The Imperium of the Colonial Tongue?
Evidence on Language Policy Preferences in Zambia

Rajesh
Ramachandran
Monash University
Malaysia

Christopher
Rauh
University of
Cambridge

Abstract

Sub-Saharan Africa stands out as a part of the world that primarily uses, as its official languages, former colonial languages that are neither spoken at home nor in the community. In this paper, we elicit preferences for colonial versus local languages and analyze the role of perceived costs and returns to different languages. In order to do so, we elicit beliefs about the effects of hypothetical changes to Zambia's language policy on schooling outcomes, income, and social cohesion. Our results show overwhelming support for the use of the colonial language to act as official. Looking at the determinants, we find that fears of being disadvantaged by the installation of another group's language, high perceived costs of learning in another group's language, and lack of association between retaining the elite language and socioeconomic inequality as crucial factors in affecting preferences over language policies.

Reference Details

The Imperium of the Colonial Tongue? Evidence on Language Policy Preferences in Zambia

Rajesh Ramachandran and Christopher Rauh

December 2021

Abstract

Sub-Saharan Africa stands out as a part of the world that primarily uses, as its official languages, former colonial languages that are neither spoken at home nor in the community. In this paper, we elicit preferences for colonial versus local languages and analyze the role of perceived costs and returns to different languages. In order to do so, we elicit beliefs about the effects of hypothetical changes to Zambia's language policy on schooling outcomes, income, and social cohesion. Our results show overwhelming support for the use of the colonial language to act as official. Looking at the determinants, we find that fears of being disadvantaged by the installation of another group's language, high perceived costs of learning in another group's language, and lack of association between retaining the elite language and socioeconomic inequality as crucial factors in affecting preferences over language policies.

Keywords: ethnic cleavages, language policy, stigmatization

JEL: H41, I28, Z13

1 Introduction

One of the institutional features distinguishing Sub-Saharan Africa (SSA) from the rest of the world is the use of languages that are typically not spoken at home or in the community as the principal language of education, business and public administration, including the functioning of higher courts (Bamgbose, 1991).[1] In the context of Sub-Saharan Africa these refer to the former colonial languages, namely, English, French, Portuguese and Spanish.[2] The primacy of the former colonial languages in formal domains is highlighted by the fact that not a single country in SSA provides secondary schooling or higher education in a local language. In fact, only Eritrea, Ethiopia and Tanzania offer the entire span of primary schooling in a non-colonial language (Albaugh, 2014).

The choice of the former colonial language to act as official, has the benefit of enabling easier integration into the global economy (Ku and Zussman, 2010; Egger and Lassmann, 2012). This is also supported by micro evidence that suggests high labor market returns to knowledge of the former colonial language in postcolonial states (Angrist and Lavy, 1997; Azam et al., 2013). This, however, comes at the higher cost of both obtaining the necessary language skills and forming human capital through the use of a non-indigenous language, as evidenced in both postcolonial states (Eriksson, 2014; Ramachandran, 2017; Laitin and Ramachandran, 2016 Taylor and von Fintel, 2016), and from immigrant experiences in the industrialized world (Bleakley and Chin, 2004; Dustmann et al., 2010; Isphording and Otten, 2013). A further key aspect is related to the link between language and identity. Based on the European experience, the existence of a common shared language has been highlighted as a key factor in the process of creating 'imagined communities', and nation building (Anderson, 2006).

The use of a widely spoken and understood language, as in the industrialized world, can help reduce transaction costs associated with navigating the education, legal and political system. Table 1 shows the limited spread of official languages in SSA. It draws from the 4th round of the Afrobarometer conducted in 2008 and shows the self-reported ability to "speak well" in the official colonial language, as well as the language of the largest linguistic group, for a set of nineteen countries. The data shows limited spread in the knowledge of the official language, despite 60 years of use as the principal language of state institutions including primary education. Across the nineteen countries, on an average, 47 percent of respondents report being able to speak well in the official colonial language. On the other hand, 66 percent report being able to speak the largest indigenous language. Moreover, on average 3.17 languages are reported to be spoken well by individuals living across these 19

---

[1]North America and Latin America also use former colonial languages as official languages. However, the former colonial languages are spoken by the majority of the populations at home, as well as employed for societal communication.

[2]There is a small set of elites who do employ the former colonial languages in the home environment; for instance, data from the Southern and Eastern Africa Consortium for Monitoring Educational Quality (2004) show that 22 percent of 6th grade children report speaking English at home "often."

countries, and at least two languages are spoken on average.

The question that arises is what explains the continued use of colonial languages in SSA? Do differences in *perceived* costs and returns to using the global language play a role? Or, if not just monetary costs and benefits, then what other factors affect preferences over official language choice in SSA? To enable an understanding of the factors influencing language choice in SSA, we explore the importance of two other factors besides the perceived cost and returns to obtaining human capital in local languages as compared to the colonial language.

The first issue arises due to the negative attitudes individuals hold about the suitability of indigenous languages as vehicles for science, and knowledge creation and dissemination in society. These negative attitudes arise in part due to the indigenous languages in SSA remaining largely oral languages, with little investment since post-independence to promote standardization, and consequently having not been employed in formal domains (Albaugh, 2014). The experiences of Europe and the Indian subcontinent show the challenges that erstwhile spoken tongues faced in overcoming the label of corrupted vulgar speech, in comparison to the 'perfection and purity' of Latin and Sanskrit, before finally replacing them in all formal domains (Burke et al., 2004; Pollock et al., 2006). Thus, it is crucial to understand whether individuals tend to hold negative attitudes about the suitability of indigenous languages for use in formal domains resulting in preference for the former colonial language (Bourdieu, 1991).

The second motivation is related to the extent of linguistic diversity and the resulting coordination problem (Laitin and Ramachandran, 2020). SSA is characterized by extremely high levels of linguistic diversity (Easterly and Levine, 1997; Michalopoulos, 2012); moreover, a history of state formation where country frontiers were determined by arbitrary border drawing by the colonial powers, resulting in splitting of ethnolinguistic groups across country borders further augmenting levels of linguistic diversity (Asiwaju, 1985; Alesina et al., 2011). In the presence of multiple language groups, the fear that a group whose language is chosen discriminates against other groups could be a reason why individuals in multilingual societies exhibit a preference for the colonial language despite the high barriers it imposes to participating in the economic and political life. The role of competing group claims is further complicated by the relative salience of the category of ethnicity or class in society. The preference for the former colonial language might be more relevant for societies where ethnicity is the central cleavage. However, in societies where class forms the dividing line, indigenous languages might be preferred as they do not favor the (linguistic) elites.

To gauge the relevance of the outlined mechanisms, we design and implement a survey collecting elicited beliefs about the effects of a hypothetical change in language policy. We ask nearly 200 respondents about their expectations concerning outcomes of hypothetical agents under scenarios of different language policies (see Section 2.4 for more details). The survey is implemented in an urban and rural site in Zambia.

The data shows overwhelming preference for the colonial language to act as official in Zambia. In our sample, 71 percent of the individuals specify English as their only preferred official language, whereas a mere 19 percent report a preference for only an indigenous language, and 10 percent prefer the use of both English and indigenous language(s).

The beliefs of the cost and returns to installing English or an indigenous language to act as official, and its correlation with preference over official language choice reveal several patterns. First, individuals seem to systematically believe that learning in another group's indigenous language is more costly than learning in English, though data on language repertoires suggest the contrary; in Zambia, the knowledge of the largest indigenous language, Bemba, is much more widespread among non-Bemba speakers than knowledge of English though Bemba is not taught in any educational institutions; 48 percent speak Bemba as compared to 36 percent reporting speaking English (see Table 1). Second, less than a third of the sample report that a child would find it easier to learn Math in an indigenous language as compared to in English. However, individuals who believe so are almost 20 percentage points more in favor of installing an indigenous language as official. Thus, individuals do not seem to systematically believe that the use of indigenous languages will reduce costs of human capital acquisition. However, data from The SACMEQ-III show that only 28% of students can read for meaning or attain higher reading skills, and as low as 8% of students achieve beginning numeracy or higher mathematical skills, by the end of Grade 6.[3]

Turning to labor market returns, surprisingly the individuals do not seem to expect that using English as the official language will result in higher earnings, as compared to a situation where an indigenous language is used both in education and other formal domains. However, a situation where education is provided in an indigenous language but administration and jobs remain in English are seen to reduce expected earnings by almost a third. The fact that most language policy proposals that have been debated have tried to address only the language of education, and not explicitly tackled the question of language of administration and jobs, might underlie the limited support for indigenous languages in SSA. Put differently, unless indigenous languages have an important role in the formal labor market, the support is likely to remain low as individuals then believe that formation of human capital through the indigenous medium might not result in commensurate returns on the labor market (Mufwene and Vigouroux, 2008).

The perceived costs and returns, however, are not the only factors driving preferences over official language choice. Turning to the role of attitudes, we find that a significant share of individuals believe that indigenous languages are not capable of fulfilling the role of an official language. More specifically, 40 percent of the sample is seen to agree with the statement that countries require English, French, or Portuguese as the language of education and government to be economically successful. Similarly, a third of our sample agrees with the statements that English is the language of the intelligent people and English is the only

---

[3]See http://www.sacmeq.org/?q=sacmeq-members/zambia/reading-and-math-achievement-levels.

3

language in which knowledge is useful. These attitudes might be due to misplaced beliefs about policy choices of other economically successfully countries. 58 and 31 percent of the sample believe that Sweden and South Korea, respectively, employ English or French as the language of education and government. However, our data does not exhibit any systematic links between these attitudes and preferences over official language choice.

Finally, turning to the third class of explanations, we find that ethnic and class cleavages play an important role in the decision making of individuals. We find not only that more than four-fifths of the sample concurs with the statement that a group whose language is not chosen will be disadvantaged and discriminated in society, but moreover, that these beliefs are seen to be predictive of a preference for the colonial tongue to act as official. We find more than 90 percent of the sample agrees that the gap between the rich and poor is a problem. As Piper et al. (2016, pg. 782) note: "Naming a local language as an official language of instruction—or, conversely, not doing so— can be a powerful political decision. Some stakeholders fear that promoting local languages might deepen ethnic divisions, despite evidence that one-language policies do not guarantee cohesion across ethnic groups."

Finally, individuals who believe that the use of indigenous languages in education and government administration would reduce the gap between rich and poor are almost 30 percentage points more likely to be in favor of their use. However, only a small share (22 percent) believe that use of indigenous languages is a policy tool to combat socioeconomic inequality generated by linguistic capital. Thus, both ethnic and class concerns are very prevalent in society, though only the former seem to have a strong influence on language policy preferences.

There is a body of literature exploring the choice of official language in a plurilingual polity (Thorburn, 1971; Vaillancourt, 1983; Green, 1987; Pool, 1987). One key question has been whether a fair and efficient language policy is feasible, with Pool (1991) answering the question in the affirmative. Our contribution is to highlight how the demand for fairness in one dimension, namely ensuring group parity, might result in overlooking the implications for class parity. The paper also relates to the work of Laitin and Ramachandran (2020) who show that linguistically diverse states are more likely to exclusively install the former colonial language to act as official, and that reliance on it has large negative effects on human capital outcomes. They argue that multiple language groups create competing claims and thus countries retain the colonial language to act as an 'ethnically neutral' language to assuage competing group interests. We provide explicit evidence for this channel by using microdata on individuals beliefs on the role of ethnolinguistic competition and preferences over language policy.

Our work also relates to the literature which has tried to measure preference for the use of the colonial language and how it relates to attitudes regarding the suitability of the indigenous varieties for use in formal domains, as well as perceived returns to using the

4

'global' colonial language. Using data primarily from Nigeria, Adegbija (1994) provides an overview of language attitudes in SSA, and finds that individuals believe that colonial languages, compared to indigenous languages, are more suitable for use in the formal domain. Skattum (2008) discusses the phenomenon of *diglossia*, i.e. the difference in prestige and usage between French and local languages in the context of the former French colonies. French is considered to be a 'high' language or the language of prestige and suitable for education, government and business. On the other hand, local languages are considered to be the 'low' languages and more suitable for informal daily functions such as interacting with friends and family. As Skattum (2008, pg. 174) observes "This functional difference both stems and is reflected in people's attitudes, and to a large extent explains why ordinary people as well as government officials harbor negative attitudes towards their own languages - be it languages of education or written languages in general." In the context of Zimbabwe, Chiwome et al. (1992) and Mparutsa et al. (1992) show that students display a strong preference for the continued use of English as the medium of instruction, with a lot of students stressing the importance of English for international communication. Analyzing the roles of Swahili and English in urban Kenya, Mukhwana (2014) finds that the respondents clearly reject Swahili as a means to achieving social mobility, though not its role as the language of social interaction. Similarly, Laitin (1994) in the context of Ghana finds strong preference for the colonial language with more than a majority of the respondents favoring English as the dominant language for official communication, and as the medium of instruction in primary schools. The respondents highlight the importance of English for formal domains, especially in determining job opportunities, and highlight the insufficiency of Ghanaian languages for science and official communication. Piper et al. (2016) evaluate a mother-tongue instruction program in primary schools in Kenya and are confronted with some of the concerns we document. Teachers and parents, anxious about about children's long-run outcomes, such as exam performance, access to higher education access, and employment prospects, as well as the high prestige they accord to European languages, resisted during the implementation of the mother-tongue program.

More generally, our paper is related to a body of work that tries to situate language policies of postcolonial states with their history of colonialism. As Phillipson (1992, pg. 127-8) notes, "Despite differences in the articulation of policies in the French and British empires, what they had in common was the low status accorded to dominated languages, whether these were ignored or used in the early years of education; a very small proportion of the population in formal education, especially after the lower grades; local traditions and educational practice being ignored; unsuitable education being provided; an explicit policy of "civilizing the natives," and the master language being attributed civilizing properties." These attitudes are aptly captured in the minute on education by Macaulay in 1835, a member of the Supreme Council of India between 1834 and 1838, arguing for the introduction education in English when he stated, "I have no knowledge of either Sanscrit or Arabic. [...] I am quite ready to take the oriental learning at the valuation of the orientalists themselves. I have never found one among them who could deny that a single shelf of a good European

library was worth the whole native literature of India and Arabia. The intrinsic superiority of the Western literature is indeed fully admitted by those members of the committee who support the oriental plan of education."[4] Along with the denigration of the status of indigenous languages, the knowledge of the colonial language was essential to achieving socioeconomic mobility in the colonial state. Thus, in the postcolonial era, individuals who possessed linguistic capital in the colonial language were direct beneficiaries of positions in the administrative and bureaucratic, and not surprisingly have been keen to protect their rents through continuation of language policy that favors the colonial over the indigenous (Tollefson and Tsui, 2003). A combination of elite interests (Weinstein, 1983), combined with continuing stigmatization of indigenous languages (Rannut, 2010) and inertia on part of the policy leaders (Albaugh, 2014) have thus been conjectured to underlie the continued dominance of colonial languages in the postcolonial world in general, and Sub-Saharan Africa in particular.

We add to this body of work by explicitly putting to test the role of perceived costs and returns to the use of the colonial language, as well as the prevalent beliefs about importance of the colonial language for labor market success, and successful participation in the global economy. Our results suggest that the importance of group competition and fears of discrimination by installing another groups' language as strong drivers of language choice. The group competition results in the retention of an ethnically 'neutral' choice and levels the playing field between the different groups, but at the same time results in negatively affecting the educational outcomes by imposing high costs of obtaining the necessary linguistic capital.

2   The setting, the channels and survey design

We collect data from a rural and an urban site in Zambia, namely around Mpumba in the Muchinga province and Lusaka (the capital). Zambia is a landlocked country located in Southern Africa. It shares its borders with the Democratic Republic of Congo in the north, with Botswana, Mozambique, Namibia and Zimbabwe in the south, with Tanzania to the north-east, Malawi to the east, and with Angola to the west. It became independent from British rule in 1964 and has a population of around 13 million. It has a per capita income of $1721 measured in current US$ and around 64.4 percent of the population lives below the $1.90 a day poverty line measured in 2011 international prices (World Bank Group, 2012).

---

[4]http://www.columbia.edu/itc/mealac/pritchett/00generallinks/macaulay/txt_minute_education_1835.html

2.1 Why Zambia?

The choice of Zambia as the site location is based on three factors. First, Zambia, like 28 other countries in SSA, exclusively uses the colonial language in all formal domains (Laitin and Ramachandran, 2020). Second, multiple language groups are present and ethnic politics is a central aspect of political life in Zambia (Posner, 2005). Third, though there are numerous language groups in Zambia, all of the indigenous languages are Bantu languages and come from the Niger-Congo language family (Kashoki and Mann, 1978). The relative similarity implies that knowledge of the large indigenous language is widespread; as Table 1 shows even members who do not belong to the largest linguistic group of Zambia, that is non-Bemba speakers, are more likely to self-report fluency in Bemba (48 percent) than in English (36 percent).

2.2 Languages and language use in Zambia

The Bantu speaking people settled in different parts of Zambia during the Bantu expansion from the regions of Cameroon and Nigeria starting around the 12th century AD (Fagan, 1967). The colonial legacy meant that English became the official language in 1964, and is the only language so identified in the 1991 constitution. English is the dominant language of education, business, administration, and government; schooling in local languages is typically available up until the first few grades of primary schooling, and secondary and tertiary schooling is available exclusively in English (Albaugh, 2014).

How many indigenous languages are spoken in Zambia is not an easy question to answer, with estimates ranging from 20 to 80 languages, as it is notoriously difficult to distinguish between what is classified as a language opposed to a dialect (Marten and Kula, 2008). The 1991 constitution, however, designated seven indigenous languages as national languages, namely, Bemba, Nyanja, Tonga, Lozi, Kaonde, Luvale, and Lunda. These seven languages are the more important languages for wider communication in the country, and the first four account for the large majority of the first and second-language speakers. Recognition as national languages meant that these languages along with English are supposed to be used in the early years of primary schooling, though in practice this still remains restricted (Gordon, 2014). Efforts have been made by the government to create a common orthography and publish some key government documents in these seven languages. Table 2 shows the proportion of people estimated to use the seven national languages, and the official language, as their first and second language, respectively.

Table 2 shows that Bemba, Nyanja, Tonga, and English are spoken by more than 10 percent of the population as their first or second language. Bemba is the most widely spoken language with 50 percent of the population reporting that they use the language as a first or a second language; this is within a reasonable range of the 63 percent shown in Table 1,

7

where not just the first or second but the entire language repertoire is taken into account. English in turn is spoken by 1.7 percent of the population as their first language, and by 26 percent of the population as a second language. This number is 'broadly' within the range of 21 provided by Albaugh (2014) or the 37 percent shown in Table 1, where the entire language repertoire is taken into account. These are self-reported language repertoires and provide no information on the actual level of fluency that the individuals possess. Data from the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) raise the concern that the self-reported ability to speak English might be overestimated. The data from Zambia on 6th Grade students show that only 20 percent of students reach the minimum and only 5 percent the desirable reading level.

Figure 1 shows the spatial distribution of the seven national languages of the country. Each language has a specific regional base, where it is predominantly employed. Bemba is the main language of the Northern Province, Luapula, Muchinga and the Copperbelt, and, to a lesser extent, of the Central Provinces too. Nyanja is the main language of the Eastern Province, as well as of the province of Lusaka where Bemba and English are also widely used. The regional base of Tonga lies in the Southern Province, whereas Lozi is spoken mainly in the Western Province. Lunda, Luvale, and Kaonde are spoken in the North-Western Province which does not have a single dominant language.

Using the Afrobarometer to look at knowledge of Bemba by language group shows 51, 32, 18, 33, 54 and 76 percent of Nyanja, Tonga, Lozi, Luvale, Lunda and Kaonde speakers report speaking Bemba; whereas the corresponding figures for English are 49, 40, 48, 36, 38 and 47. Thus, except for the Tonga and Lozi, all major non-Bemba language group members report greater proficiency in Bemba than English.

2.3   Sites of collection

Data was collected in August 2015 in two sites in Zambia, i.e. in Lusaka and Mbumpa.[5] The two sites were chosen to have a rural and an urban representation of the Zambian population. All enumerators were in command of English as well as the prevalent local language(s).

The data representing the urban population were collected from the national capital, Lusaka. The capital is the largest city of Zambia with a population of around 2 million. The sampling frame was created by identifying nine representative neighborhoods from which data was collected.[6] Through neighbourhood selection, we oversampled more educated

---

[5]The approximate location of the sites is shown in Figure 2.

[6]The nine neighborhoods include: Chelston (medium income area with low population density), Kabwata (middle income area with medium population density), Kalikalinga and Kamanga (low income areas with high population density), Kaunda square (medium income area with high population density), Mutendere (low income areas with high population density), Northmead, Rhodes Park and Shilenje (high income areas

individuals as we wanted to obtain the preferences of elites who often have undue influence on policy in such settings, and have an important stake in preserving the status-quo due to their linguistic capital. While elites tend to be fluent in English, Nyanja is the dominant local language amongst non-elite residents in Lusaka. The urban sample consists of 109 respondents with females comprising 47 percent.

The data representing the rural population were collected from the Mpumba and surrounding villages, located in the Mpika district, in the newly created Muchinga province. Mpumba is a remote village located off the highway connecting Lusaka with Tanzania. Most of the villages are without electricity, reliant on the village fountain as a source of water, and are populated with subsistence farmers with low education and limited knowledge of English. The Mpika district according to the 2000 Zambian census had a population of 146,196 people. The individuals interviewed were randomly selected households in the villages.[7] The rural sample consists of 93 individuals with females comprising 48 percent.

2.4   The channels and the survey methodology

To test beliefs regarding perceived costs and benefits, we ask respondents about their expectations concerning outcomes of hypothetical agents under scenarios of different language policies. More specifically, to understand the role of learning costs, we ask them "Imagine 7 children who are 7 years old. Their parents speak only indigenous language a and no English. Therefore, they do not know any English when entering school. How many of these children will finish secondary schooling if it is provided in English or indigenous language a. We elicit not just a point estimate but confidence intervals by asking for the least and the most number of children who would finish secondary school. We also elicit beliefs about costs of learning in an indigenous relative to English that is not the mother tongue of the children. The survey was designed so that for half of the respondents the language spoken at home coincides with the hypothesized indigenous language to be provided at school, and for the other half, home and hypothesized indigenous language to be provided in school differ.

To further probe the perceived costs imposed by indigenous versus colonial language, we also elicit confidence intervals for the hours of study required every day for a hypothetical child to become fluent in English compared to becoming fluent in another indigenous language that is not the mother tongue of the child.

Using the same methodology, we also elicit confidence intervals on the average earnings per month in Kwachas on the expected future monthly earnings given the following three language scenarios: (1) education were provided in English and government administration

---

with medium population density).

[7]The protocol followed was to knock on every second household door.

and jobs were in English; (2) education were provided in an indigenous language and government administration and jobs were in an indigenous language; and (3) education were provided in an indigenous language and government administration and jobs were in English. To further understand the perceived labor market returns to the colonial language, we also elicit beliefs on whether knowledge of Math and science or English is a more crucial determinant of labor market earnings.

The use of hypothetical agents in the scenarios minimizes the potential impact of private information respondents have about themselves and allows us to isolate the channels we are interested in.[8] Another advantage of our methodology is that the indirect way of separately eliciting beliefs about outcomes and preferences for language policy allow us to back out the relative importance of different channels.

To understand the role of attitudes regarding the suitability of indigenous languages for use in formal domains, we collect information using a Likert scale on the stated belief on the need for a country to use English, French or Portuguese as an official language to be economically successful. Adegbija (1994) shows there are disproportionate attitudes of superiority toward European languages in Sub-Saharan Africa, while there are attitudes of low esteem and inferiority toward indigenous African languages. In a similar vein, in the context of South Africa, Prah (2006, pg. 18) notes, "It is unfortunate that most parents still believe that speaking eloquent English necessarily means you are intelligent." We operationalize the concept of low esteem by again employing a Likert scale and elicit beliefs about whether individuals are likely to believe that English is the only language in which knowledge is useful and whether English is the language of the intelligent people.

To gauge the importance of ethnic politics, we again employ a Likert scale to elicit beliefs on whether use of indigenous language a as official would result in member of group b facing discrimination in finding jobs or disadvantages the members of group b. To understand the importance of class in relation to ethnicity, we use the Likert scale to elicit beliefs on perceptions of the gap between the rich and the poor being a problem in the context of Zambia, whether they think it is a bigger concern than competition between ethnic groups, and whether the use of indigenous languages as the language of education and government would help close the gap between the rich and the poor in Zambia.

Our key outcome variable is the stated response of the individual regarding the most preferred official language/s of education and government in Zambia. Additionally, we also ask whether the government should rethink language policy and why, and whether voting is a good mechanism for aggregating language policy preferences in society. Finally, we also collect information on basic demographic characteristics, as well as self-reported fluency in the major indigenous languages and English, as well as the beliefs about the knowledge of these languages among other members of the population.

---

[8]The survey methodology builds on Attanasio (2009), Delavande et al. (2011), and Delavande (2014) concerning the elicitation of beliefs in developing countries.

2.5   Key characteristics of the sample

Table 3 shows the descriptive statistics of the urban and the rural sample. The mean age of the respondents is 36 years. The percentages of rural and urban respondents who have completed secondary schooling are 15 percent and 48 percent, respectively.

Furthermore, 10 percent of the rural sample and 27 percent of the urban sample are university graduates. The individuals in our sample are more educated than the average individual in the country. For instance, data from the Central Statistical Office (CSO), Zambia (2007) shows that 44 and 6 percent of the males aged between 18 and 36 have completed secondary schooling and higher education, respectively, in Zambia. As mentioned before, the reason for oversampling more educated individuals was to have a representation of the language preferences of elites, who might have undue influence on charting the course of language choices in society. Around 60 percent of our sample are married and 37 percent of the individuals are employed, which corresponds to the low levels of employment encountered in most of Zambia.

Table A1 in the Appendix shows the ethnic distribution and the distribution of the languages spoken at home. The village of Mpumba lies in the Bemba speaking Muchinga province, thus not surprisingly Bemba speakers comprise 82 percent of the rural sample. 67 percent of the individuals report using Bemba at home, comparable to 52 percent reported in the 2000 census.

Nyanja speakers account for 7 and 22 percent of the rural and urban sample, respectively, and 41 percent of individuals report using Nyanja at home. Tonga is used by 14 percent of individuals at home and ethnic Tonga form around 9 percent of the sample. Finally, about 5 percent of the sample are Lozi speakers, with 4 percent reporting as using Lozi at home. As for the official language English, 4 percent of the rural respondents and 47 percent of the urban respondents report using English at home, with the overall average being 27 percent. This number is again comparable to the 28 percent who report speaking English as a first or second language in the 2000 census.

The proportion who report to have good knowledge of English is 48 percent, whereas 34 and 18 percent report having fair and poor skills, respectively. This implies that around 80 percent of the population believes they have fair or good knowledge of the English language. Unfortunately, we were not able to conduct proficiency tests of their English skills to be able to obtain an objective measure that could be contrasted with the self-reported assessment. This remains an important task for the future as the proportion of people who report having good or fair English skills is more than four times the number of Grade 6 children reaching the minimum reading level. This leads us to suspect that people tend to overestimate their command and ability to function effectively in the English language.

3 Method and results

3.1 Method

Our key outcome variable is a dummy indicating whether the respondent mentioned at least one local language when asked "What language(s) do you think should be the official language(s) of education and government in Zambia?"[9] All estimations are carried out with Ordinary Least Squares, wherefore the coefficients are interpretable as percentage-point increases related to a one unit increase in the independent variable.[10] Moreover, all regressions include a constant, as well as controls for age, age squared, gender dummy, urban dummy, earnings, employment dummy, dummy for completion of secondary education, fair or good English skills dummy, an ethnic Bemba dummy, and dummies for whether the local language in the scenario was Bemba or Nyanja.

3.2 Language policy preferences and its correlates

Panel A of Table 4 summarizes the language policy preferences of the individuals in our sample. The percentage of individuals who report wanting only English as the official language of education and government in the country is 71 percent. In contrast, only 19 percent of the sample expresses a preference for using a local language exclusively as the language of education and government. 29 percent of the individuals express a preference for the use of both local language/s and English as the language of education and government.

Before turning to the factors affecting official language choice, we look at which personal characteristics are correlated with preference for (i) the use of local language(s) in government and education, (ii) the use of only local language(s). The results are shown in Table A2 in the Appendix. It is seen that income, education, and self-assessed English skills are not significantly related to language-choice preferences. This suggests that preference for the elite language is not correlated with socioeconomic status. The only two personal characteristics correlated with preference for installing an indigenous language are age (u-shaped) and whether the individual belongs to the majority language group.[11] Moreover, we find that majority language speakers have a lower rather than higher demand for the installation of using an indigenous language.

---

[9]Note that this means that respondents potentially mentioned English in addition. The results for whether the respondent mentioned exclusively local language(s) are very similar.

[10]Using logit or probit models provided qualitatively similar results. Due to the ease of interpreting coefficients, we chose to present estimations of linear probability models.

[11]Given that we only have cross-sectional data without a panel dimension, we cannot distinguish between an age or cohort effect, which could be driven, for instance, by different experiences in formative years.

3.3 Perceptions of costs and benefits and language preference

One of the crucial arguments for using the colonial language is the benefits stemming from the global linkages to education, labor and trade markets that knowledge of English, French or Portuguese facilitate. However, these need to be balanced against the potentially increased learning costs imposed by the use of a non-indigenous language. We now examine evidence on the relation between perceived economic costs and benefits and language policy preferences.

3.3.1 The role of costs

Panel B of Table 4 shows that the mean number of children expected to pass secondary schooling when using English, an indigenous language that is not one's own, and mother tongue are 4.46, 4.29 and 5.27, respectively. A rank sum test shows that the distribution of children expected to finish secondary schooling when using the mother tongue is significantly different from when using another group's indigenous language. Surprisingly, a test of equality of means shows that the average number of children who they believe will finish secondary schooling is significantly higher using English as compared to another group's mother tongue (4.46 vs 4.29). On the other hand, on average there is no perceived differences between the hours necessary to learn an indigenous language that is not their mother tongue compared to English. In both cases an estimated three hours every day is required to become fluent by the end of primary school.

The key relationship of interest is between these beliefs on costs of obtaining education via different languages and preferences concerning the official language. We next regress our key dependent variable, a dummy taking the value one when a respondent expresses a preference for using an indigenous language as official, on the two proxies of cost - number of children expected to finish secondary schooling under different languages and the hours it takes to learn a language - plus the set of controls outlined in Section 3.1.

The results in Table 5 are shown for the full sample, as well for two sub-samples, which is indicated by the scenario language being the 'Same' or 'Differ', respectively. The column titled 'Same' refers to the sample for whom the choice faced is between using either the colonial language or an indigenous language in schools, and the indigenous language is one's own mother tongue. The columns titled 'Differ' refer to the sample for whom the choice is between the colonial language and an indigenous language, which is not one's own mother tongue. This allows us to capture whether individuals' cost perceptions of obtaining education differ systematically between their own mother tongue and some other indigenous language.

Table 5 shows that the expected number of children that would finish secondary school-

ing to be not significantly related to the language preference. Thus, though individuals on average do believe that greater educational success is possible through use of mother tongue in schooling, this does not seem to translate into higher support for use of indigenous languages to act as official.

On the other hand, Columns (3) and (4) of Table 5 show that the preference for an indigenous language is increasing in the expected hours it would take to learn English and decreasing in the expected hours it takes to master the non-mother tongue indigenous language. This means that respondents believing that English is more difficult to learn and indigenous languages are easier to learn are more likely to support the idea of using indigenous languages in education and government. However, as seen in Panel B of Table 4, individuals on an average do not believe that learning English takes more hours than learning another indigenous language.

To further probe the importance of the cost channel, besides these indirect methods of eliciting the relative ease/difficulty of learning in English vs. indigenous languages, we also asked respondents directly whether they believe it to be easier to learn other subjects in an indigenous language. Surprisingly, only 28 percent of the sample believes that it is easier to learn Math in an indigenous language as compared to English. When asked more specifically about learning in the mother tongue, respondents who report that learning in English would be easier is equal to the proportion who report learning in the mother tongue would be easier.

This perception of costs imposed by different languages, captured by the three different proxies, seems to be at odds with the actual English knowledge of the people summarized in Table 1; in particular, as Table 1 shows, in Zambia 48 percent of individuals not belonging to the ethnic group Bemba report speaking it well, a language which plays no role in any official domain, whereas only 37 percent report speaking English well though it has been the official language in all spheres for more than 60 years.[12] In other words, the self reported repertoires suggest individuals underestimate the cost of learning in English and overestimate the cost of learning in another groups' language. Column (5) of Table 5 shows that individuals who believe that learning Math in an indigenous language is easier are a whole 15 percentage points more likely to report a preference for the indigenous language to act as official.

3.3.2 The role of returns

We next turn to the link between expected earnings and the choice of language of education, government administration and jobs. Panel B of Table 4 shows the expected earnings

---

[12]Also refer to Section 4 which using data on revealed preference rather than stated repertoires again suggests much higher knowledge of the other groups' language than English.

when (a) English is used as the medium of instruction (MOI) as well as the language of administration and jobs, or (b) the indigenous language is used as the MOI as well as the language of administration and jobs, and finally, when (c) an indigenous language is used in education but not used for administration and jobs. Surprisingly, we see that on an average individuals do not seem to expect that using English as the official language will result in higher earnings, as compared to a situation where an indigenous language is used both in education and other formal domains. However, it can be seen that the situation where education is provided in an indigenous language, but administration and jobs remain in English, are seen to reduce expected earnings by more than 30 percent.

We now turn to examining the correlations between the proxies for labor market returns to using different languages and preferences over official language. Columns (1) and (2) of Table 6 show that the higher the expected earnings arising from using English is negatively correlated with the preference for the use of an indigenous languages. On the other hand, for the case of provision of education in the indigenous language, higher expected earnings are positively related to the expressed preference for indigenous languages. When including only the two payoffs, that is, from the English scenario and when the indigenous and home language coincides, 20 percent of the variation in language preference is accounted for. This means that beliefs about changes in earnings alone account for more than one-fifth of the variation in preferences for the use of local languages. The expected earnings are reported in units of 1000 Kwacha (about 100 US$) per year. Therefore, an increase in expected monthly earnings of 1000 Kwacha from indigenous language provision is associated with a 8 percentage-point increase in the likelihood of having a preference for the usage of indigenous languages in education and government. In Column (2), we add the expected earnings when education is provided in the indigenous language, whereas government jobs and administration remain in English. This coefficient is significant and negative, hinting to the idea that respondents attach a lot of importance to the match in language used in education and the job market.

In Columns (3) and (4) home and language of education differ from each other in the indigenous language scenario. Again we find that the higher the expected earnings in English, the lower the support for the use of indigenous languages in education and government. In Column (5), we use the entire sample, add a dummy for the case in which home and indigenous language of instruction coincide, and interact this dummy with expected earnings. Here we find that an increase of expected earnings when learning in English of 1000 Kwacha is associated with a 13 percentage point lower preference for local language use.

Table 6 further probes the role of expected labor market returns to using English by examining whether individuals rate the knowledge of Math and science or English as a more crucial determinant of labor market earnings. To this end, we include two dummy variables dummy which takes the value 1 if individuals think Math and Science skills are more important in determining earnings and Math and Science skills are more important for obtaining good jobs, respectively, as compared to English skills. The dummies, however,

15

appear not to be significantly correlated to language policy preferences.

3.4   Attitudes towards indigenous languages

The second key class of mechanism we want to analyze are the views respondents hold regarding the suitability of using the indigenous language to act as official in formal domains. Panel C of Table 4 shows 40 percent of the sample believe that usage of the colonial language is necessary for economic success suggesting that a significant share of the population think that indigenous languages might not be suitable for usage in formal domains.

Panel C of Table 4 also shows that nearly half of the sample report that English speakers are more intelligent than speakers of indigenous languages. For those who agree that English speakers are more intelligent, we want to gain a deeper understanding of this perception and ask respondents whether they agree with any of the two statements.[13] First, whether English is the only language in which knowledge is useful, and, second, whether English is the language of the intelligent. With the first question, we want to capture whether respondents think it even useful to be intelligent if one does not know English, which many do not seem to agree with. 60 percent of the rural and urban respondents, respectively, agree with the statement that English is the only language in which knowledge is useful; in other words, around a quarter of the sample thinks that English speakers are more intelligent as English is the only language in which knowledge is useful. With the second question, we want to capture whether respondents believe one can be intelligent without being able to speak English. This does not seem to be the case for the majority of those who think that English speakers are more intelligent, as 68 percent believe that English speakers are more intelligent because English is the language of the intelligent people; in other words, around a third of the sample thinks that English speakers are more intelligent as English is the only language in which knowledge is useful. Thus, around 30 percent of the sample seem to conflate knowledge of English with advancement and intelligence. This suggests that a potential source of institutionalized negative attitudes and low esteem attached to indigenous language is that people associate knowledge as being inseparable from the medium of knowledge.

When we examine relation between language preference and attitudes, we can see that out of the individuals who do not agree that a country needs to use English, French or Portuguese as official to be successful, 30 percent of them exhibit support for use of the indigenous language as official, and this reduces to 26 percent for individuals who do not believe so. Out of the 50 percent of individuals who do not agree with the statement that English speakers are more intelligent, 29 percent express support for use of the indigenous language as official, and this reduces to 27 percent for individuals who do not believe English

---

[13]This implies that individuals can agree with neither, one of the two, or both the statements. In other words, they can list multiple responses.

speakers are more intelligent. 25 percent of individuals, who either agree or disagree with the statement, English is the language of the intelligent, want the installation of indigenous language to act as official. To summarize, a non-negligible share of individuals believe that the use of English to act as official is necessary for economic success, that knowledge of English is a mark of intelligence and that English is the only language in which knowledge is useful. However, negative attitudes seem to only weakly correlate with a stronger preference for the colonial language.

Given the prevalence of these beliefs it is interesting to understand how individuals think about nations that are economically successful but do not rely on English, French or Portuguese. To explore this we ask individuals their beliefs on the choice of medium of instruction in schools in India and Malaysia for mathematics and science, and the official language of Sweden and South Korea.

The summary stats are shown in Table 7; 58 percent of respondents believe that Sweden uses English or French as the official language and more than half assume that Malaysia and India rely exclusively on English for educational purposes.[14] We, however, do not find systematic evidence concerning how (lack of) information might affect preference over language choice, but this seems an important avenue to pursue to understand the overwhelming preference for the former colonial language.

3.5 Ethnic and class cleavages

The last class of explanations pertain to the role of linguistic diversity, and the relative importance of the class and the ethnolinguistic cleavage in society. The fear of discrimination by others might be a reason for the preference exhibited for the use of the colonial language as the language of commerce, education, and government (Laitin and Ramachandran, 2020). People might believe that choosing any one group's language would result in the other ethnic groups being marginalized or facing discrimination in society. Consistent with this reasoning, Panel D of Table 4 shows that nearly 82 percent of the sample concurs with the statement that a group whose language is not chosen will be disadvantaged in society. Moreover, 75 percent believe that this disadvantage will be manifested through discrimination on the job market. Similarly, when asked if competition between ethnic groups is a problem, a whole 85 percent agree with the statement.

Turning to the importance of class, almost 92 percent of the respondents concur with the statement that gap between the rich and poor is a problem, though only around 20 percent seem to think that official language choice could help redress this gap.

We now examine the correlations between importance of ethnic and class cleavages and

---

[14]Both India and Malaysia primarily use indigenous languages in education.

preferences over official language choice. Column (1) of Table 8 shows that the belief that a group whose language is not chosen will disadvantage the group reduces the support for the installation of an indigenous language by around 15 percentage points; in Column (2), the belief that group whose language is not chosen will face discrimination, though negatively correlated with preference for use of an indigenous language, is not statistically significant. Surprisingly, the belief that competition between ethnic groups is a problem increases support for the use of the indigenous language. The pattern here suggests that individuals believe that choice of their own language can actually promote group interests, thus increasing support for use of indigenous languages as official. On the other hand, framing it as a loss, that is, a scenario where their language is not utilized, leads them to favor a policy that implements the colonial language. Finally, Columns (5) and (6) show that associating use of indigenous languages as being pro-poor increases the support by around 28 percentage points. Thus, ethnic concerns not only are widespread, they seem to strongly correlate with preferences over official language choice.

In the presence of multiple linguistic groups, a concern about a unifying identity could be an important determinant of the preferred language for education and government administration. Table A3 in the Appendix shows that despite 88 percent expressing a feeling of belonging to their linguistic group, only 15 percent see their linguistic group as their primary identity, in contrast to 51 percent and 30 percent identifying themselves primarily as Africans and Zambians, respectively.[15] When regressing the preference for an indigenous language on expressed identity dummies while controlling for personal characteristics, we find no statistically significant relationships. The results in Table 5 suggest that social identities are not correlated to language preferences. However, Column (6) of Table A4 shows that concerns about local languages weakening the national identity are indeed significant and negatively correlated with the support for local languages.[16]

4   Discussion

The analysis of the survey data shows the overwhelming support for English can be attributed to several factors. The first of these is the perception that learning in English is not necessarily more difficult with only 28 percent of the sample reporting that they would find it easier to learn mathematics in an indigenous language as compared to using English as the medium of instruction. The cost of learning through the medium of English could potentially be underestimated as evidenced by the limited spread in the ability to speak in

---

[15]The 4th round Afrobarometer data are based on a nationally representative sample and exhibit a very similar picture, with only 13 percent of the Zambian sample identifying themselves primarily with their ethnic group.

[16]The variable "Weaken national identity" in Table A4 is a dummy that takes the value one when individuals either completely or mostly agree with the statement that "The national identity would be weakened because of the lack of an uniting common language."

Table 1. Despite English being the sole official language for more than half-a-century in Zambia, as Table 1 shows, non-Bemba speakers are 33 percent, 12 percentage points, more likely to report speaking Bemba, as compared to English (48 percent speak Bemba and 36 percent English), suggesting a greater ease in learning the majority group language. This is also supported by the poor educational outcomes of Zambian students, who rely exclusively on the use of English. Nonetheless, people at odds with their own experience, still seem to consider it to be easier to use English as a medium of instruction. One criticism that could be leveled at the data is that the language abilities are self-reported and should be treated with caution while drawing inferences about actual competence. To further explore the ability to *actually* function in languages besides the mother tongue, we employ Zambian data from the 2013-14 Demographic and Health Survey (DHS). The data reports both the native language of the respondent, as well as the language in which the survey interview was conducted. We have a sample of 31,184 individuals, out of which 53 percent, or around 16,500 individuals, answer the survey questions in a language different from their mother tongue. The languages used by the individuals, when not answering the survey questions in their mother tongue, shed light on the revealed, rather than stated, repertoires of individuals. Whereas only 14 percent of the sample uses English as the language of the interview, the rest 86 percent use other indigenous languages, with 43 and 30 percent using Bemba and Nyanja. The above seems inconsistent with the individual's own skill set or language abilities and the perceived difficulty of learning in English as compared to learning in the indigenous language. The lack of exposure to schooling systems using indigenous language might be one of the reasons for the belief distributions concerning the cost imposed by the use of different languages. Exposure may change individuals beliefs as the colonial language can also be learned without being the medium of instruction (Cummins, 2000; Laitin et al., 2019).

The DHS data also allows us to explore the link between socioeconomic status and language skills. The DHS classifies individuals, based on their wealth, into quintiles labeled as 'poorest', 'poorer', 'middle', 'richer' and 'richest'. On the one hand, for the individuals whom the survey interview is conducted in English, only 5 percent belong to the categories of 'poorest', 'poorer' and 'middle'. On the other hand, for Bemba and Nyanja the corresponding figures are 55 and 31 percent, respectively. The data thus indicates a strong gradient between English speaking ability and socioeconomic status. This is consistent with a large body of sociolinguistic literature that has argued that retention of a colonial language acts as a barrier for the majority of the population and allows economic elites to perpetuate their dominance through the language capital they wield (Laitin, 1977; Weinstein, 1983; Tollefson and Tsui, 2003). Thus, it remains an open question as to why the large majority of individuals (79 percent in our case) tend to consider the choice of English to be class neutral. In contrast, 80 percent of the sample reports that using another group's language would result in a disadvantage. This seems at odds with the actual ability of the individuals to be able to learn and function in other indigenous languages.

The above is analogous to a puzzle that Esteban and Ray (2008) try to answer in their

work, that is, the salience of ethnic conflict especially in societies with marked economic inequalities. They show that the 'rich' may prefer peace overall, yet propose an ethnic alliance in order to prevent a class conflict initiated by the 'poor' (Esteban and Ray, 2008, 2186). Similarly, in the realm of language policy, by highlighting the ethnic dimension involved in installing an indigenous language, the elites can successfully convert what is essentially a class issue into an ethnic issue. The Sri Lankan experience suggests something very similar; though the initial demand was to replace English with indigenous languages, the elites were successful in turning what was a class conflict into an ethnic conflict where the elites emerged largely unscathed and de-facto English remained the language of power and bureaucracy in society (Horowitz, 1973; Gunasekera, 1996; Narayan Swamy, 1994). The situation is further complicated by the fact that most indigenous languages in this part of the world have oral histories, and switching to using indigenous languages in formal domains is not just a question of nominating one or the other language but requires investments by the government to create standardized orthographies, dictionaries, and the necessary literature to be used in schools and institutions of higher learning before they can be installed as official. Laitin (1977, 2000) discusses in the context of Somalia and Sri Lanka, the role of the bureaucracy, whose position is guaranteed due to their linguistic capital, in imposing barriers and overstating the difficulty of transitioning to African languages, instead of performing their job of aiding the process of transition. In sum, the fear of domination by certain ethnic groups, and lack of association between elite interests the use of the colonial language result in the status-quo dominance of colonial language being perpetuated in SSA. Thus, though retention of colonial language is neutral from the perspective of ethnicity, it has the potential of favoring elites.

Finally, it is important to note that our results are based upon a small sample based in Zambia. Though, we validate several features of the sample using additional data from the Afrobarometer and the DHS, it would be important to undertake surveys that prove both greater spatial coverage of Africa and larger sample sizes to be able to draw any definitive conclusions.

5   Conclusion

Sub-Saharan Africa is characterized by the use of former colonial languages in education and the government despite these languages not being spoken by the majority of the population. In this paper we investigate whether people are in favor of this policy and what drives their preferences. In order to do so we collect data from an urban and rural site of Zambia, a country with seven major local languages and English as official language. We use hypothetical scenarios in order to gain an idea of how a policy change would be perceived to affect outcomes.

We find that in both the urban and rural setting more than three-fourths of the respon-

dents prefer English as the official language. Accounting for the importance of learning costs, we find results that are at odds with respondents actual linguistic repertoires or language use: more than two-thirds of the sample does not report that a child would find it easier to learn Math in an indigenous language as compared to in English. However, individuals who believe so are almost 20 percentage points more in favor of installing an indigenous language as official. Turning to returns, we find that individuals consider it important that indigenous language be used not only in education but also as language of administration and jobs. They report that a situation where education is provided in an indigenous language but administration and jobs remain in English will reduce expected earnings by more than 30 percent. As most language policy proposals that have been debated in the context of SSA have primarily dealt with the question of language of education but has not addressed the question of language of administration and jobs, might underlie the limited support for indigenous languages in SSA.

One of the key issues associated with local languages seem to be related to the concerns of ethnic domination. More than 80 percent of the individuals report that installing another group's language as official would result in individuals from other groups facing discrimination in the job market. Moreover, individuals who express this attitude are much more likely to support the use of the colonial languages to act as official. This is despite the fact that individuals seem to have greater proficiency in other indigenous language than the former colonial language.

The dominating preference for the former colonial language suggests that it is of first-order importance to improve its knowledge across the population to improve human capital and participation in the economic and political life. If, however, this preference stems from biases in perceived returns or over estimating costs of learning and using other indigenous languages, it could be beneficial to reconsider language policy. More understanding about the actual returns is required but is left for future research.

References

Adegbija, E. E. (1994). *Language attitudes in Sub-Saharan Africa: A sociolinguistic overview*. Number 103. Multilingual matters.

Albaugh, E. A. (2014). *State-building and multilingual education in Africa*. Cambridge University Press.

Alesina, A., W. Easterly, and J. Matuszeski (2011). Artificial states. *Journal of the European Economic Association 9*(2), 246–277.

Anderson, B. (2006). *Imagined communities: Reflections on the origin and spread of nationalism*. Verso books.

Angrist, J. D. and V. Lavy (1997). The effect of a change in language of instruction on the returns to schooling in Morocco. *Journal of Labor Economics*, S48–S76.

Asiwaju, A. I. (1985). *Partitioned Africans: Ethnic relations across Africa's international boundaries, 1884-1984*. C. Hurst & Co. Publishers.

Attanasio, O. P. (2009). Expectations and perceptions in developing countries: Their measurement and their use. *The American Economic Review 99*(2), 87–92.

Azam, M., A. Chin, and N. Prakash (2013). The returns to english-language skills in India. *Economic Development and Cultural Change 61*(2), 335–367.

Bamgbose, A. (1991). *Language and the nation: The language question in sub-Saharan Africa*. Edinburgh University Press.

Bleakley, H. and A. Chin (2004). Language skills and earnings: Evidence from childhood immigrants. *Review of Economics and Statistics 86*(2), 481–496.

Bourdieu, P. (1991). *Language and symbolic power*. Harvard University Press.

Burke, P. et al. (2004). *Languages and communities in early modern Europe*. Cambridge University Press.

Central Statistical Office (CSO), Zambia (2007). Zambia Demographic and Health Survey 2013-14. Rockville, Maryland, USA: Central Statistical Office, Ministry of Health, and ICF International.

Chiwome, E., J. Thondhlana, and R. Herbert (1992). Sociolinguistics and education: A survey concerning attitudes on the teaching of Shona through the media of Shona and English. *Language and Society in Africa*, 247–263.

Cummins, J. (2000). Bilingual children's mother tongue: Why is it important for education? sprogforum nr. 19, 2001, 15-20.

Delavande, A. (2014). Probabilistic expectations in developing countries. *Annual Review of Economics 6*(1), 1–20.

Delavande, A., X. Giné, and D. McKenzie (2011). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics 94*(2), 151–163.

Dustmann, C., S. Machin, and U. Schönberg (2010). Ethnicity and educational achievement in compulsory schooling. *The Economic Journal 120*(546), F272–F297.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics*, 1203–1250.

Egger, P. H. and A. Lassmann (2012). The language effect in international trade: A meta-analysis. *Economics Letters 116*(2), 221–224.

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. and D. Ray (2008). On the salience of ethnic conflict. *The American Economic Review*, 2185–2202.

Fagan, B. M. (1967). *A short history of Zambia: From the earliest times until AD 1900.* Oxford University Press.

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth 8*(2), 195–222.

Gordon, R. (2014). Language of education planning in Zambia. *Linguistic Portfolios 3*(1), Article 6.

Green, L. (1987). Are language rights fundamental. *Osgoode Hall LJ 25*, 639.

Gunasekera, R. O. (1996). The implementation of the official language policy, 1956–1970. *In National Language Policy in Sri Lanka*.

Horowitz, D. L. (1973). Direct, displaced, and cumulative ethnic aggression. *Comparative Politics 6*(1), 1–16.

Isphording, I. E. and S. Otten (2013). The costs of babylon — linguistic distance in applied economics. *Review of International Economics 21*(2), 354–369.

Kashoki, M. E. and M. Mann (1978). A general sketch of the Bantu languages of Zambia. *Language in Zambia*, 47–100.

Ku, H. and A. Zussman (2010). Lingua franca: The role of English in international trade. *Journal of Economic Behavior & Organization 75*(2), 250–260.

Laitin, D. D. (1977). *Politics, language, and thought: The Somali experience.* University of Chicago Press.

Laitin, D. D. (1994). The tower of babel as a coordination game: Political linguistics in Ghana. *American Political Science Review 88*(03), 622–634.

Laitin, D. D. (2000). Language conflict and violence: The straw that strengthens the camel's back. *European Journal of Sociology 41*(01), 97–137.

Laitin, D. D. and R. Ramachandran (2016). Language policy and human development. *The American Political Science Review 110*(3), 457.

Laitin, D. D. and R. Ramachandran (2020). Linguistic diversity, official language choice and human capital. Working paper.

Laitin, D. D., R. Ramachandran, and S. L. Walter (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), 239–72.

Marten, L. and N. C. Kula (2008). Zambia: One Zambia, one nation, many languages. *Language and national identity in Africa*, 291–313.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. *The American Economic Review 102*(4), 1508–1539.

Mparutsa, C., J. Thondhlana, and N. Crawhall (1992). An initial investigation into language attitudes of secondary school students in Zimbabwe. *Language and to Society in Africa: The Theory and Practice of Sociolinguistics, ed. by Robert K. Herbert*, 235–245.

Mufwene, S. and C. B. Vigouroux (2008). Colonization, globalization and language vitality in africa: An introduction. *Globalization and language vitality: Perspectives from Africa 1*, 31.

Mukhwana, A. (2014). Attitude towards Kiswahili in urban Kenya. *International Journal of Scientific Research and Innovative Technology 1*(3).

Narayan Swamy, M. (1994). Tigers of Lanka: From boys to guerrillas. *New Delhi, India: Konark*.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford, England: Oxford University Press.

Piper, B., S. S. Zuilkowski, and S. Ong'ele (2016). Implementing mother tongue instruction in the real world: Results from a medium-scale randomized controlled trial in kenya. *Comparative Education Review 60*(4), 776–807.

Pollock, S. et al. (2006). *The language of the gods in the world of men: Sanskrit, culture, and power in premodern India*. Univ of California Press.

Pool, J. (1987). Thinking about linguistic discrimination. *Language Problems and Language Planning 11*(1), 3–21.

Pool, J. (1991). The official language problem. *The American Political Science Review*, 495–514.

Posner, D. N. (2005). *Institutions and ethnic politics in Africa*. Cambridge University Press.

Prah, K. K. (2006). Challenges to the promotion of indigenous languages in South Africa. *Review Commissioned by the Foundation for Human Rights in South Africa. October – November 2006*.

Ramachandran, R. (2017). Language use in education and human capital formation: Evidence from the Ethiopian educational reform. *World Development* (98), 195–213.

Rannut, M. (2010). *Linguistic human rights: Overcoming linguistic discrimination*, Volume 67. Walter de Gruyter.

Skattum, I. (2008). French or national languages as means of instruction. *Language and Power: The Implications of Language for Peace and Development 172*.

Taylor, S. and M. von Fintel (2016). Estimating the impact of language of instruction in South African primary schools: A fixed effects approach. *Economics of Education Review 50*, 75–89.

Thorburn, T. (1971). In R. Joan and J. Bjorn H. (Eds.), *Can language be planned?*, pp. 239. Honolulu: University of Hawaii Press.

Tollefson, J. W. and A. B. Tsui (2003). *Medium of instruction policies: Which agenda? Whose agenda?* Routledge.

Vaillancourt, F. (1983). The economics of language and language planning. *Language Problems and Language Planning 7*(2), 162–178.

Weinstein, B. (1983). *The civic tongue: Political consequences of language choices*. Longman New York.

World Bank Group (2012). *World Development Indicators 2012*. World Bank Publications.

Table 1: Language spread in Sub-Saharan Africa

Country	Average number of languages spoken	Size of the largest group	Proportion speaking the largest indigenous language		Proportion speaking the former colonial language		
			Countrywide	Minorities	Countrywide	Majority	Minorities
Benin	3.15	0.33	0.57	0.37	0.46	0.5	0.44
Botswana	2.83	0.76	0.99	0.96	0.42	0.43	0.36
Burkina Faso	2.96	0.51	0.67	0.31	0.32	0.28	0.35
Cape Verde	2.67	1	1	1	0.47	0.47	1
Ghana	3.36	0.48	0.56	0.15	0.48	0.48	0.48
Kenya	3.72	0.19	0.22	0.04	0.61	0.55	0.63
Lesotho	2.53	0.98	1	1	0.26	0.26	0.65
Liberia	2.8	0.2	0.28	0.11	0.74	0.59	0.78
Madagascar	2.63	1	1	1	0.24	0.24	1
Malawi	2.91	0.43	0.92	0.86	0.24	0.21	0.26
Mali	3.59	0.5	0.85	0.71	0.22	0.23	0.22
Mozambique	3.23	0.26	0.33	0.09	0.79	0.72	0.81
Namibia	3.84	0.51	0.56	0.1	0.76	0.73	0.78
Nigeria	3.59	0.23	0.41	0.24	0.68	0.32	0.78
Senegal	3.28	0.58	0.92	0.81	0.29	0.27	0.32
South Africa	3.68	0.17	0.29	0.15	0.7	0.67	0.7
Uganda	3.18	0.22	0.48	0.33	0.5	0.45	0.51
Zambia	3.53	0.29	0.63	0.48	0.37	0.38	0.36
Zimbabwe	2.82	0.81	0.9	0.48	0.45	0.45	0.45
Average	3.17	0.5	0.66	0.48	0.47	0.43	0.57

Notes: The estimates are based on the author calculations of the reported language repertoires in the 4th round of the Afrobarometer conducted in the year 2008. An individual is coded as being able to speak the language when they self-report that they are able to "speak well" the language under consideration. The largest indigenous language is the language of the largest linguistic group as coded by Fearon (2003). 'Majority' refers to a member of the largest linguistic group, and minorities refers to individuals who are not part of the largest linguistic group in the country.

Table 2: Language repertoires in Zambia

	Percentage of population using it as	
Language	First language	Second language
Bemba	30.1	20.2
Nyanja	10.7	19.5
Tonga	10.6	4.4
Lozi	5.7	5.2
Lunda	2.2	1.3
Kaonde	2.0	1.8
Luvale	1.7	1.9
English	1.7	26.3

Source: Zambia - 2000 Census of Population and Housing

Table 3: Key characteristics of the sample

Country	Rural	Urban	Total	SD
Age	40.78	31.47	35.82	[12.17]
Female	.48	.47	.48	[.50]
Completed primary	.77	.94	.87	[.34]
Completed secondary	.15	.48	.33	[.47]
University graduate	.10	.27	.19	[.39]
Fair/good English	.72	.91	.82	[.38]
Employed	.30	.43	.37	[.48]
Married	.71	.50	.60	[.49]
Number of children	4.63	1.83	3.15	[2.75]
Income	943.33	1985.57	1483.96	[1429.72]
Observations	93	109	202	

Source: Authors' calculations.

Table 4: Language policy preferences of sample

	Rural sample	Urban sample	Total
Panel A	**Language preference**		
Want local language(s) as official language	.190	.361	.286
Want only local language(s) as official language	.143	.231	.193
Want English as official language	.774	.761	.767
Want only English as official language	.810	.639	.714
Panel B	**Cost and return perceptions**		
Mean no. of children finishing second. school:			
(a) MOI - English	4.03	4.82	4.46
(b) MOI - Indigenous (both MT and not MT)	4.41	5.17	4.82
(c) MOI - Indigenous but not MT	3.85	4.62	4.30
(d) MOI - Indigenous and MT	4.82	5.75	5.27
Everyday study hours to become fluent in:			
(a) English	2.89	3.26	3.09
(b) Indigenous lang. not MT	2.97	3.19	3.10
Expected earnings in 1000s of Kwachas:			
(a) English in education and jobs	2.27	2.63	2.46
(b) Indigenous in education and jobs	2.26	2.70	2.50
(c) Indigenous only in education not jobs	1.57	1.48	1.52
Panel C	**Attitudes towards indigenous**		
	and colonial language		
A country needs to use Eng., Fr. or Port. to be economically successful	.495	.320	.403
Beliefs about English speakers			
English speakers are more intelligent	.674	.324	.485
Why are they more intelligent?			
Language of intelligent	.746	.531	.680
English is only language in which knowledge is useful	.600	.594	.598
Panel D	**Ethnolinguistic and class cleavages**		
What would happen to groups whose language is not chosen			
(a) Disadvantaged	.800	.832	.817
(b) Discrimination on job market	.774	.736	.754
Competition between ethnic group is a problem	.843	.854	.849
Gap between rich and poor is a problem	.924	.908	.915
Use of local language would reduce gap between rich and poor	.211	.222	.217

Source: Authors' calculations. MT refers to mother tongue and MOI to the medium of instruction.

Table 5: Ease of learning in different languages and language policy preferences

Dependent variable: Preference for a local language					
Scenario language:	Same		Differ		Full
	(1)	(2)	(3)	(4)	(5)
Children finishing secondary (English)	-0.032	0.038		0.044	-0.029
	(0.029)	(0.040)		(0.040)	(0.025)
Children finishing secondary (Local)	-0.016	0.004		-0.011	0.028
	(0.033)	(0.037)		(0.036)	(0.031)
Easier to learn in local language	0.079	0.199	0.105	0.134	0.155*
	(0.097)	(0.164)	(0.169)	(0.168)	(0.089)
Hours it takes to learn English			0.101*	0.115**	
			(0.051)	(0.051)	
Hours it takes to learn local language			-0.102**	-0.117**	
			(0.048)	(0.048)	
Same lang. scenario x Children finishing secondary (Local)					-0.037
					(0.044)
Same language scenario					-0.036
					(0.221)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	81	79	79	78	160
R^2	0.407	0.331	0.353	0.384	0.193

Notes: *, ** and *** significant at 10, 5, and 1 % significance level, respectively. Standard errors are in parentheses. All regressions include a constant, age, age squared, gender dummy, urban dummy, earnings, employment dummy, dummy for completion of secondary education, fair or good English skills dummy, an ethnic Bemba dummy, and dummies for whether the local language in the scenario was Bemba or Nyanja. In Column (5), we interact the dummy whether the languages coincide in the local language scenarios with the number of children expected to complete secondary school under the local language scenario. The column titled 'Same' refers to the sample for whom the choice faced is between using either the colonial language or an indigenous language in schools, and the indigenous language is one's own mother tongue. The columns titled 'Differ' refer to the sample for whom the choice is between the colonial language and an indigenous language, which is not one's own mother tongue. The column titled 'Full' covers the full sample.

Table 6: Expected earnings and language policy preferences

Dependent variable: Preference for a local language					
Scenario language:	Same		Differ		Full
	(1)	(2)	(3)	(4)	(5)
Perceived earnings					
English (educ. and govern.)	-0.092***	-0.084**	-0.085	-0.097*	-0.127***
	(0.033)	(0.033)	(0.055)	(0.056)	(0.030)
Local language (educ. and govern.)	0.080**	0.092**	0.042	0.043	0.051
	(0.037)	(0.037)	(0.053)	(0.053)	(0.044)
Local lang. (only educ.)		-0.077*		0.057	
		(0.040)		(0.062)	
Other factors					
Math more important for jobs	0.097	0.101	0.077	0.085	0.015
	(0.086)	(0.084)	(0.122)	(0.122)	(0.074)
Maths skills pay more	0.059	0.025	-0.170	-0.150	-0.086
	(0.076)	(0.078)	(0.115)	(0.117)	(0.067)
Same language scenario					-0.269
					(0.176)
Same lang. × Perceived earn. local lang.					0.063
					(0.060)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	79	78	80	80	159
R^2	0.562	0.591	0.326	0.335	0.304

Notes: *, ** and *** significant at 10, 5, and 1 % significance level, respectively. Standard errors are in parentheses. All regressions include a constant, age, age squared, gender dummy, urban dummy, earnings, employment dummy, dummy for completion of secondary education, fair or good English skills dummy, an ethnic Bemba dummy, and dummies for whether the local language in the scenario was Bemba or Nyanja. In Column (5), we interact the dummy whether the languages coincide in the local language scenarios with the number of children expected to complete secondary school under the local language scenario. The column titled 'Same' refers to the sample for whom the choice faced is between using either the colonial language or an indigenous language in schools, and the indigenous language is one's own mother tongue. The columns titled 'Differ' refer to the sample for whom the choice is between the colonial language and an indigenous language, which is not one's own mother tongue. The column titled 'Full' covers the full sample.

Table 7: Respondents' knowledge of language use in other countries

	Rural	Urban	Total
Only using English in education in:			
Malaysia	.424	.776	.613
India	.290	.785	.555
French or English are official language in:			
South Korea	.387	.248	.312
Sweden	.688	.495	.584

Notes: Source: Authors' calculations.

Table 8: Ethnolinguistic and class cleavages and language policy preferences

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable: Preference for a local language							
Disadvantage for groups whose language not chosen	-0.154* (0.087)			-0.150* (0.090)			
Groups whose language not chosen would face discrimination		-0.092 (0.079)		-0.043 (0.083)			
Competition between ethnic groups is a problem			0.264*** (0.095)	0.272*** (0.095)			
Local language would reduce gap between rich and poor					0.278*** (0.081)	0.283*** (0.081)	
Gap between rich and poor is a problem						-0.111 (0.118)	-0.089 (0.122)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	171	173	166	162	173	173	174
R^2	0.163	0.166	0.197	0.217	0.199	0.203	0.143

Notes: *, ** and *** significant at 10, 5, and 1 % significance level, respectively. Standard errors are in parentheses. All regressions include a constant, age, age squared, gender dummy, urban dummy, earnings, employment dummy, dummy for completion of secondary education, fair or good English skills dummy, an ethnic Bemba dummy, and dummies for whether the local language in the scenario was Bemba or Nyanja.

Northern Province
Bemba Speakers

Muchinga Province
Bemba Speakers

Luapulal Province
Bemba
Speakers

North Western Province
Lunda, Luvale
and Kaonde Speakers

Copperbelt
Bemba
Speakers

Eastern Province
Nyanja Speakers

Central Province
Bemba Speakers

Western Province
Lozi Speakers

Lusaka
Nyanja Speakers

Southern Province
Tonga Speakers
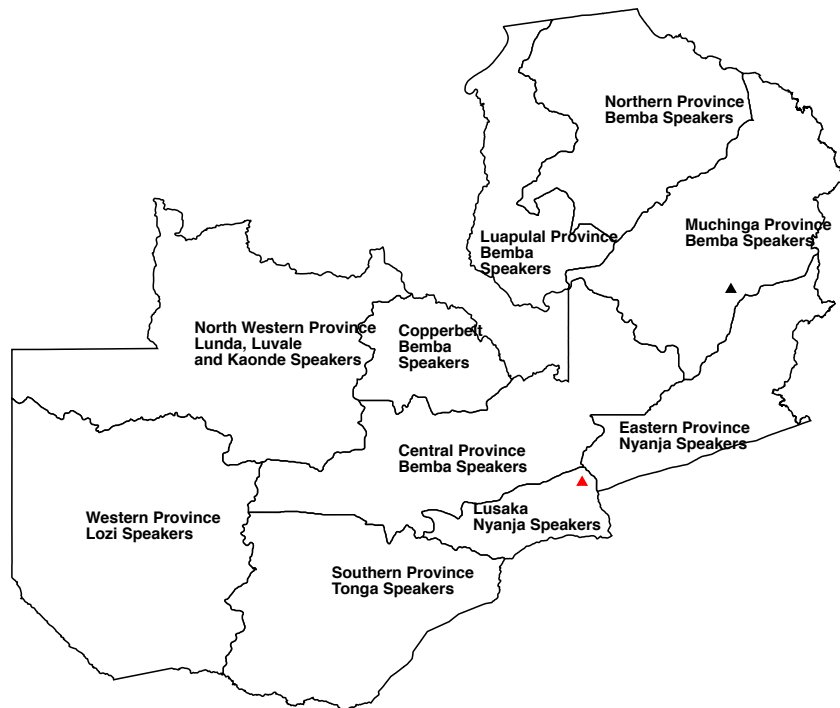
The black and red triangles indicate the approximate geographical
location of the rural and urban site of data collection, respectively.

Notes: Source: Authors' construction

Figure 1: The geographical distribution of the main languages of Zambia.

Online Appendix

Please turn to the next page.

Table A1: Ethnic and linguistic distribution of sample

	Rural	Urban	Total
Ethnicity			
Bemba	.826	.275	.527
Nyanja	.065	.22	.149
Tonga	.011	.156	.09
Bisa	.141	0	.065
Silozi	0	.092	.05
Luvale	.011	.046	.03
Kikaonde	0	.055	.03
Namwanga	.022	.009	.015
Tumbuka	.011	.009	.01
Lungu	.011	0	.005
English	.011	0	.005
Lala	.011	0	.005
Other	.022	.138	.084
Language spoken at home			
Bemba	.913	.459	.667
Nyanja	.043	.734	.418
English	.043	.468	.274
Tonga	.043	.211	.134
Silozi	0	.073	.04
Bisa	.076	0	.035
Kikaonde	.011	.037	.025
Tumbuka	.011	.009	.01
Lunda	0	.009	.005
Luvale	0	.009	.005
Other	0	.009	.005
Namwanga	.011	0	.005
Lala	0	0	0
Lungu	0	0	0
Observations	93	109	202

Notes: Ethnicities and language spoken at home can sum to more than 1 as respondents can provide multiple replies. Source: Authors' calculations.

Table A2: Relating personal characteristics to language policy preferences

Dependent variables related to language policy (specified in column header)						
	(A local)	(Only local)	(Vote)	(Vote)	(Rethink)	(Rethink)
Age	-0.044***	-0.044***	-0.020	-0.013	-0.012	0.002
	(0.016)	(0.014)	(0.017)	(0.017)	(0.017)	(0.017)
Age^2 / 1000	0.496***	0.496***	0.198	0.112	0.131	-0.024
	(0.176)	(0.158)	(0.186)	(0.191)	(0.191)	(0.193)
Female	0.071	0.049	0.066	0.034	-0.004	-0.025
	(0.067)	(0.060)	(0.071)	(0.071)	(0.072)	(0.071)
Urban	0.094	-0.026	-0.327***	-0.402***	-0.226**	-0.238**
	(0.092)	(0.082)	(0.092)	(0.096)	(0.095)	(0.097)
Earnings	-0.004	-0.026	-0.015	-0.007	0.029	0.027
	(0.034)	(0.031)	(0.037)	(0.037)	(0.037)	(0.037)
Employed	-0.009	0.060	-0.073	-0.039	-0.152*	-0.132
	(0.082)	(0.074)	(0.088)	(0.087)	(0.090)	(0.087)
Completed secondary	-0.051	0.007	0.052	0.061	0.083	0.092
	(0.090)	(0.081)	(0.098)	(0.096)	(0.101)	(0.097)
Fair/good English	-0.023	0.019	-0.132	-0.145	-0.021	-0.039
	(0.090)	(0.081)	(0.097)	(0.095)	(0.101)	(0.098)
Ethnic Bemba	-0.145*	-0.177**	0.055	0.044	0.057	0.086
	(0.083)	(0.074)	(0.086)	(0.088)	(0.089)	(0.088)
Favors local language(s)				0.259***		0.301***
				(0.083)		(0.085)
Constant	1.188***	1.158***	1.085***	0.924**	0.659*	0.297
	(0.355)	(0.319)	(0.377)	(0.390)	(0.393)	(0.399)
Observations	175	175	177	169	169	162
R^2	0.118	0.117	0.185	0.237	0.086	0.151

Notes: *, ** and *** significant at 10, 5, and 1 % significance level, respectively. Standard errors are in parentheses. 'A local' is a dummy taking the value 1 if the respondent has a preference for local language(s) (but potentially for English as well), 'Only local' takes the value 1 if the respondent has a preference for local language(s) only, 'Vote' takes the value 1 if the respondent agrees or strongly agrees that the government should allow the people to vote on language policy, and 'Rethink' takes the value 1 if the respondent agrees or strongly agrees that the government should rethink language policy.

Table A3: Identity choices of sample

	Rural	Urban	Total
Feel Zambian	.989	1	.995
Feel African	.925	.963	.946
Feel belong to linguistic group	.86	.898	.881
Feel primarily African	.556	.481	.515
Feel primarily Zambian	.222	.37	.303
Feel primarily belong to linguistic group	.156	.148	.152

Notes: Source: Authors' calculations.

Table A4: Identity and language policy preferences

Dependent variable: Preference for a local language						
	(1)	(2)	(3)	(4)	(5)	(6)
African	-0.079			-0.003	-0.017	
	(0.158)			(0.183)	(0.187)	
Zambian		0.219		0.321	0.371	
		(0.440)		(0.483)	(0.494)	
Linguistic group			-0.106	-0.094	-0.088	
			(0.105)	(0.112)	(0.113)	
Primarily African					-0.032	
					(0.077)	
Primarily linguistic group					-0.024	
					(0.104)	
Weaken national identity						-0.132*
						(0.079)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	175	174	174	173	170	161
R^2	0.120	0.116	0.126	0.123	0.125	0.148

Notes: *, ** and *** significant at 10, 5, and 1 % significance level, respectively. Standard errors are in parentheses. All regressions include a constant, age, age squared, gender dummy, urban dummy, earnings, employment dummy, dummy for completion of secondary education, fair or good English skills dummy, whether the home and language of instruction coincided in the scenarios, and an ethnic Bemba dummy. The variables 'African', 'Zambian' and 'Linguistic group' are dummies that take the value 1 when individuals completely or mostly agree with the statement "I see myself as" each of the three identities. Thus, it is possible to identity with multiple social identities. The variables 'Primarily African' and 'Primarily linguistic group' are dummies that take the value one when individuals out of the categories of Zambian, African and Member of my linguistic group choose one of the two. Thus, here individuals have to choose one among the three primary identities presented. The variable 'Weaken national identity' is a dummy that takes the value one when individuals either completely or mostly agree with the statement that "The national identity would be weakened because of the lack of an uniting common language."