** (0.0814)	-0.189** (0.0830)
Ln (Avg. GDP)				-0.0367 (0.0298)	-0.0432 (0.0290)	-0.0395 (0.0275)	-0.0397 (0.0276)	-0.0388 (0.0283)
Ln (Avg. Arable land area)					0.196*** (0.0438)	0.147*** (0.0443)	0.143*** (0.0448)	0.144*** (0.0451)
Avg. perc. oil reserves						-0.0346*** (0.0072)	-0.0356*** (0.0076)	-0.0358*** (0.0076)
Avg. external conflict							1.004 (1.297)	1.011 (1.296)
Linguistic fractionalization a/c for distance								0.0770 (0.403)
Constant	0.167 (0.119)	0.344** (0.138)	1.588* (0.825)	1.333 (0.840)	2.101** (0.837)	2.082** (0.832)	2.136** (0.844)	2.146** (0.846)
PANEL B: Reduced form	Avg. Tax/GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Avg. WGI (vae)	13.19*** (2.397)	8.888*** (1.463)	8.890*** (1.416)	8.685*** (1.480)	9.240*** (1.470)	9.295*** (1.539)	9.320*** (1.557)	7.071*** (2.152)
Socialist legal origin		8.982*** (1.820)	8.981*** (1.853)	8.836*** (1.870)	8.452*** (1.948)	8.550*** (2.001)	8.556*** (2.012)	7.320*** (2.048)
British legal origin		-1.874 (1.285)	-1.875 (1.277)	-2.049 (1.280)	-1.903 (1.314)	-1.787 (1.311)	-1.715 (1.340)	-1.061 (1.344)
Ln (Population)			-0.00437 (0.419)	0.359 (0.523)	-0.523 (0.740)	-0.553 (0.742)	-0.548 (0.742)	-0.662 (0.695)
Ln (Avg. GDP)				-0.297 (0.297)	-0.263 (0.309)	-0.264 (0.309)	-0.261 (0.310)	-0.451 (0.318)
Ln (Avg. Arable land area)					0.762 (0.494)	0.792 (0.501)	0.820 (0.506)	1.054** (0.527)
Avg. perc. oil reserves						0.0290 (0.0840)	0.0394 (0.0901)	-0.0224 (0.0902)
Avg. external conflict							-9.099 (12.22)	-7.695 (9.542)
Linguistic fractionalization a/c for distance								-8.369** (4.132)
Constant	22.37*** (0.972)	20.20*** (0.881)	20.27*** (7.027)	18.54*** (6.979)	21.59*** (7.374)	21.49*** (7.434)	20.95*** (7.662)	24.72*** (7.237)
Observations	147	147	147	147	146	146	146	146
F-Stat	28.45	43.89	44.31	40.63	53.59	53.21	50.73	32.56
R2	0.174	0.553	0.553	0.560	0.551	0.551	0.551	0.594

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$



Table 4: Conley Test (Local to zero approach)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: CI 2sigma=0								
Avg. WGI (vae)	13.19*** (2.397)	8.888*** (1.463)	8.890*** (1.416)	8.685*** (1.480)	9.240*** (1.470)	9.295*** (1.539)	9.320*** (1.557)	7.071*** (2.152)
Lower bound	8.494	6.020	6.115	5.784	6.359	6.280	6.267	2.853
Upper bound	17.89	11.76	11.67	11.59	12.12	12.31	12.37	11.29
Panel B: CI 2sigma=10								
Avg. WGI (vae)	13.19*** (2.499)	8.888*** (1.563)	8.890*** (1.512)	8.685*** (1.576)	9.240*** (1.546)	9.295*** (1.621)	9.320*** (1.639)	7.071*** (2.206)
Lower bound	8.294	5.824	5.926	5.596	6.210	6.119	6.108	2.746
Upper bound	18.09	11.95	11.85	11.77	12.27	12.47	12.53	11.40
Panel C: CI 2sigma=20								
Avg. WGI (vae)	13.19*** (2.782)	8.888*** (1.831)	8.890*** (1.769)	8.685*** (1.834)	9.240*** (1.755)	9.295*** (1.845)	9.320*** (1.861)	7.071*** (2.361)
Lower bound	7.738	5.300	5.423	5.092	5.800	5.679	5.673	2.443
Upper bound	18.64	12.48	12.36	12.28	12.68	12.91	12.97	11.70
Panel D: CI 2sigma=30								
Avg. WGI (vae)	13.19*** (3.200)	8.888*** (2.206)	8.890*** (2.130)	8.685*** (2.197)	9.240*** (2.057)	9.295*** (2.168)	9.320*** (2.181)	7.071*** (2.599)
Lower bound	6.920	4.565	4.716	4.380	5.209	5.046	5.044	1.976
Upper bound	19.46	13.21	13.06	12.99	13.27	13.54	13.59	12.17
Panel E: CI 2sigma=40								
Avg. WGI (vae)	13.19*** (3.706)	8.888*** (2.643)	8.890*** (2.550)	8.685*** (2.622)	9.240*** (2.416)	9.295*** (2.552)	9.320*** (2.563)	7.071** (2.900)
Lower bound	5.928	3.709	3.892	3.547	4.504	4.293	4.295	1.388
Upper bound	20.46	14.07	13.89	13.82	13.98	14.30	14.34	12.75
Panel F: CI 2sigma=50								
Avg. WGI (vae)	13.19*** (4.270)	8.888*** (3.116)	8.890*** (3.005)	8.685*** (3.083)	9.240*** (2.812)	9.295*** (2.974)	9.320*** (2.984)	7.071** (3.245)
Lower bound	4.823	2.782	3.000	2.642	3.728	3.466	3.471	0.710
Upper bound	21.56	15.00	14.78	14.73	14.75	15.12	15.17	13.43
Panel G: CI 2sigma=60								
Avg. WGI (vae)	13.19*** (4.871)	8.888** (3.611)	8.890** (3.482)	8.685** (3.567)	9.240*** (3.231)	9.295*** (3.420)	9.320*** (3.428)	7.071* (3.623)
Lower bound	3.645	1.812	2.066	1.693	2.908	2.592	2.600	-0.0307
Upper bound	22.74	15.97	15.72	15.68	15.57	16.00	16.04	14.17
Panel H: CI 2sigma=70								
Avg. WGI (vae)	13.19** (5.497)	8.888** (4.120)	8.890** (3.972)	8.685** (4.066)	9.240** (3.664)	9.295** (3.882)	9.320** (3.889)	7.071* (4.025)
Lower bound	2.417	0.814	1.105	0.717	2.058	1.688	1.698	-0.817
Upper bound	23.97	16.96	16.68	16.65	16.42	16.90	16.94	14.96
Panel I: CI 2sigma=80								
Avg. WGI (vae)	13.19** (6.141)	8.888* (4.638)	8.890** (4.472)	8.685* (4.574)	9.240** (4.108)	9.295** (4.354)	9.320** (4.360)	7.071 (4.443)
Lower bound	1.155	-0.202	0.126	-0.279	1.189	0.762	0.775	-1.636
Upper bound	25.23	17.98	17.65	17.65	17.29	17.83	17.86	15.78

CI stands for "Confidence intervals". In Column (1) there are no additional controls present in the regression. Column (2) includes controls for legal origins (Socialist and British). Column (3) includes controls for legal origins (Socialist and British) and the population in logarithm. Column (4) includes controls for legal origins (Socialist and British), the population in logarithm, and the average GDP in logarithm. Column (5) includes controls for legal origins (Socialist and British), the population in logarithm, the average GDP in logarithm, and the average arable land area in logarithm. Column (6) includes controls for legal origins (Socialist and British), the population in logarithm, the average GDP in logarithm, the average arable land area in logarithm, and the average percentage of oil reserves. Column (7) includes controls for legal origins (Socialist and British), the population in logarithm, the average GDP in logarithm, the average arable land area in logarithm, the average percentage of oil reserves and the average period in external wars. Column (8) includes controls for legal origins (Socialist and British), the population in logarithm, the average GDP in logarithm, the average arable land area in logarithm, the average percentage of oil reserves, the average period in external wars and the linguistic fractionalization  $a/c$  for distance. Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 5: IV regressions using Executive Constraints

PANEL A: First stage reg.	Avg. executive constraints							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ADOL	-1.627*** (0.349)	-1.952*** (0.358)	-1.974*** (0.370)	-1.961*** (0.371)	-2.243*** (0.358)	-2.018*** (0.309)	-2.008*** (0.316)	-1.869*** (0.446)
Socialist legal origin		-0.623 (0.441)	-0.645 (0.445)	-0.644 (0.446)	-0.995** (0.436)	-1.152*** (0.376)	-1.147*** (0.375)	-1.111*** (0.386)
British legal origin		0.305 (0.341)	0.300 (0.344)	0.290 (0.345)	0.488 (0.323)	0.105 (0.290)	0.0819 (0.296)	0.106 (0.304)
Ln (Population)			-0.0404 (0.104)	-0.0210 (0.124)	-0.434** (0.171)	-0.305* (0.174)	-0.305* (0.176)	-0.290 (0.179)
Ln (Avg. GDP)				-0.0153 (0.0628)	-0.0304 (0.0607)	-0.0216 (0.0536)	-0.0221 (0.0535)	-0.0272 (0.0565)
Ln (Avg. Arable land area)					0.388*** (0.0978)	0.266** (0.106)	0.256** (0.109)	0.252** (0.109)
Avg. perc. oil reserves						-0.0867*** (0.0144)	-0.0897*** (0.0140)	-0.0888*** (0.0142)
Avg. external conflict							2.882 (2.432)	2.841 (2.448)
Linguistic fractionalization a/c for distance								-0.407 (0.909)
Constant	5.513*** (0.216)	5.676*** (0.250)	6.351*** (1.796)	6.247*** (1.798)	7.733*** (1.795)	7.694*** (1.717)	7.849*** (1.752)	7.801*** (1.764)
PANEL B: Reduced form	Avg. Tax/GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Avg. executive constraints	7.469*** (1.600)	5.481*** (1.099)	5.597*** (1.101)	5.379*** (1.105)	5.600*** (1.052)	5.849*** (1.115)	5.877*** (1.134)	4.915*** (1.484)
Socialist legal origin		6.983*** (2.310)	6.833*** (2.404)	6.697*** (2.366)	6.443*** (2.471)	7.081*** (2.533)	7.085*** (2.542)	6.474*** (2.439)
British legal origin		-2.061 (1.693)	-2.144 (1.677)	-2.434 (1.656)	-2.503 (1.699)	-1.674 (1.728)	-1.540 (1.769)	-1.125 (1.673)
Ln (Population)			-0.411 (0.524)	0.227 (0.718)	-0.289 (1.034)	-0.501 (1.085)	-0.490 (1.091)	-0.577 (1.000)
Ln (Avg. GDP)				-0.510 (0.372)	-0.471 (0.389)	-0.485 (0.392)	-0.481 (0.392)	-0.573 (0.374)
Ln (Avg. Arable land area)					0.390 (0.640)	0.596 (0.710)	0.647 (0.724)	0.834 (0.714)
Avg. perc. oil reserves						0.215* (0.121)	0.235* (0.124)	0.161 (0.137)
Avg. external conflict							-16.78 (16.35)	-14.57 (14.37)
Linguistic fractionalization a/c for distance								-5.672 (5.024)
Constant	-16.77** (7.882)	-7.885 (5.588)	-1.679 (10.01)	-3.759 (9.912)	-2.588 (10.59)	-4.417 (10.88)	-5.537 (11.36)	1.344 (12.88)
Observations	146	146	146	146	145	145	145	145
F-stat	21.67	29.72	28.49	28.00	39.20	42.71	40.44	17.60
R2	-0.248	0.229	0.216	0.257	0.222	0.220	0.220	0.332

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 6: IV Regressions including *Federal*

PANEL A: First stage reg.	Avg. executive constraints							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ADOL	-0.933*** (0.176)	-1.197*** (0.186)	-1.219*** (0.189)	-1.190*** (0.193)	-1.342*** (0.190)	-1.223*** (0.176)	-1.224*** (0.177)	-1.236*** (0.229)
Federal	0.219 (0.250)	0.0452 (0.248)	0.162 (0.271)	0.158 (0.265)	0.0923 (0.254)	0.258 (0.225)	0.239 (0.239)	0.235 (0.244)
Socialist legal or.		-0.527** (0.222)	-0.564** (0.218)	-0.568** (0.218)	-0.752*** (0.218)	-0.802*** (0.196)	-0.803*** (0.197)	-0.806*** (0.203)
British legal or.		0.160 (0.170)	0.131 (0.173)	0.108 (0.173)	0.213 (0.166)	0.0247 (0.151)	0.0243 (0.152)	0.0230 (0.154)
Ln(Population)			-0.0790 (0.0525)	-0.0350 (0.0656)	-0.236*** (0.0822)	-0.193** (0.0823)	-0.192** (0.0829)	-0.193** (0.0839)
Ln(Avg. GDP)				-0.0345 (0.0303)	-0.0411 (0.0295)	-0.0366 (0.0281)	-0.0368 (0.0283)	-0.0364 (0.0286)
Ln(Avg. Arable land area)					0.192*** (0.0439)	0.139*** (0.0435)	0.138*** (0.0439)	0.138*** (0.0443)
Avg. perc. oil reserves						-0.036*** (0.00714)	-0.036*** (0.00752)	-0.037*** (0.00752)
Avg. external conflict							0.519 (1.389)	0.529 (1.392)
Linguistic fractionalization a/c for distance								0.0329 (0.406)
Constant	0.150 (0.123)	0.334** (0.142)	1.633* (0.880)	1.394 (0.907)	2.085** (0.898)	2.230** (0.866)	2.239** (0.870)	2.239** (0.873)
PANEL B: Reduced form	Avg. Tax/GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Avg. WGI	13.48*** (2.446)	8.843*** (1.503)	8.833*** (1.473)	8.638*** (1.534)	9.206*** (1.518)	9.276*** (1.626)	9.258*** (1.633)	6.750*** (2.307)
Federal	-3.659 (2.349)	0.0844 (1.688)	0.0210 (1.860)	0.0200 (1.831)	-0.148 (1.871)	-0.284 (2.008)	0.0548 (2.017)	1.615 (2.196)
Socialist legal or.		9.255*** (1.815)	9.269*** (1.849)	9.126*** (1.867)	8.712*** (1.950)	8.804*** (2.010)	8.801*** (2.015)	7.532*** (2.034)
British legal or.		-1.886 (1.339)	-1.868 (1.342)	-2.035 (1.348)	-1.865 (1.381)	-1.732 (1.420)	-1.723 (1.428)	-1.329 (1.393)
Ln(Population)			0.0431 (0.457)	0.390 (0.545)	-0.466 (0.741)	-0.484 (0.742)	-0.506 (0.739)	-0.740 (0.705)
Ln(Avg. GDP)				-0.284 (0.296)	-0.252 (0.307)	-0.253 (0.308)	-0.251 (0.309)	-0.444 (0.315)
Ln(Avg. Arable land area)					0.748 (0.489)	0.776 (0.500)	0.805 (0.501)	1.053** (0.513)
Avg. perc. oil reserves						0.0282 (0.0893)	0.0359 (0.0933)	-0.0389 (0.0945)
Avg. external conflict							-9.366 (12.39)	-10.71 (9.851)
Linguistic fractionalization a/c for distance								-8.599** (4.193)
F-Stat	27.95	41.23	41.42	38.17	49.81	48.26	47.89	29.13
Observations	146	146	146	146	145	145	145	145
R2	0.149	0.552	0.552	0.559	0.549	0.548	0.549	0.595

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 7: IV Regressions excluding the non-former colonies

PANEL A: First stage reg.	Avg. WGI (vae)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ADOL (delta 0.5)	-0.745*** (0.183)	-0.979*** (0.193)	-0.991*** (0.196)	-0.959*** (0.201)	-1.108*** (0.198)	-1.040*** (0.185)	-1.034*** (0.191)	-1.092*** (0.245)
Socialist legal or.		-0.278 (0.231)	-0.309 (0.233)	-0.307 (0.232)	-0.493** (0.235)	-0.567*** (0.216)	-0.565** (0.216)	-0.582** (0.225)
British legal or.		0.377** (0.165)	0.370** (0.167)	0.352** (0.168)	0.445*** (0.154)	0.279* (0.158)	0.270* (0.163)	0.262 (0.164)
Ln (Population)			-0.0478 (0.0507)	-0.00210 (0.0624)	-0.192** (0.0829)	-0.143* (0.0835)	-0.143* (0.0840)	-0.151* (0.0868)
Ln (Avg. GDP)				-0.0347 (0.0327)	-0.0429 (0.0317)	-0.0453 (0.0305)	-0.0455 (0.0307)	-0.0432 (0.0314)
Ln (Avg. Arable land area)					0.177*** (0.0441)	0.136*** (0.0446)	0.133*** (0.0453)	0.135*** (0.0458)
Avg. perc. oil reserves						-0.0276*** (0.00723)	-0.0286*** (0.00788)	-0.0289*** (0.00783)
Avg. external conflict							0.824 (1.301)	0.832 (1.293)
Linguistic fractionalization a/c for distance								0.161 (0.411)
Constant	0.0262 (0.126)	0.0761 (0.149)	0.863 (0.856)	0.597 (0.857)	1.340 (0.883)	1.327 (0.866)	1.383 (0.881)	1.418 (0.888)
PANEL B: Reduced form	Avg. Tax/GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Avg. WGI (vae)	12.79*** (3.228)	6.846*** (1.887)	6.982*** (1.840)	6.651*** (1.914)	7.475*** (1.828)	7.477*** (1.903)	7.517*** (1.931)	4.915* (2.850)
Socialist legal or.		9.584*** (1.687)	9.290*** (1.775)	9.213*** (1.776)	8.558*** (1.819)	8.560*** (1.857)	8.565*** (1.864)	7.881*** (1.827)
British legal or.		-0.102 (1.567)	-0.229 (1.510)	-0.284 (1.506)	-0.311 (1.515)	-0.308 (1.442)	-0.259 (1.460)	0.824 (1.762)
Ln (Population)			-0.515 (0.432)	-0.0627 (0.531)	-1.032 (0.731)	-1.033 (0.721)	-1.022 (0.725)	-1.041 (0.706)
Ln (Avg. GDP)				-0.355 (0.304)	-0.323 (0.325)	-0.323 (0.328)	-0.320 (0.329)	-0.542 (0.353)
Ln (Avg. Arable land area)					0.822* (0.493)	0.822* (0.489)	0.838* (0.493)	1.066* (0.544)
Avg. perc. oil reserves						0.000569 (0.0886)	0.00851 (0.0965)	-0.0541 (0.0964)
Avg. external conflict							-5.595 (11.21)	-3.838 (9.199)
Linguistic fractionalization a/c for distance								-7.483* (4.308)
Constant	22.22*** (1.339)	18.37*** (1.120)	26.84*** (7.100)	24.40*** (7.068)	28.28*** (7.141)	28.28*** (7.187)	27.85*** (7.535)	29.83*** (7.149)
F-Stat	16.62	25.73	25.63	22.76	31.40	31.51	29.39	19.84
Observations	127	127	127	127	126	126	126	126
R2	0.00196	0.499	0.504	0.512	0.507	0.507	0.507	0.537

Robust standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

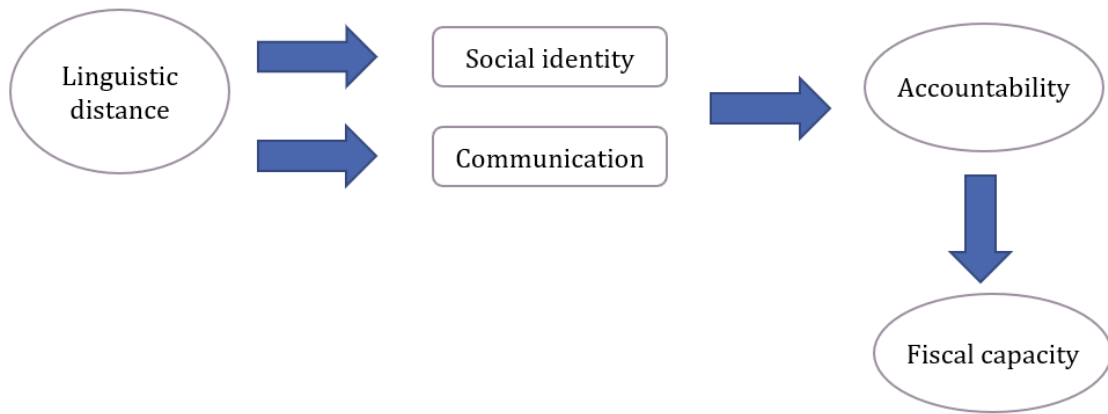


Figure 1: *The effect of linguistic distance*

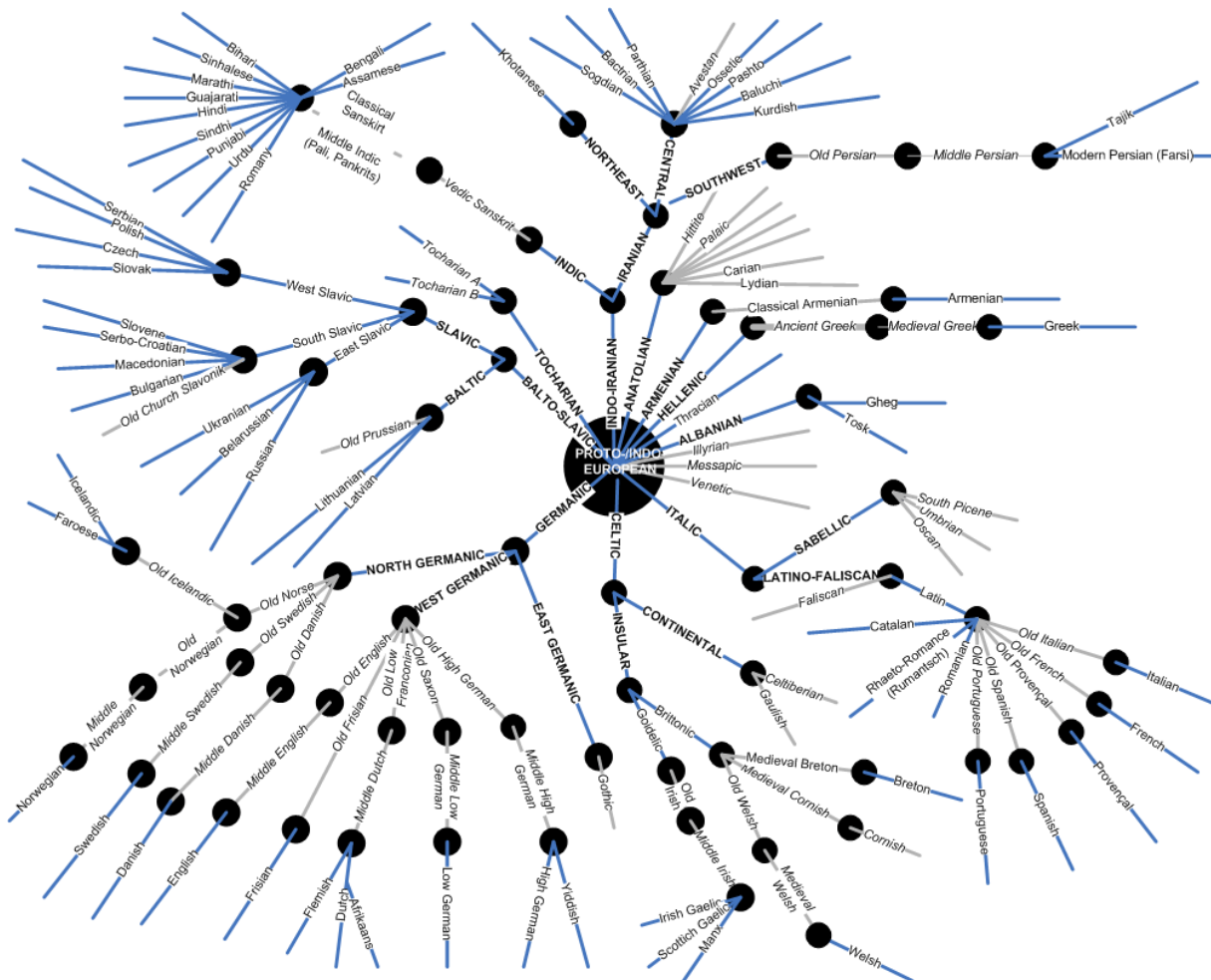


Figure 2: *The Proto-Indo-European Language Tree* (Source: Intersol, Inc. Translation and Localization<sup>18</sup>)

<sup>18</sup> <https://intersolinc.wordpress.com/2015/01/27/differences-and-similarities-between-the-tree-of-life-and-the-tree-of-languages/>

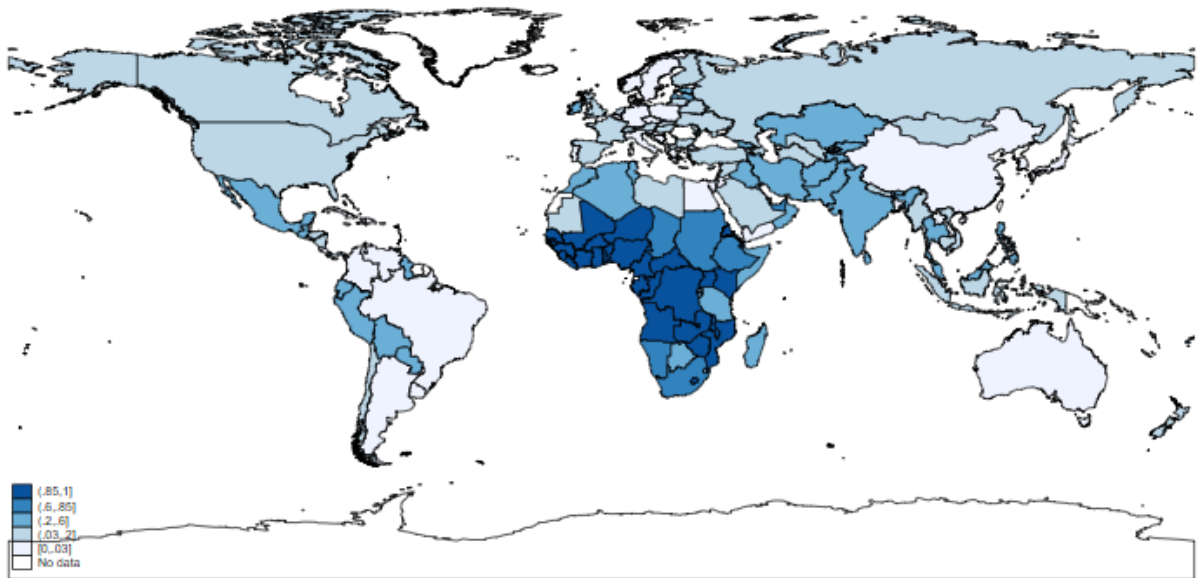


Figure 3: ADOL across the World

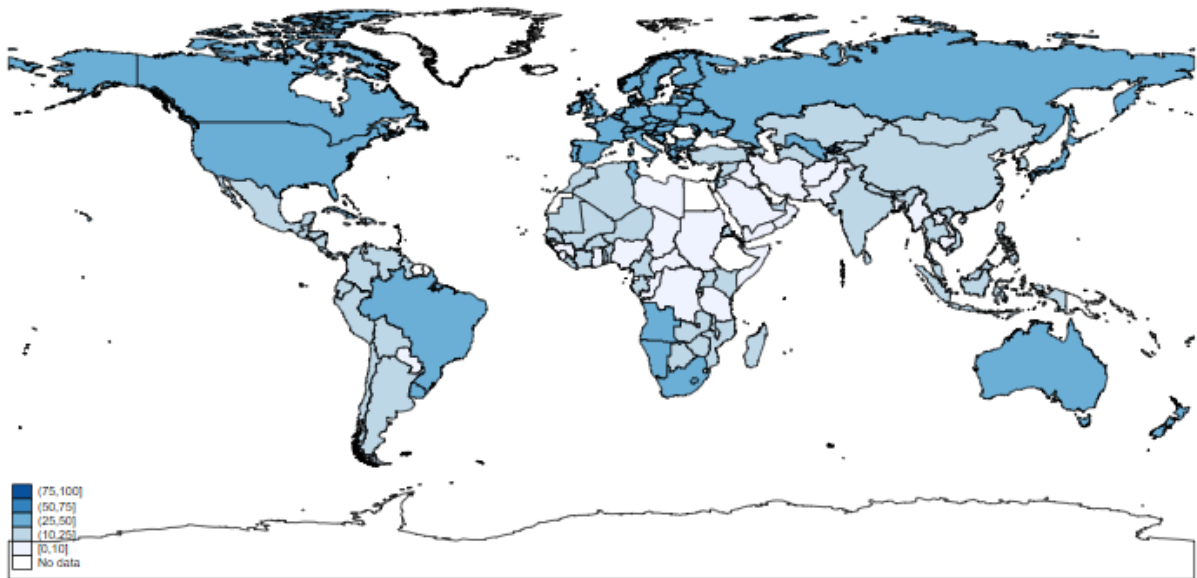


Figure 4: Fiscal Capacity across the World



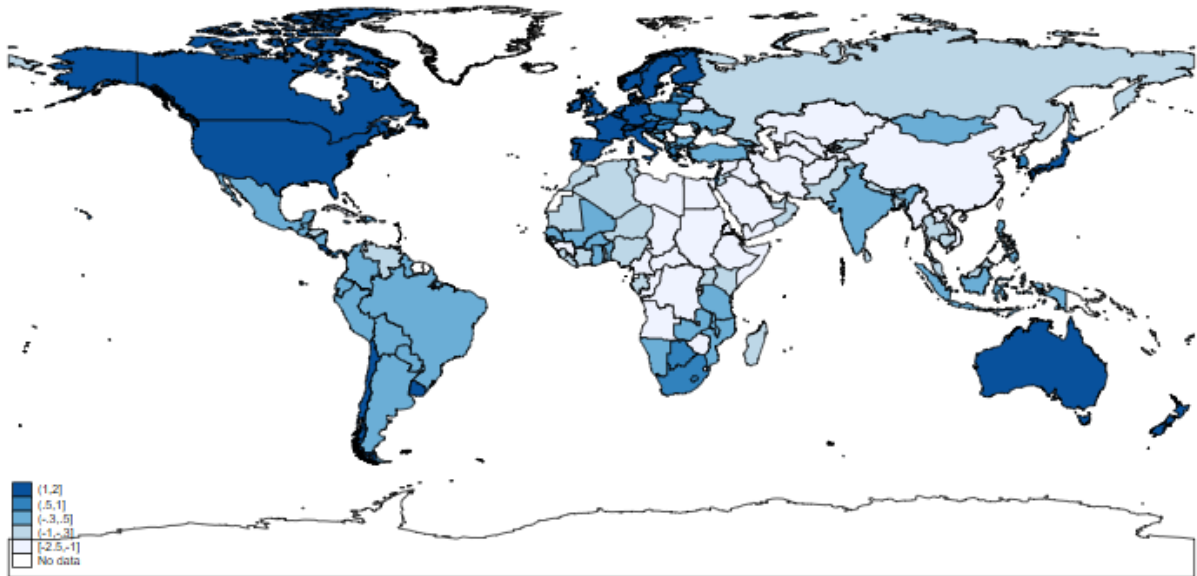


Figure 5: Accountability index across the World