Unveiling the global determinants and effects of bilingual education policies in Africa

Mathilde Col*

July 2, 2024

Abstract

Most African students were taught in the colonial language at the independence. The vast majority of African countries have addressed the issue by changing the Language of Instruction (LoI) in primary education through official laws, replacing the colonial language with a local one. However, little is known about these linguistic policies' context and long-term effects. This paper proposes a data-driven cross-country comparison of all linguistic policies targeting LoI to highlight some stylized facts about the political and linguistic context surrounding the reforms, using a novel dataset on linguistic reforms in Africa. Past and present political features seem to matter when explaining the extensive margin. Using micro-data on eighteen African countries, I provide descriptive aggregated evidence on the effect of these bilingual education reforms. Results suggest that schooling and learning increase with the introduction of local languages in education, especially for women and students who have finished primary education.

*PhD student. University of Bordeaux and Paris School of Economics. https://mathildecol.github.io/

1 Introduction

As advocated by UNESCO and research in linguistics over the past decades, using a familiar language is critical for efficient learning (Benson, 2002; Trudell, 2023); learning in a known language is associated with better mathematics and language skills test scores. This policy recommendation is particularly relevant in post-colonial Africa, where learning levels are deficient (Bold et al., 2017), and 80% of the students are taught in the ex-colonial language they do not speak or understand (World Bank, 2018).

However, these statistics should not hinder the reforming wave since the independence years that tried to address the linguistic issue by replacing the colonial language or adding local languages as medium of instruction (MoI), i.e., languages teachers use to teach the curriculum in schools[1]. Countries experienced different paths regarding the choice of medium of instruction, and little is known about the underlying motivations of policymakers. David Laitin and Rajesh Ramachandran (Laitin and Ramachandran, 2022) pinpoint the possession of a writing tradition as a determinant factor in promoting one language as official. However, choosing a national language as official does not automatically translate into using it as a medium of instruction; in the Central African Republic, Sango has been co-official since 1995, while French is still used in primary schools nowadays.

Only a few papers looked at the context in which bilingual education programs were implemented. Some highlighted implementation issues (Piper, Zuilkowski, and Ong'ele, 2016 for Kenya) or lack of political motivation (Kébé and Diop, 2021 for Senegal). Among political motives, the high ethnolinguistic diversity that defines many African states is widely seen as an obstacle to using local languages in education, as argued by David Laitin and Rajesh Ramachandran (Laitin and Ramachandran, 2022). Indeed, promoting one language rather than another would promote one ethnic group associated with this language and is a risky policy move; in the last round of the Afrobarometer survey (2021), almost half of the respondents (44%) said they do not trust people from other ethnic groups. However, rigorous data analysis has yet to be done to prove this correlation.

To my knowledge, only one large study tried to identify the plurality of motives behind reforming the use of the ex-colonial language in education in Sub-Saharan Africa in a

[1] I will refer also to Languages of Instruction (LoI) as a synonym.

book resulting from almost 20 years of intensive work on the topic (Albaugh, 2014). Erika Albaugh proposes an original measure called the "Intensity of Local Language Use in Education" (ILLED) for all African states at three points in time: 1960 (more or less at the independence), 2000, and 2010. I build on this work by providing a more granular overview of the use of local languages in education by coding every linguistic reform in all post-colonial African countries that affected the MoI and official languages. Except for Tanzania, the ex-colonial language is used in all secondary and tertiary education systems, reducing the scope of my work to primary education only. I find that the ex-colonizer's identity (French vs Britain) matters a lot, even when controlling for pre-existing differences and other potential confounders. Only some measures of ethnolinguistic diversity are good predictors of whether a country replaced the colonizer language with a local one in education. I also add to Albaugh's study by providing evidence of additional motives, such as the importance of communism as a determinant for promoting local language use in education.

As for the context surrounding linguistic policies, the literature on the effect of providing education in a known language in Sub-Saharan Africa is scarce in economics. Two papers (Laitin, Ramachandran, and Walter, 2019; Piper, Zuilkowski, and Ong'ele, 2016) provided evidence of large and positive significant impacts on test scores in controlled experiments in Cameroon and Kenya. Very recently, Mohohlwane and co-authors (Mohohlwane et al., 2023) showed similar effects and positive transfers to the second language in South Africa using two randomized controlled trials. While these findings are enlightening from an academic viewpoint, they do not tell much about the effect of providing education in the local language rather than the ex-colonial language when scaled up at the country level. In their paper, Benjamin Piper and his co-authors highlighted the implementation obstacles they encountered that are inherent to a countrywide policy, for instance, low demand from parents, scarcity of teachers proficient in the MoI. This paper provides new evidence on the long-term effects of introducing local languages as MoI using micro data on eighteen post-colonial African countries. Rather than providing causal estimates of bilingual education, I intend to give an overview of the effects of language policies that have been implemented on a large scale. I find that the effect on schooling and learning is overall positive. These results are especially large for women and increase

with the number of schooling years. It supports the official recommendation that a child should learn in a familiar language for at least five or six years (Ball, 2011; Trudell, 2023).

The paper proceeds as follows. Section 2 details the novel database listing all linguistic reforms affecting MoI and official languages in all post-colonial African states. In Section 3, I use these data and cross them with many different determinants that can affect the probability of replacing the ex-colonial language with a local one in primary education. I highlight potential underlying motives behind the choice of the MoI in post-colonial Africa. In Section 4, I provide evidence on the long-term effects of at-scale bilingual education policies using micro data, and link these findings to the implementation side.

2 Data

2.1 Exhaustive list of linguistic reforms in post-colonial Africa

Data description. For this study, I built a newly comprehensive dataset listing all the linguistic reforms in Africa that affected the languages used in primary education or administration[2]. For every country, I hand-coded every linguistic laws since the independence, indicating for every reform the year the official law passed and the language(s) newly introduced. I specified whether the reform was about languages used in education or official/national languages. The dataset contains additional details only for education reforms: the grade in which teachers theoretically switch from the newly introduced local language to the colonial one, and the second language(s) taught at school.

The time range of the dataset covers the independence years to 2023 for 54 countries. Figure 1 shows full geographical and time coverage. The last coded reform is the Senegalese law imposing the use of the local languages for the first two years of primary education, supposedly implemented at the country level in 2027. The main caveat of this dataset is the lack of information about the implementation of the reform. This dataset is a compilation of *de jure* laws only, and does not take a look at the *de facto* aspect.

Sources. I used various sources to build this new dataset, from newspapers to official policy briefs. As much as possible, I relied on high-quality policy reports, such as

[2] Available upon request.

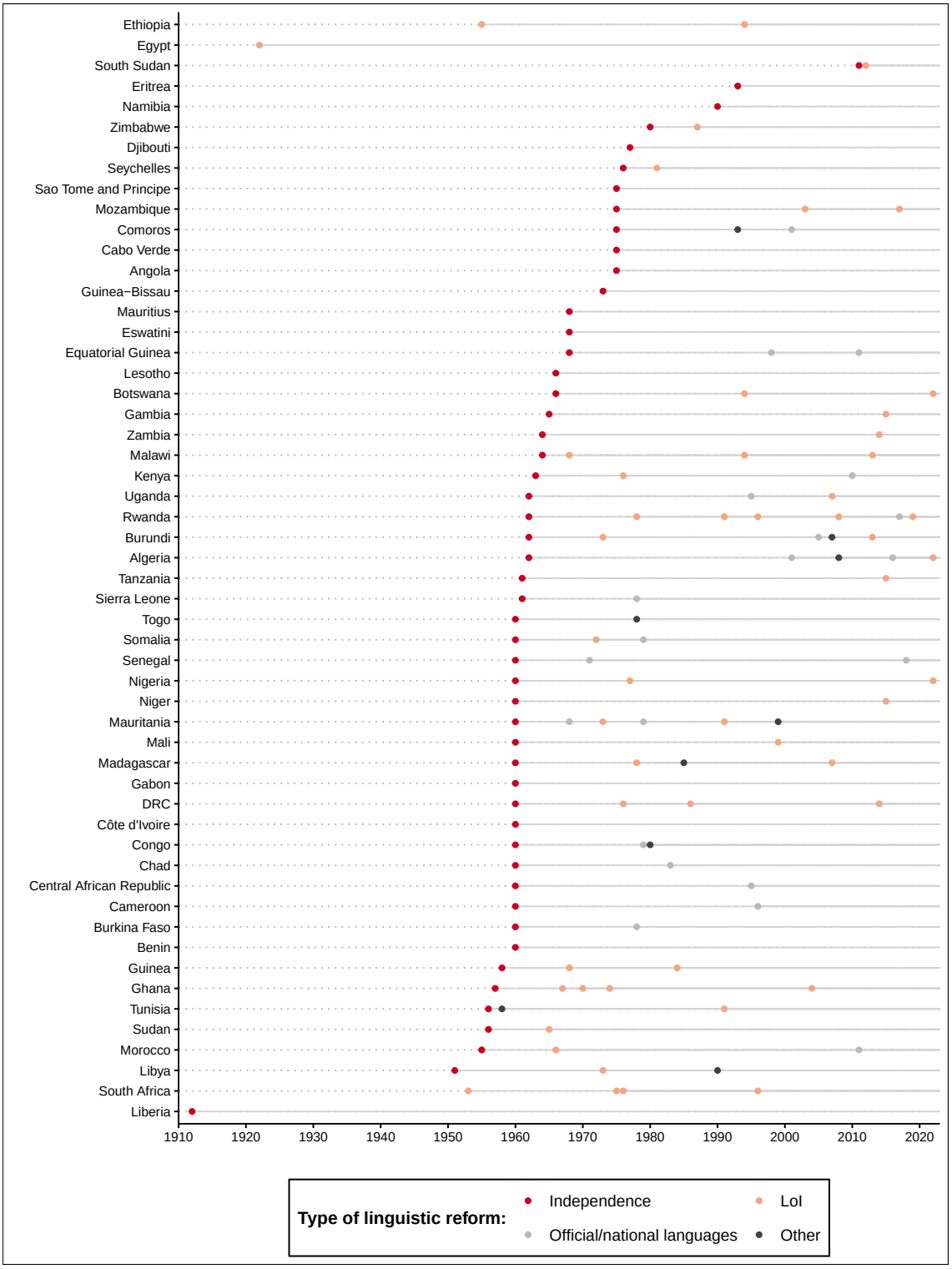

Figure 1: Country and time coverage of the linguistic reform dataset

Notes: All data hand-coded compiled by the author. The 2027 bilingual education law in Senegal does not appear because the x-axis ends in 2023. For countries not colonized during the nineteenth, the reference period starts in 1900.

UNICEF reports on language policy in Eastern and Southern African countries (Eastern, Southern Africa Regional Office (ESARO), and Section, 2016). The country profiles on LoI published by USAID on almost all English-speaking African countries were also extensively used. For the other countries, the main information came from official reports from the International Office of the Francophonie.

Data quality. Ericka Albaugh published in 2014 a book (Albaugh, 2014) in which she compiled data on the extent of the use of local language in instruction in every African country. She coded it only at three points: at the independence, in 1980, and 2010. I extended the time range of her work, with which I crosschecked the consistency of the dataset I built. I also crosschecked the dataset with information given by Jacques Leclerc (UC Laval), who graciously built a highly detailed and sourced linguistic profile for every country in the world[3].

Descriptive statistics. Eighty-four linguistic reforms are listed for the 54 African countries post-colonization period. I also hand-coded the situation at independence for every African countries that went under colonization. Out of these 84 reforms, more than 60% (53) focus only on the language of instruction. These LoI policy changes are not uniformly distributed: almost half of the countries never experienced any change in the language of instruction since their independence. Most countries kept the colonial language as the education medium, except Eritrea, Eswatini, Lesotho, Mauritius, and Namibia. All of these countries (except Mauritius) experienced a late decolonization process, and the replacement of the colonial language with a local one in education was agreed before the independence.

In 2023, eighteen African countries still used the colonial language as the primary medium for school entry. For countries that switched toward integrating local languages in primary education, there is no shared rule on when the transition towards the foreign language should happen. Figure A.1 shows that no policy consensus appears clearly. On this matter, UNESCO linguistic experts recommend a late-exit transition model if the objective is to enhance student test scores (Trudell, 2023).

[3]Available at https://www.axl.cefan.ulaval.ca/

2.2 List of pilot programs

I built a companion dataset listing the pilot programs that theoretically prepare the countrywide policy replacing the colonial language with a local one in all African countries. Indeed, changing the medium used by teachers required preparation. In Senegal, for instance, the bilingual education policy has been tested and scaled up very gradually since 2022. In addition to the variables collected in the primary dataset, the dataset contains details about the number of schools concerned by the pilot program, the start and end year of the program, and whether an evaluation officially took place. Contrary to the main dataset, the listing of pilot programs was not intended to be comprehensive.

3 Underlying determinants for policy change

3.1 Colonial legacy: France vs Great Britain

Existing work in comparative development often praised the British colonial rule over the French one (Bertocchi and Canova, 2002; La Porta, 1999). Some empirical studies even argued over the "benign" to positive role of the British administration in explaining long-term development trends. Julia Bolt and Dirk Bezemer (Bolt and Bezemer, 2009) provided empirical evidence that the higher growth rates in British colonies after decolonization can be explained by colonial education.

Naive comparison. I test whether the identity of the ex-colonizer (France vs United Kingdom) can predict whether a country is more willing to experience a change in the languages used in primary public schools. I expect the British colonies to exhibit a higher likelihood of such a change. Indeed, during the twentieth century, the French education system was more under regulation by the colonial administration than the British one (Cogneau and Moradi, 2014). French was the only language of instruction allowed in public schools, and this rule was enforced. One example of punishment used is the so-called symbol, inherited from the intensive linguistic unification campaign in France during the nineteenth century: the student who dared to speak an African language at school, even during the break, has to wear a disgusting item such as a necklace made of rabbit bones (De Gaston, 2011). In the British colonial empire, the extent of regulation that

the administration could implement was limited by the importance of missionary schools compared to state-ruled ones. The administration was also less reluctant to the use of vernacular languages if it could lead quicker to a semi-skilled African labor force that could be integrated into the colonial administration[4]. For example, Swahili, which is nowadays the most widely spoken language in Africa, was partly used and spread by the colonial administration for communication purposes (Chanson, 2012). Using a novel dataset on reforms previously detailed, I regress naively ex colonial empire belonging on the contemporary linguistic policy change. Empirically, we do observe in Table 1 that British ex-colonies are more likely to introduce local languages in primary education compared to French ex-colonies.

Controlling for pre-existing differences. However, this naive comparison is biased. Indeed, Ewout Frankema (Frankema, 2012) showed that British colonies are associated with higher educational outcomes because the controlled areas were less prone to Muslim presence and more fertile. The small Muslim presence led to a more extensive spread of Christian missions in the British Empire compared to French territories. Previous results are robust to adding these controls, i.e., the share of Muslims in the population in 2000 and geographical variables (Column 2 of Table 1). Interestingly, the presence of Muslims seems to be an essential factor: the more the share of Muslims in the country, the less reluctant the government is to replace the colonial language with a local one for education provision[5]. The rationale is that Muslims resisted greatly the Christian conversion practices during the colonial period (Daun, 2000); and one major component of the Christian presence in the African colonies was the provision of education.

Additional controls are related to ethnolinguistic diversity: the number of languages and ethnic groups could affect the government's willingness to change the MoI. Indeed, with many different languages and ethnic groups, the optimal decision could be to keep the

[4] "The first task of education is to raise the standard alike of character and efficiency of the bulk of people, but provisions must also be made for the training of those who are required to fill posts in the administrative and technical services, as well as those who as chiefs will occupy positions of exceptional trust and responsibility" Advisory Committee on Native Education in the British Tropical Dependencies, Education British Tropical Africa (London, 1925), pp. 4-5

[5] The full regression table can be found in Appendix Table A.1.

colonial language as the language used at school for two main reasons (Laitin, 1994): (i) Keeping the linguistic *status quo* in primary schools implies not choosing between different languages associated with specific ethnic groups. Hence, not choosing a language is similar to not promoting one ethnic group above the others. (ii) In some cases, the colonial language acts as a *lingua franca* i.e. a communication language (as in Côte d'Ivoire). Results when controlling for the number of languages in the country and the number of ethnic groups are shown in the third column of Table 1. The number of languages is taken from Glottolog. This open-source database provides a comprehensive catalog of the world's languages, language families, and dialects[6]. The number of ethnic groups is inferred from the Ethnic Power Relationship dataset (Vogt et al., 2015). Results are robust to these additional control variables, comforting that French ex-colonies are likelier to keep the linguistic *status quo* in education.

Mechanisms. I investigate a few mechanisms underlying the colonial legacy on contemporary linguistic policy choices. The first channel is the crucial role of precolonial missions in the spread, documentation, and early use of local languages in education. Table A.2 in Appendix shows the unequal distribution of missions across colonial empires: on average, the number of Protestant missions per country in 1925 was 53 in British-controlled countries and 9 in French territories. The use of local languages by teachers at school started with the missionaries (Berman, 1974; Johnson, 1967): speaking the local language was seen as an advantage in the Christianization process paired with education provision. The Protestant missions particularly emphasized the use of African languages by contributing actively to the writing development of these languages. Their final objective was to translate the Bible into the local languages to spread God's words more easily (Eisenstein, 1980)[7].

Another channel is the language documentation that happened under colonial rule. Indeed, besides the role of precolonial missions, we observe a differential trend in the number

[6]Hammarström, Harald & Forkel, Robert & Haspelmath, Martin & Bank, Sebastian. 2024. Glottolog 5.0. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://glottolog.org)

[7]Groves (Groves, 1958) listed 394 Bible translations into African languages made between 1805 and 1954 by Protestant missionaries.

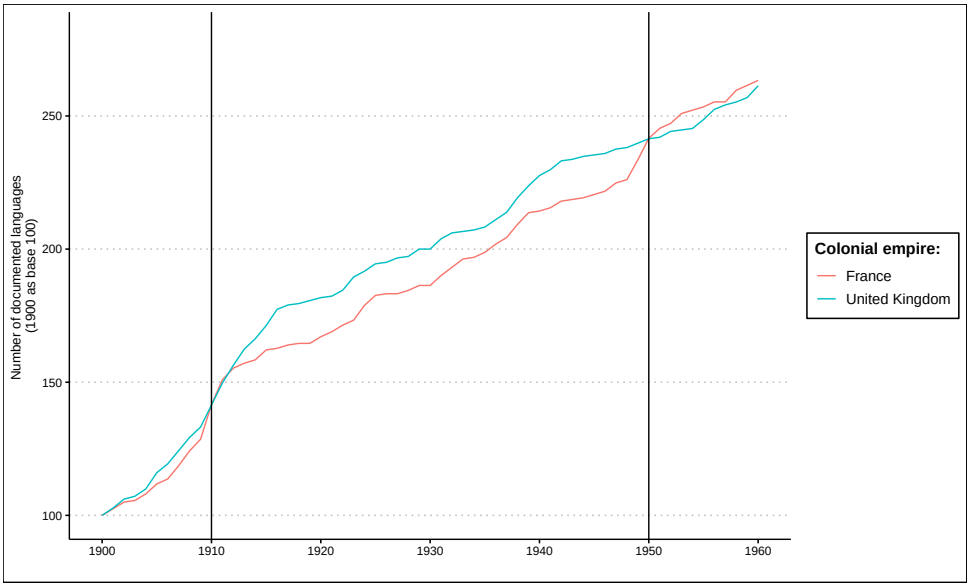

Figure 2: Linguistic production rate in the French and British colonial empires in Africa between 1900 and 1960

Notes: Data on languages and development come from Glottolog 5.0. The year 1900 is taken as a base 100. Similar results are found when 1990 is taken as base 100.

of fully documented language among the French and British colonial empires. Figure 2 is obtained using the Glottolog database; I consider a language as fully documented (or developed) in a year i if a grammar reference was published in the given year. Taking 1900 as a base 100 (controlling for linguistic development due to precolonial missions), we observe a systematic higher rate of linguistic production in the British colonial empire until 1950.

To isolate only the contemporary colonial legacy, I control for the number of missions in 1925 per country using newly published data by Bastian Becker (Becker, 2022) and the share of fully documented languages at the independence (taking 1900 for never-colonized countries) in Column (4) of Table 1. I still find a positive effect of being under British colonial rule compared to the French administration on introducing local languages in education. This effect is even larger than for the naive comparison.

Dependent Variable:	Dummy for introducing local languages in education			
Model:	(1)	(2)	(3)	(4)
Variables				
Colonizer dummy:				
Great Britain	3.20*** (1.15)	5.60** (2.23)	5.00** (2.19)	4.47* (2.39)
Other colonizer	0.201 (0.799)	1.75 (1.63)	1.22 (2.09)	0.049 (2.35)
Not colonized	0.894 (1.34)	0.517 (5.19)	-0.447 (14.0)	-0.135 (53.9)
Controls: Differential endowments		Yes	Yes	Yes
Controls: Ethnic + Language			Yes	Yes
Controls: Mission + Documentation				Yes
Fit statistics				
Observations	54	53	51	51
Pseudo R^2	0.22541	0.54082	0.61569	0.65217

IID standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The dependent variable is a dummy that is equal to 1 if the country replaced the use of the colonial language in education by a local language. The model estimated is a binomial logistic regression model. Differential endowments controls stand for the log of the population, the population density, the population share of muslim in 2000, latitude and longitude, and a dummy for landlock countries. Ethnic and language controls stand for the number of ethnic groups (EPR dataset) and the number of languages (Glottolog) in the country in 2000. Mission and documentation controls stand for the number of mission in 2025 (Becker, 2018) and the share of fully documented languages at the independence (1900 as a reference for non-colonized coutries).

Table 1: Colonial legacy on the bilingual education policy choice

3.2 Ethnolinguistic diversity

Ethnolinguistic diversity is another potential determinant of contemporary policy choices in Africa that has been thoroughly investigated in the literature. Alberto Alesina and Eliana La Ferrara provide a review of channels linking ethnolinguistic diversity[8] to poor economic growth (Alesina and La Ferrara, 2005). It can be particularly salient when studying the language of instruction choices. Indeed, promoting one (or more) local languages as a medium of instruction is equal to promoting the associated ethnolinguistic group. The fear of increasing ethnic salience is a solid political motive that could prevent the introduction of national languages in education. David Laitin and Rajesh Ramachandran provided a conceptual framework modeling the choice of official languages in a recent paper (Laitin and Ramachandran, 2022). They directly map high ethnolinguistic diversity with high likelihood of retaining the colonial language as official or co-official. They provide empirical evidence on this link, using different definitions of ethnolinguistic diversity. I am to provide similar evidence for the linguistic choices made by African government regarding the medium of instruction.

Measures and Data. As highlighted by Alberto Alesina and Eliana La Ferrara (Alesina and La Ferrara, 2005), ethnolinguistic diversity is a concept encompassing very diverse definitions. I start by following the same methodology as David Laitin and Rajesh Ramachandran and test my hypothesis using their selected measures of ethnolinguistic diversity:

1.1 An ethnolinguistic fractionalization (ELF) index: this index is taken from previous work done by Klaus Desmet and co-authors (Desmet, Weber, and Ortuño-Ortín, 2009), who define it as "the probability of two randomly chosen individuals being from different groups; [ELF] does not take into account the distances between the different groups". I consider another measure of this index derived from the recent work of Klaus Desmet and co-authors (Desmet, Gomes, and Ortuño-Ortín, 2020); it is obtained by averaging local ELF indexes. This computation method gives slightly different results;

[8] For consistency, I will refer throughout this paper to ethnolinguistic diversity instead of ethnolinguistic fractionalization, but the reader should be aware that the two are considered alike in this study.

1.2 Greenberg Index (GI): this measure was introduced by Greenberg in 1956 (Greenberg, 1956). It is widely seen as a generalization of the ELF, because it takes into account distance between languages.

I complete the list by adding additional measures of ethnolinguistic diversity:

2.1 A polarization index that does not consider distance between languages (Reynal-Querol, 2002). Compared to the other diversity indexes, it gives more weight to the relative size of each group;

2.2 The same polarization index, but with distance between languages included (Esteban and Ray, 1994);

2.3 A peripheral heterogeneity index (Desmet, Weber, and Ortuño-Ortín, 2009), which is a variant of the GI that takes into account polarization[9];

2.4 The number of languages, as given in the Glottolog database.

Finally, I also consider other measures related to ethnicity and not only languages, using the Ethnic Power Relationship dataset[10]. Indeed, even if the linguistic map in Africa is highly correlated with the distribution of ethnic groups, it does not capture the same reality. In Rwanda, Hutus and Tutsis share the same language, Kinyarwanda, but belong to different ethnic groups.

3.1 The number of politically relevant ethnic groups (EPR);

3.2 An index of pure ethnic fractionalization (EPR);

3.3 The share of the country covered by the largest ethnic group (EPR).

Results. Regression tables are in Table 2. Using a logistic regression model, I find close results to what David Laitin and Rajesh Ramachandran found for official languages with a linear probability model: it seems that linguistic diversity seems to matter when

[9]For more details on these different measures and computations, see the work of Klaus Desmet, Ignacio Ortuño-Ortìn and Shlomo Weber (Desmet, Weber, and Ortuño-Ortín, 2009)

[10]The number of ethnic groups has been computed in many different projects and studies. I use here this dataset because it captures the "politically relevant ethnic groups".

policymakers decide to promote bilingual education at scale (Columns (1) et (2)). The more linguistically diverse a country is, the less likely the medium of instruction is not a colonial language. However, distance between languages is not perceived as a determinant factor for policy making decisions (Columns (3), (5) and (6)). Similarly, pure ethnic diversity, even if positively correlated with linguistic diversity, is not an underlying motive for maintaining the linguistic *status quo* (Columns (8), (9), and (10)).

Dependent Variable:	Dummy for introducing local languages in education									
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables										
Ethnolinguistic diversity variables:										
ELF (Desmet at al., 2009)	-14.7*									
	(8.82)									
ELF (Desmet et al., 2020)		-7.21*								
		(3.89)								
Greenberg Index			-3.38							
			(3.40)							
Polarization index				-22.3						
				(17.6)						
Polarization index with distance					25.2					
					(29.5)					
Peripheral heterogeneity index						1.70				
						(7.30)				
Number of languages							-0.019*			
							(0.011)			
Number of ethnic groups								0.128		
								(0.277)		
Ethnic ELF									8.29	
									(6.60)	
Area share of the largest ethnic group										1.40
										(3.90)
Controls: Colonizer dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls: Differential endowments	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls: Mission + Documentation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics										
Observations	53	52	53	53	53	52	53	51	51	51
Pseudo R^2	0.74806	0.62454	0.56169	0.58148	0.55974	0.54505	0.59948	0.61466	0.67706	0.61290

IID standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The dependent variable is a dummy that is equal to 1 if the country replaced the use of the colonial language in education by a local language. The model estimated is a binomial logistic regression model. Colonizer dummy stand for Great Britain dummy, other colonizer dummy and non-colonized dummy (French as a reference). Differential endowments controls stand for the log of the population, the population density, the population share of muslim in 2000, latitude and longitude, and a dummy for landlock countries. Mission and documentation controls stand for the number of mission in 2025 (Becker, 2018) and the share of fully documented languages at the independence (1900 as a reference for non-colonized coutries). All the variables on ethnicity are computed from the EPR dataset. The number of languages for each country is computed using the Glottolog database. All the other variables on ethnolinguistic diversity are taken from datasets used for papers published by Klaus Desmet and his co-authors in 2009 and 2020.

Table 2: Ethnolinguistic diversity on the bilingual education policy choice

3.3 Contemporary political features

3.3.1 Communism

One feature of the post-colonial political area in Africa that is majorly understudied in empirical economic history is communism. The spread of the communist thought throughout the continent started right after WWII, with the beginning of the Cold War. The independence movements matched partly Soviet anti-imperialism at that time. Sékou Touré for Guinea, Modibo Keïta in Mali, and Kwame Nkrumah in Ghana were even awarded a Lenin Peace Prize between 1961 and 1963. These so-called socialist countries, very close to the USSR, promoted the use and development of national languages in all domains. In Mali, for instance, an International UNESCO Conference was organized in 1966 to fix the syntax and grammar of the different Malian languages with the objective of scaling up bilingual education afterwards.

Measures. Using different data sources, I examine the association between communism and linguistic reforms promoting bilingual education. I test whether the socialist movement was a factor in replacing colonial languages in education. I first use the UN votes: for every vote since the independence, I looked at whether the vote of each African country was similar to the USSR or its ex-colonizer, or none of the two, from 1960 to 1991. I use this measure as a proxy for being part of the East Block, part of the West Block, and part of the non-aligned movement during the Cold War. The second measure is a dummy indicating if a country c is declared as a communist country for the year y in the Bjørnskov-Rhode dataset (Bjørnskov and Rode, 2020). Contrary to the estimations above that were done at the country level, I use here a dataset with one observation per year per country, covering all African countries from independence to 2023 (1900 is taken as a reference for countries that were not colonized).

Results. I weigh all observations so that every country is equally represented in the final dataset, and cluster the standard errors at the country level. I use the same control variables as before and allow variation for time-varying ones (such as population, for instance). Table 3 shows the results. Odd columns display the results when controlling for the ethnolinguistic fractionalization index, which was previously shown to be a determinant

in explaining linguistic policy choices. The two measures are positively correlated to the introduction of bilingual education. Even if the coefficient associated with being communist is reduced in magnitude when adding ELF as a control, it is still positive and significant. Hence, communism as a shared political movement during the post-colonial area in Africa is positively associated with the progressive introduction of bilingual education.

3.3.2 Democracy

To investigate the role of political institutions more in-depth, I do the same work using the Political Regime Characteristic Database (PolityV) hosted by the World Bank. I explore whether the country's autocratic or democratic features is correlated with policymakers' decisions regarding the languages spoken in public schools by using the polity score, which ranges from -10 (very autocratic) to +10 (very democratic).

Table A.3 in the Appendix displays the results. The democracy score does not seem to be a good predictor of the probability of introducing bilingual education at scale. When dichotomizing the polity score into a dummy indicating democracy (higher or equal to 0) and autocracy (below 0) only, I do not observe robust results to the introduction of ELF as a control variable.

4 Aggregated effects on education

To the best of my knowledge, there is very limited evidence on the long-term effects of bilingual education at scale. Rajesh Ramachandran (Ramachandran, 2017) and Yared Seid (Seid, 2016) use the 1994 mother-tongue reform in Ethiopia to assess the impact on education outcomes. They find striking positive results regarding quality (literacy and reading skills) and quantity (schooling) education measures. However, these findings could be improved by a more robust identification strategy.

4.1 Data and Coverage

Coverage. I provide here aggregated long-term descriptive effects on eighteen countries. Out of the 54 African countries, I selected all countries satisfying the following criteria: (i)

Dependent Variable:	Dummy for introducing local languages in education			
Model:	(1)	(2)	(3)	(4)
Variables				
Share of UN votes similar to USSR	1.72*** (0.593)	2.22*** (0.696)		
Communist dummy			1.31 (1.14)	2.39*** (0.522)
Controls: Ethnolinguistic diversity		Yes		Yes
Controls: Colonizer dummy	Yes	Yes	Yes	Yes
Controls: Differential endowments	Yes	Yes	Yes	Yes
Controls: Ethnic + Language	Yes	Yes	Yes	Yes
Controls: Mission + Documentation	Yes	Yes	Yes	Yes
Fit statistics				
Observations	2,635	2,635	2,716	2,716
Pseudo R^2	1	1	1	1

Clustered (country) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The dependent variable is a dummy that is equal to 1 if the country replaced the use of the colonial language in education by a local language. The model estimated is a binomial logistic regression model. Colonizer dummy stand for Great Britain dummy, other colonizer dummy and non-colonized dummy (French as a reference). Differential endowments controls stand for the log of the population, the population density, the population share of muslim in 2000, latitude and longitude, and a dummy for landlock countries. Ethnic and language controls stand for the number of ethnic groups (EPR dataset) and the number of languages (Glottolog) in the country in 2000. Mission and documentation controls stand for the number of mission in 2025 (Becker, 2018) and the share of fully documented languages at the independence (1900 as a reference for non-colonized coutries). The communist dummy is equal to 1 if the government is communist according to the Bjornskov-Rhode dataset.

Table 3: Salience of communism on the bilingual education policy choice

experienced a change in the languages used in education, (ii) for which data on education were available, and (iii) with sufficiently long lead time[11]. Indeed for every country in the analysis, I look at birth cohorts from 10 years minimum before the implementation of the bilingual education reform up to 10 years maximum after the reform, that are adult at the time of the survey. The ten-year range is motivated by two conflicting reasons. On one side, I want to put a limit on the extent to which the estimation can capture other reforms' effects that happened roughly at the same time. On the other side, education reforms take time to be implemented, and only a few years around the reform included in the study would be not enough to capture the effect.

I consider a birth cohort as treated when people in this cohort supposedly started primary education in a local language[12]. That means that if a person is 9 years old when the bilingual education is passed, I do not consider this person as treated. Finally, I restrict the sample to people from 15 to 49 years old to match the Demographic and Health Survey (DHS) sample age range. The exact time coverage for each country is detailed in the Appendix in Figure A.2.

Data. I use census data that is publicly available on IPUMS and DHS data when the latter is missing. Figure A.2 details the survey type in the Appendix. I end up with a sample of almost three million people and eighteen countries. For analysis purposes, I reweighed all observations so that the weight sum is one for each country.

I use education outcomes that are comparable across censuses and DHS data. On the demand side, I expect bilingual education to expand the education coverage for people who were out of school and to increase the level of schooling for people already in school, as stated in the literature (Laitin, Ramachandran, and Walter, 2019, Piper, Zuilkowski, and Ong'ele, 2016). On the supply side, pilot studies highlighted positive results on literacy and test scores (Benson, 2000, Benson, 2002). Following previous work on the topic, I use as outcomes school attendance, years spent at school, and literacy (both self-reported and tested[13]).

[11] I exclude Sudan from this analysis because of a lack of state capacity in the country that hinders the policy implementation of the bilingual education law.

[12] I use the entry age in primary school for every country given by UNESCO.

[13] In many surveys, respondents self-reported literacy. Even for recent DHS surveys that test for liter-

4.2 Results

Estimation strategy. I investigate long-term shared trends of education outcomes before and after implementing a bilingual education law. As stated previously, this study focuses only on the *de jure* aspect of the law, regardless of its actual implementation. The section 4.4 will give more details about implementation.

For every education outcomes Y included in this analysis, I estimate the following regression:

$$Y_{itc} = \alpha + \beta * T_{t,c} + \theta_1 c * time + \theta_2 c * time^2 + \theta_3 c * time^3 + \eta * X_{itc} + \rho_c + \omega_t + \epsilon_{itc}$$

The time variable here ranges from 0 to 20 and is equal to 10 when one country's first birth cohort is considered fully exposed to bilingual education. I allow the relationship between education and time to be non-linear, controlling for second and third-order polynomials. I use strata[14] and year fixed effects.

The estimated model is a binomial logistic probability model when the outcome takes the values 0 and 1. This choice makes interpreting the coefficient difficult, as it is limited to the direction and significance of the coefficients. For the other outcomes, I use an OLS model.

Global results. Table 4 shows the results for all education outcomes. Except for the number of schooling years, the introduction of local languages in education correlates with higher literacy and school attendance trends. I can't interpret these results as causal: many other reforms can happen simultaneously, and my estimates can capture other reform effects as well. Moreover, the estimates can also capture the impact of a shared time shock across countries. Covering many different countries at different periods makes these results more robust to these concerns.

acy skills, people are assumed to be literate after a certain number of schooling years. Justin Sandefur wrote an enlightening blog post in CGDEV about this issue and plotted striking literacy trends in the case of Nigeria: https://www.cgdev.org/blog/how-usaids-demographic-and-health-surveys-overestimate-literacy-around-world

[14]Strata are either the strata given in the DHS, or the lowest geographical level provided in the census

Dependent Variables:	Literacy	Tested literacy	School attendance	Years of schooling
Model:	(1)	(2)	(3)	(4)
	Logit	Logit	Logit	OLS
Variables				
After the bilingual education reform	0.079**	0.072**	0.073**	0.058
	(0.031)	(0.035)	(0.032)	(0.044)
Time (from 0 to 20)	0.077***	0.069***	0.101***	0.129***
	(0.009)	(0.010)	(0.010)	(0.012)
$Time^2$	-0.004***	-0.004**	-0.006***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.002)
$Time^3$	0.0001***	0.0001**	0.0002***	0.0003***
	(4.17×10^{-5})	(4.64×10^{-5})	(4.38×10^{-5})	(6.17×10^{-5})
Controls: Sex	Yes	Yes	Yes	Yes
Fixed-effects				
Strata FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Fit statistics				
Observations	2,910,406	2,894,785	2,987,337	2,988,397
Pseudo R^2	1	1	1	-153,103.1
R^2				0.41849

Clustered (at the country level) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The model estimated is a binomial logistic regression model for the 3 first columns, and an OLS for the last one. Strata fixed effects stands for the strata given in the DHS surveys and the lowest geographical level for the census surveys. Cluster at the country level stand for DHS cluster for the DHS surveys and the household level for the census surveys. All dependant variables are dummies, except for years of schooling.

Table 4: Descriptive global effects of bilingual education

Respondents' heterogeneity. Heterogeneous analysis gives insight into what drives these positive results. Results can be found in Tables A.4 and A.5 in Appendix. Women experience more positive returns from bilingual education. This result is consistent with evidence on the positive interaction between gender and mother-tongue education (Benson and Wong, 2019). Among people who attended school, we observe larger effects for people who completed primary education. Figure 3 details the effect by years of schooling completed: the higher the schooling level, the more positive the impact of bilingual education. It is in line with the UNESCO recommendation for a late-exit bilingual education model (Ball, 2011, Trudell, 2023). In their paper on official languages, David Laitin and Rajesh Ramachandran (Laitin and Ramachandran, 2022) find empirically similar heterogeneity with respect to the number of schooling years (figure 9). It is also noticeable that literacy returns are negative or null for people who have less than two years of education. Hence, there is a limit above which bilingual education seems to be beneficial for students. This conclusion is enlightening when knowing that in seven African countries, the transition between the local language(s) and the colonial one happens at maximum the third grade (Figure A.1 in Appendix).

Policy characteristics. Finally, I look at different features of the bilingual education reforms that can partially explain positive or negative results at the country level (Tables A.7 and A.8 in Appendix). So far, I considered all bilingual education laws as identical and comparable. However, they differ in many dimensions. Figure A.1 shows the wide range of grades at which the colonial language is used according to these laws. Laws also differ regarding the languages introduced: some countries have introduced only one language in education, like Tanzania with Swahili. In other countries, the law states that education is supposedly given in the student's mother tongue, as in Nigeria or Kenya. I detail languages used theoretically at the beginning of primary education and the grade at which the LoI switches for the colonial language for every country considered in this analysis in Table A.6 in the Appendix.

When doing a heterogeneity analysis of policy features, I observed that only mother-tongue policies correlate with more student learning at scale. I find similar results for policies with a transition before the start of secondary education.

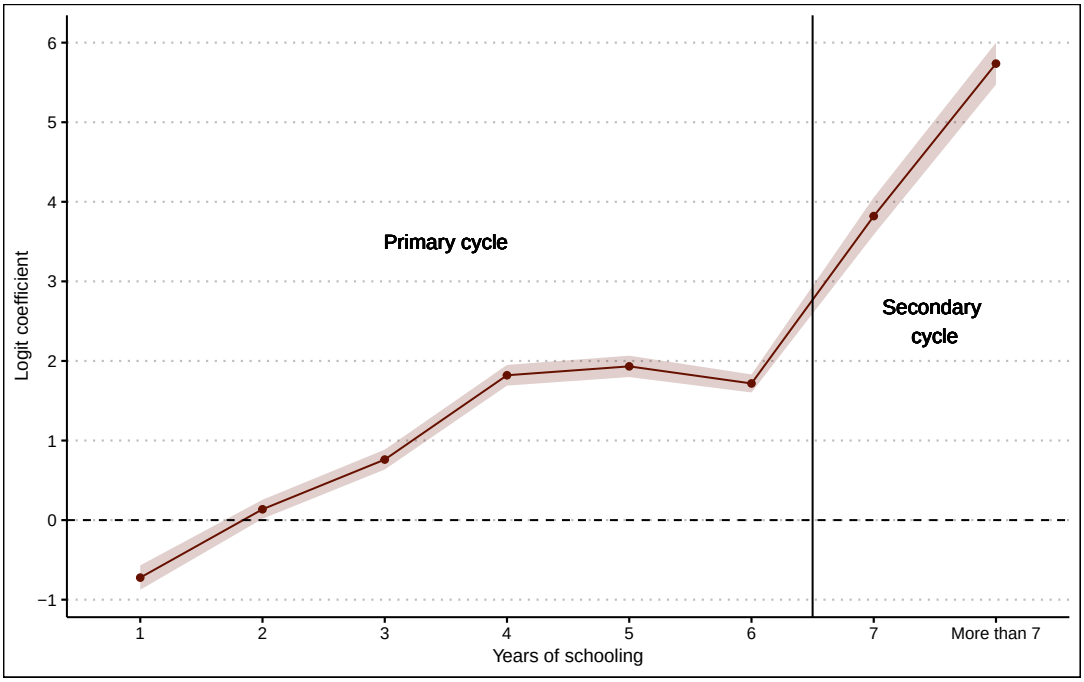

Figure 3: Heterogeneous effects of bilingual education on literacy by years of schooling

Notes: Point estimates are estimated through a binomial logistic regression regressing exposure to bilingual education on literacy, with the sex as control variable, same fixed effects and cluster standards errors. The exposure to bilingual education is interacted with the number of schooling year completed by the respondent. 95% confidence intervals are represented by shadowed areas. The rupture between primary and secondary cycle is estimated between the sixth and seventh years of education, but this limit can vary across countries.

4.3 Robustness checks

Effects on the no-school sample. I estimate the same effect on the sample of people who did not attend school. Results in Table A.9 in the Appendix show no impact on literacy using this subsample. It indicates that people attracted to school because of the bilingual model were likely not to be the top literacy performers among people who did not go to school.

Intermediate years. It takes time for a country-wide policy to be implemented. I test it by removing the intermediate birth cohorts from the analysis, for which implementation is partial. For instance, I consider that people in grade 2 when the bilingual education passed were not affected by the law for the previous data analysis. I remove these inter-

mediate birth cohorts who entered school up to 5 years before implementing the bilingual education law. I find more robust results, both in magnitude and significance (Table A.10 in Appendix).

Country level. As an additional robustness check, I drop one country at a time and look at how the estimates vary (Figure A.3 in Appendix). The average effects are pretty stable, except for the dropping of two countries, Ghana and Tanzania. In these two countries, the analysis at the country level (Figure A.4 in Appendix) gives similar results: the bilingual education laws seem to be correlated with largely positive results in terms of education quantity (attendance) and quality (learning). To better understand these results, a better overview of the implementation of these educational policies is needed.

4.4 Implementation

Salience. Little is said about implementing the reforms in the literature assessing the impact of mother-tongue education. Some recent papers point to the salience of implementation when studying education policies. Benjamin Piper and his co-authors (Piper, Zuilkowski, and Ong'ele, 2016) stressed it in Kenya: "Implementation of the MT[mother-tongue] program faced challenges because many educators were not speakers of the languages, some communities resisted mother tongue instruction, and some areas were more language heterogeneous." A working paper published by Noam Agrist and Rachael Meager (Angrist and Meager, 2023) investigates the channels explaining the heterogeneity of effects in education programs. They show that implementation (defined as the comparison between intention-to-treat and treatment-on-the-treated) is key. In the case of Mali, for example, Samba Traoré (Traoré, 2001) detailed the expected implementation phases in a UNESCO policy report in 2001, while Ag Agouzoum Alou documented the realized implementation aspect in 2020 (Alou, 2020). Both do not fully coincide, showing that things did not go perfectly as planned.

Piloting. Because implementation is key, piloting a policy before its scale-up is a must-have for policymakers. However, I do not see such a pattern when exploiting the companion database I built for this work (see section 2.2). I display in Figure A.5 in the Appendix

all pilot programs I have listed in the database: pilot programs do not precede the laws or program evaluation. One example of unprepared linguistic education law is Madagascar. In 1975, the newly elected President Didier Ratsiraka promoted many reforms against the French influence. One of them was the exclusive use of Malagasy in education. However, the law was passed with no additional training for teachers nor the proper design of a new curriculum. Confronted by the failure of the education reform, the government reintroduced French to the curriculum in 1985. Years after, literacy rates for these birth cohorts were still lower for people who went to school during the Ratsiraka governance (Rapanoël, 2008, Blum, 2011).

Linguistic composition. Collecting accurate qualitative evidence about implementation quality would be costly and challenging to gather. I use the change in the linguistic composition here as a proxy for the quality of the implementation of the linguistic reform of interest. The linguistic composition of one country is the fraction of the population that speaks a language for every language spoken within the country. If one language is promoted at school, I rationally expect the fraction of people speaking this language to increase.

I do not find any evidence of a change in the linguistic composition for the four countries for which data on languages spoken are available[15]. As shown in Figure A.6 in the Appendix, the share of each language is pretty stable over time.

5 Conclusion

Very little is known about the importance of the languages of instruction in Africa. This paper has highlighted striking determinants explaining why some African countries replaced the colonial language with a local one in primary education. Using a novel dataset detailing linguistic reforms since independence, I provide data-driven evidence on colonial legacy and present-day characteristics that drive policymakers regarding the LoI choice.

In the last section, I provide descriptive evidence of the positive effects of bilingual education on schooling and learning using large micro datasets. These estimates are not

[15]Refined data analysis is ongoing at the moment of the writing.

causal but give precious insight into at-scale linguistic policies. Further research is needed to understand better what works and what does not work in terms of bilingual education.

References

Albaugh, Ericka A. (Apr. 2014). *State-Building and Multilingual Education in Africa*. 1st ed. Cambridge University Press. DOI: 10.1017/CBO9781107323735.

Alesina, Alberto and Eliana La Ferrara (Sept. 2005). "Ethnic Diversity and Economic Performance". en. In: *Journal of Economic Literature* 43.3, pp. 762–800. DOI: 10.1257/002205105774431243.

Alou, Ag Agouzoum (Dec. 2020). "Problématique de l'introduction des langues maternelles dans le système éducatif". fr. In: *Asinag* 15. Number: 15 Publisher: Institut Royal de la Culture Amazighe (IRCAM), pp. 11–31.

Angrist, Noam and Rachael Meager (2023). "Implementation Matters: Generalizing Treatment Effects in Education". en. In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.4487496.

Ball, Jessica (2011). *Enhancing learning of children from diverse language backgrounds: Mother tongue-based bilingual or multilingual education in the early years*. Tech. rep. UNESCO.

Becker, Bastian (2022). "The Empire Within: Longitudinal Evidence on the Expansion of Christian Missions in Colonial Africa". en. In: *Journal of Historical Political Economy* 2.2, pp. 333–362. DOI: 10.1561/115.00000032.

Benson, Carol and Kevin M. Wong (Feb. 2019). "Effectiveness of policy development and implementation of L1-based multilingual education in Cambodia". en. In: *International Journal of Bilingual Education and Bilingualism* 22.2, pp. 250–265. DOI: 10.1080/13670050.2017.1313191.

Benson, Carolyn J. (Sept. 2000). "The Primary Bilingual Education Experiment in Mozambique, 1993 to 1997". In: *International Journal of Bilingual Education and Bilingualism* 3.3. Publisher: Routledge _eprint: https://doi.org/10.1080/13670050008667704, pp. 149–166. DOI: 10.1080/13670050008667704.

— (Dec. 2002). "Real and Potential Benefits of Bilingual Programmes in Developing Countries". In: *International Journal of Bilingual Education and Bilingualism* 5.6. Publisher: Routledge _eprint: https://doi.org/10.1080/13670050208667764, pp. 303–317. DOI: 10.1080/13670050208667764.

Berman, Edward H. (Dec. 1974). "African Responses to Christian Mission Education". en. In: *African Studies Review* 17.3, pp. 527–540. DOI: 10.2307/523799.

Bertocchi, Graziella and Fabio Canova (Dec. 2002). "Did colonization matter for growth?" en. In: *European Economic Review* 46.10, pp. 1851–1871. DOI: 10.1016/S0014-2921(01)00195-7.

Bjørnskov, Christian and Martin Rode (Apr. 2020). "Regime types and regime change: A new dataset on democracy, coups, and political institutions". en. In: *The Review of International Organizations* 15.2, pp. 531–551. DOI: 10.1007/s11558-019-09345-1.

Blum, Françoise (2011). "Madagascar 1972 : l'autre indépendance: Une révolution contre les accords de coopération". fr. In: *Le Mouvement Social* 236.3, p. 61. DOI: 10.3917/lms.236.0061.

Bold, Tessa et al. (Jan. 2017). *What Do Teachers Know and Do? Does It Matter? Evidence from Primary Schools in Africa*. en. World Bank, Washington, DC. DOI: 10.1596/1813-9450-7956.

Bolt, Jutta and Dirk Bezemer (Jan. 2009). "Understanding Long-Run African Growth: Colonial Institutions or Colonial Education?" en. In: *The Journal of Development Studies* 45.1, pp. 24–54. DOI: 10.1080/00220380802468603.

Chanson, Aude (July 2012). "Le kiswahili : une politique linguistique d'exception au Tanganyika". fr. In: *Les Cahiers d'Afrique de l'Est / The East African Review* 45. Number: 45 Publisher: IFRA - Institut Français de Recherche en Afrique, pp. 49–65. DOI: 10.4000/eastafrica.492.

Cogneau, Denis and Alexander Moradi (Sept. 2014). "Borders That Divide: Education and Religion in Ghana and Togo Since Colonial Times". fr. In: *The Journal of Economic History* 74.3. Publisher: Cambridge University Press, pp. 694–729. DOI: 10.1017/S0022050714000576.

Daun, Holger (2000). "Primary Education in Sub-Saharan Africa-A Moral Issue, an Economic Matter, or Both?" In: *Comparative Education* 36.1. Publisher: Taylor & Francis, Ltd., pp. 37–53.

De Gaston, William (Oct. 2011). "Le signal communicationnel français face à la langue Évé : Pour une phénoménologie de la communication". fr. In: *Communiquer. Revue de com-*

munication sociale et publique 6. Number: 6, pp. 57–70. DOI: 10.4000/communiquer.452.

Desmet, Klaus, Joseph Flavian Gomes, and Ignacio Ortuño-Ortín (Mar. 2020). "The geography of linguistic diversity and the provision of public goods". en. In: *Journal of Development Economics* 143, p. 102384. DOI: 10.1016/j.jdeveco.2019.102384.

Desmet, Klaus, Shlomo Weber, and Ignacio Ortuño-Ortín (Dec. 2009). "Linguistic Diversity and Redistribution". In: *Journal of the European Economic Association* 7.6. Publisher: Oxford University Press / USA, pp. 1291–1318. DOI: 10.1162/JEEA.2009.7.6.1291.

Eastern, UNICEF, Basic Education Southern Africa Regional Office (ESARO), and Gender Equality (BEGE) Section (2016). *The impact of language policy and practice on children's learning: Evidence from Eastern and Southern Africa*. Tech. rep. United Nations Children's Fund (UNICEF).

Eisenstein, Elizabeth L. (1980). *The Printing Press as an Agent of Change*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781107049963.

Esteban, Joan and Debraj Ray (1994). "On the Measurement of Polarization". In: *Econometrica* 62.4, pp. 819–51.

Frankema, Ewout H.P. (Nov. 2012). "The origins of formal education in sub-Saharan Africa: was British rule more benign?" In: *European Review of Economic History* 16.4, pp. 335–355. DOI: 10.1093/ereh/hes009.

Greenberg, Joseph H. (Jan. 1956). "The Measurement of Linguistic Diversity". In: *Language* 32.1, p. 109. DOI: 10.2307/410659.

Groves, C. P. (Oct. 1958). "The Planting of Christianity in Africa. Vol. IV, 1914–1954." en. In: *London: Lutterworth Press.* 28.4, pp. 372–374. DOI: 10.2307/1157645.

Johnson, Hildegard Binder (Apr. 1967). "The Location of Christian Missions in Africa". In: *Geographical Review* 57.2, p. 168. DOI: 10.2307/213158.

Kébé, Abou Bakry and Mamadou Diop (Dec. 2021). "École et langues nationales au Sénégal. À la recherche d'une terminologie opérationnelle : regard sur le modèle bilingue pulaar-français d'A.R.E.D.:" fr. In: *Éla. Études de linguistique appliquée* N° 203.3, pp. 339–354. DOI: 10.3917/ela.203.0085.

La Porta, R (Mar. 1999). "The quality of government". In: *Journal of Law, Economics, and Organization* 15.1, pp. 222–279. DOI: 10.1093/jleo/15.1.222.

Laitin, David D. (1994). "The Tower of Babel as a Coordination Game: Political Linguistics in Ghana". In: *The American Political Science Review* 88.3. Publisher: [American Political Science Association, Cambridge University Press], pp. 622–634. DOI: 10.2307/2944799.

Laitin, David D. and Rajesh Ramachandran (May 2022). "Linguistic diversity, official language choice and human capital". In: *Journal of Development Economics* 156, p. 102811. DOI: 10.1016/j.jdeveco.2021.102811.

Laitin, David D., Rajesh Ramachandran, and Stephen L. Walter (Oct. 2019). "The Legacy of Colonial Language Policies and Their Impact on Student Learning: Evidence from an Experimental Program in Cameroon". en. In: *Economic Development and Cultural Change* 68.1, pp. 239–272. DOI: 10.1086/700617.

Mohohlwane, Nompumelelo et al. (2023). "Reading Skills Transfer Best from Home Language to a Second Language". en. In: *CGD Working Paper. Washington, DC: Center for Global Development* 647.

Piper, Benjamin, Stephanie S. Zuilkowski, and Salome Ong'ele (Nov. 2016). "Implementing Mother Tongue Instruction in the Real World: Results from a Medium-Scale Randomized Controlled Trial in Kenya". en. In: *Comparative Education Review* 60.4, pp. 776–807. DOI: 10.1086/688493.

Ramachandran, Rajesh (Oct. 2017). "Language use in education and human capital formation: Evidence from the Ethiopian educational reform". en. In: *World Development* 98, pp. 195–213. DOI: 10.1016/j.worlddev.2017.04.029.

Rapanoël, Allain Bruno Solofomiarana (2008). "Les problèmes de la langue d'enseignement à Madagascar: les dysfonctionnements de la loi 94-033". fr. In: *Kabaro, revue internationale des Sciences de l'Homme et des Sociétés*. L'Harmattan ; Université de La Réunion, Interethnicité et Interculturalité à l'île Maurice, pp. 153–171.

Reynal-Querol, Marta (Jan. 2002). "Political systems, stability and civil wars". en. In: *Defence and Peace Economics* 13.6, pp. 465–483. DOI: 10.1080/10242690214332.

Seid, Yared (Dec. 2016). "Does learning in mother tongue matter? Evidence from a natural experiment in Ethiopia". In: *Economics of Education Review* 55, pp. 21–38. DOI: 10.1016/j.econedurev.2016.08.006.

Traoré, Samba (2001). "La Pédagogie Convergente : son expérimentation au Mali et son impact sur le système éducatif". fr. In: *UNESCO, Bureau International d'Education*. Monographies-Innodata.

Trudell, Barbara (2023). *Early-exit language transitioning programming: The rationale, the benefits and the limitations*. Tech. rep. UNESCO.

Vogt, Manuel et al. (Oct. 2015). "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family". en. In: *Journal of Conflict Resolution* 59.7, pp. 1327–1342. DOI: 10.1177/0022002715591215.

World Bank (2018). *Effective Language of Instruction Policies for Learning*. Tech. rep.

Appendix

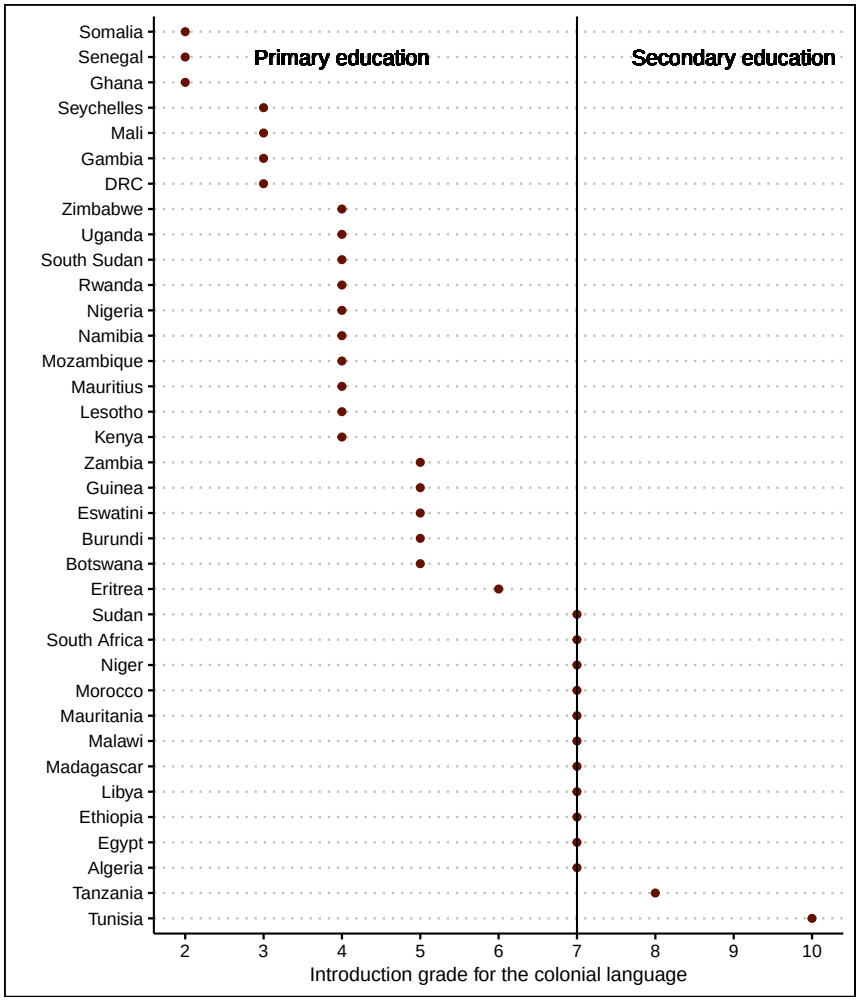

Figure A.1: Official grade of transition from the use of a local language towards the use of the colonial language

Notes: Author's computation. The transition grade indicated here is the one depict in the first law that enacted the introduction of a local language in instruction. Hence, the figure does not display the contemporary status for all countries. Secondary education does not start at the 7th grade for every country, but was taken as a reference to indicate the usual grade range for primary education.

Dependent Variable:	Dummy for introducing local languages in education
Model:	(1)
Variables	
Colonizer dummy:	
Great Britain	5.60**
	(2.23)
Other colonizer	1.75
	(1.63)
Not colonized	0.517
	(5.19)
Log population (in 2000)	1.22**
	(0.555)
Population density (in 2000)	0.003
	(0.007)
Population share of Muslims (in 2000)	0.037*
	(0.020)
Geographical controls	Yes
Fit statistics	
Observations	53
Pseudo R^2	0.54082

IID standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Geographical controls stand for latitude, longitude and a dummy for landlock countries.

Table A.1: Colonial legacy on the bilingual education policy choice - detailed

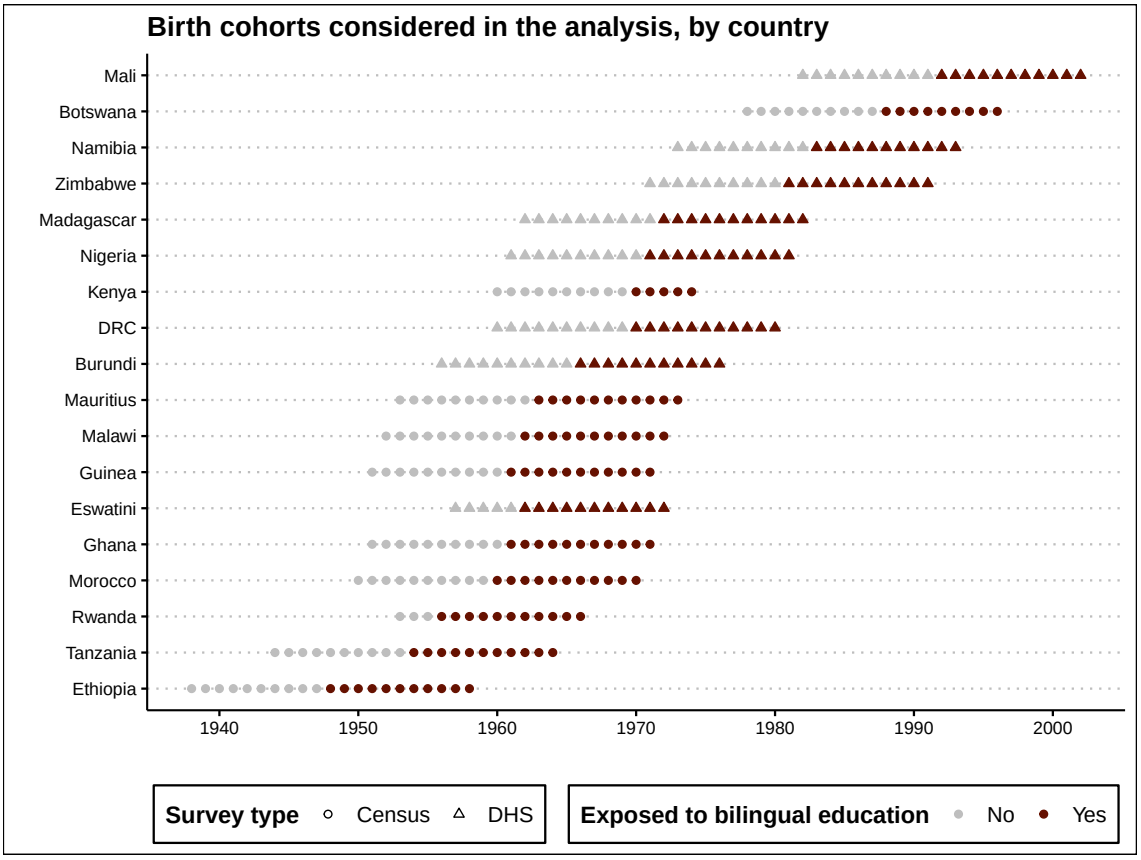

Figure A.2: Time and data coverage for the analysis presented in Section 4

Colonizer	Number of Catholic missions	Number of Protestant missions
Portugal	9	14
France	10	10
United Kingdom	17	52
Other	22	28

Table A.2: Average number of missions in 1925 per colonial empire

Dependent Variable:	Dummy for introducing local LoIs			
Model:	(1)	(2)	(3)	(4)
Variables				
PolityV score	-0.059*	-0.017		
	(0.033)	(0.045)		
Democracy dummy (= 1 if PolicyV score $\geq$ 0)			-0.637	-0.104
			(0.540)	(0.480)
Controls: Ethnolinguistic diversity		Yes		Yes
Controls: Colonizer dummy	Yes	Yes	Yes	Yes
Controls: Differential endowments	Yes	Yes	Yes	Yes
Controls: Ethnic + Language	Yes	Yes	Yes	Yes
Controls: Mission + Documentation	Yes	Yes	Yes	Yes
Fit statistics				
Observations	2,722	2,722	2,722	2,722
Pseudo R^2	1	1	1	1

Clustered (country) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The dependent variable is a dummy that is equal to 1 if the country replaced the use of the colonial language in education by a local language. The model estimated is a binomial logistic regression model. Colonizer dummy stand for Great Britain dummy, other colonizer dummy and non-colonized dummy (French as a reference). Differential endowments controls stand for the log of the population, the population density, the population share of muslim in 2000, latitude and longitude, and a dummy for landlock countries. Ethnic and language controls stand for the number of ethnic groups (EPR dataset) and the number of languages (Glottolog) in the country in 2000. Mission and documentation controls stand for the number of mission in 2025 (Becker, 2018) and the share of fully documented languages at the independence (1900 as a reference for non-colonized coutries). Ethnolinguistic diversity control stands for the ELF index computed by Klaus Desmet and his co-authors (2009).

Table A.3: Democracy determinant on the bilingual education policy choice

Dependent Variables:	Literacy	Tested literacy	School attendance	Years of schooling
Model:	(1)	(2)	(3)	(4)
	Logit	Logit	Logit	OLS
Variables				
After the education reform x Men	-0.146***	-0.152***	-0.133***	-0.129***
	(0.028)	(0.032)	(0.030)	(0.040)
After the education reform x Women	0.151***	0.149***	0.145***	0.123***
	(0.033)	(0.038)	(0.035)	(0.048)
Time (from 0 to 20)	0.079***	0.073***	0.103***	0.131***
	(0.009)	(0.010)	(0.010)	(0.012)
$Time^2$	-0.004***	-0.004***	-0.006***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.002)
$Time^3$	0.0001***	0.0001**	0.0002***	0.0003***
	(4.17×10^{-5})	(4.63×10^{-5})	(4.38×10^{-5})	(6.17×10^{-5})
Controls: Sex	Yes	Yes	Yes	Yes
Fixed-effects				
Strata FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Fit statistics				
Observations	2,910,406	2,894,785	2,987,337	2,988,397
Pseudo R^2	1	1	1	-153,105.6
R^2				0.41862

Clustered (at the country level) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The model estimated is a binomial logistic regression model for the 3 first columns, and an OLS for the last one. Strata fixed effects stands for the strata given in the DHS surveys and the lowest geographical level for the census surveys. Cluster at the country level stand for DHS cluster for the DHS surveys and the lowest geographical level for the census surveys. All dependant variables are dummies, except for years of schooling.

Table A.4: Descriptive global effects of bilingual education - Heterogeneity by sex

Dependent Variables:	Literacy	Tested literacy	Years of schooling
Model:	(1)	(2)	(3)
	Logit	Logit	OLS
Variables			
After x Did not complete primary education	-1.47***	-1.80***	-3.44***
	(0.076)	(0.089)	(0.042)
After x Completed primary education	1.41***	1.08***	1.15***
	(0.078)	(0.081)	(0.040)
Time (from 0 to 20)	0.039*	0.037	0.066***
	(0.023)	(0.024)	(0.012)
$Time^2$	-0.004	-0.002	-0.008***
	(0.003)	(0.003)	(0.002)
$Time^3$	0.0001	4.53×10^{-5}	0.0003***
	(0.0001)	(0.0001)	(5.41×10^{-5})
Controls: Sex	Yes	Yes	Yes
Fixed-effects			
Strata FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Fit statistics			
Observations	1,019,811	1,008,212	1,458,885
Pseudo R^2	1	1	-115,568.2
R^2			0.46510

Clustered (at the country level) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The model estimated is a binomial logistic regression model for the 2 first columns, and an OLS for the last one. Strata fixed effects stands for the strata given in the DHS surveys and the lowest geographical level for the census surveys. Cluster at the country level stand for DHS cluster for the DHS surveys and the lowest geographical level for the census surveys. All dependant variables are dummies, except for years of schooling.

Table A.5: Descriptive global effects of bilingual education - Heterogeneity by completion of primary education

Country	LoI	Transition grade
Botswana	Tswana	5
Burundi	Rundi	5
Democratic Republic of the Congo	Kinshasa Lingala, Congo Swahili Luba-Lulua, Eastern Kikongo	3
Eswatini	English, Swati	5
Ethiopia	Amharic	Secondary education
Ghana	Fante, Akuapem Twi, Asante Twi, Ewe, Dagbani, Adangme, Ga, Kasem, Nzima	2
Guinea	Eastern Maninkakan, Susu, Pulaar, Northern Kissi, Bassari-Tanda, Loma (Liberia), Wamey, Guinea Kpelle	5
Kenya	Mother tongue	4
Madagascar	Plateau Malagasy	Secondary education
Malawi	Nyanja	Secondary education
Mali	Bambara, Soninke, Pulaar, Western Maninkakan, Supyire Senoufo, Mamara Senoufo, Bomu, Donno So Dogon, Koyraboro Senni Songhai, Hainyaxo Bozo, Tamasheq, French	3
Mauritius	Mother tongue	4
Morocco	Standard Arabic	Secondary education
Namibia	Mother tongue	4
Nigeria	Mother tongue	4
Rwanda	Kinyarwanda	4
United Republic of Tanzania	Swahili	8
Zimbabwe	Shona, Zimbabwean Ndebele	4

Note: Information about languages are taken from Glottolog 5.0

Table A.6: Characteristic of the bilingual education reforms

Dependent Variables:	Literacy	Tested literacy	School attendance	Years of schooling
Model:	(1)	(2)	(3)	(4)
	Logit	Logit	Logit	OLS
Variables				
After x No mother tongue education	0.058*	0.046	0.034	0.041
	(0.032)	(0.037)	(0.034)	(0.044)
After x Mother tongue education	0.178***	0.184***	0.293***	0.061
	(0.035)	(0.037)	(0.037)	(0.044)
Time (from 0 to 20)	0.075***	0.066***	0.097***	0.130***
	(0.009)	(0.010)	(0.010)	(0.012)
$Time^2$	-0.004***	-0.003**	-0.006***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.002)
$Time^3$	0.0001***	9.7×10^{-5}**	0.0002***	0.0003***
	(4.17×10^{-5})	(4.64×10^{-5})	(4.38×10^{-5})	(5.94×10^{-5})
Controls: Sex	Yes	Yes	Yes	Yes
Fixed-effects				
Strata FE	Yes	Yes	Yes	
Year FE	Yes	Yes	Yes	
strata_fine_country				Yes
Fit statistics				
Observations	2,910,406	2,894,785	2,987,337	2,988,397
Pseudo R^2	1	1	1	-147,276.1
R^2				0.51415

Clustered (at the country level) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The model estimated is a binomial logistic regression model for the 3 first columns, and an OLS for the last one. Strata fixed effects stands for the strata given in the DHS surveys and the lowest geographical level for the census surveys. Cluster at the country level stand for DHS cluster for the DHS surveys and the lowest geographical level for the census surveys. All dependant variables are dummies, except for years of schooling.

Table A.7: Descriptive global effects of bilingual education - Heterogeneity by languages used at school

Dependent Variables:	Literacy	Tested literacy	School attendance	Years of schooling
Model:	(1)	(2)	(3)	(4)
	Logit	Logit	Logit	OLS
Variables				
After x Transition during prim educ	0.111***	0.098**	0.101***	0.028
	(0.034)	(0.038)	(0.035)	(0.047)
After x Transition during sec educ	0.003	0.012	0.008	0.155***
	(0.030)	(0.033)	(0.031)	(0.043)
Time (from 0 to 20)	0.077***	0.070***	0.102***	0.129***
	(0.009)	(0.010)	(0.010)	(0.012)
$Time^2$	-0.004***	-0.003**	-0.006***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.002)
$Time^3$	0.0001***	9.93×10^{-5}**	0.0002***	0.0003***
	(4.17×10^{-5})	(4.64×10^{-5})	(4.38×10^{-5})	(6.17×10^{-5})
Controls: Sex	Yes	Yes	Yes	Yes
Fixed-effects				
Strata FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Fit statistics				
Observations	2,910,406	2,894,785	2,987,337	2,988,397
Pseudo R^2	1	1	1	-153,079.1
R^2				0.41852

Clustered (at the country level) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The model estimated is a binomial logistic regression model for the 3 first columns, and an OLS for the last one. Strata fixed effects stands for the strata given in the DHS surveys and the lowest geographical level for the census surveys. Cluster at the country level stand for DHS cluster for the DHS surveys and the lowest geographical level for the census surveys. All dependant variables are dummies, except for years of schooling. In the regression, I do not add a dummy if the transition grade is below or above secondary education because it represents a bad control.

Table A.8: Descriptive global effects of bilingual education - Heterogeneity by grades of transition

Dependent Variable:	Literacy
Model:	(1)
Variables	
After the bilingual education reform	0.150
	(0.091)
Time (from 0 to 20)	-0.024
	(0.027)
$Time^2$	0.004
	(0.004)
$Time^3$	-0.0001
	(0.0001)
Controls: Sex	Yes
Fixed-effects	
Strata FE	Yes
Year FE	Yes
Fit statistics	
Observations	1,522,058
Pseudo R^2	1

Clustered (at the country level) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The model estimated is a binomial logistic regression model. Strata fixed effects stands for the strata given in the DHS surveys and the lowest geographical level for the census surveys. Cluster at the country level stand for DHS cluster for the DHS surveys and the lowest geographical level for the census surveys. The dependent variable is a dummy.

Table A.9: Robustness test for the no school sample

Dependent Variables:	Literacy	Tested literacy	School attendance	Years of schooling
Model:	(1)	(2)	(3)	(4)
	Logit	Logit	Logit	OLS
Variables				
After the bilingual education reform	0.167***	0.183***	0.154***	0.091
	(0.054)	(0.060)	(0.058)	(0.078)
Time (from 0 to 20)	0.069***	0.058***	0.092***	0.126***
	(0.010)	(0.010)	(0.010)	(0.013)
$Time^2$	-0.005***	-0.004***	-0.006***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.002)
$Time^3$	0.0002***	0.0001***	0.0002***	0.0003***
	(4.4×10^{-5})	(4.86×10^{-5})	(4.7×10^{-5})	(6.94×10^{-5})
Controls: Sex	Yes	Yes	Yes	Yes
Fixed-effects				
Strata FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Fit statistics				
Observations	2,323,856	2,312,084	2,384,338	2,385,652
Pseudo R^2	1	1	1	-152,883.6
R^2				0.41621

Clustered (at the country level) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The model estimated is a binomial logistic regression model for the 3 first columns, and an OLS for the last one. Strata fixed effects stands for the strata given in the DHS surveys and the lowest geographical level for the census surveys. Cluster at the country level stand for DHS cluster for the DHS surveys and the lowest geographical level for the census surveys. All dependant variables are dummies, except for years of schooling.

Table A.10: Descriptive global effects of bilingual education removing partially treated cohorts

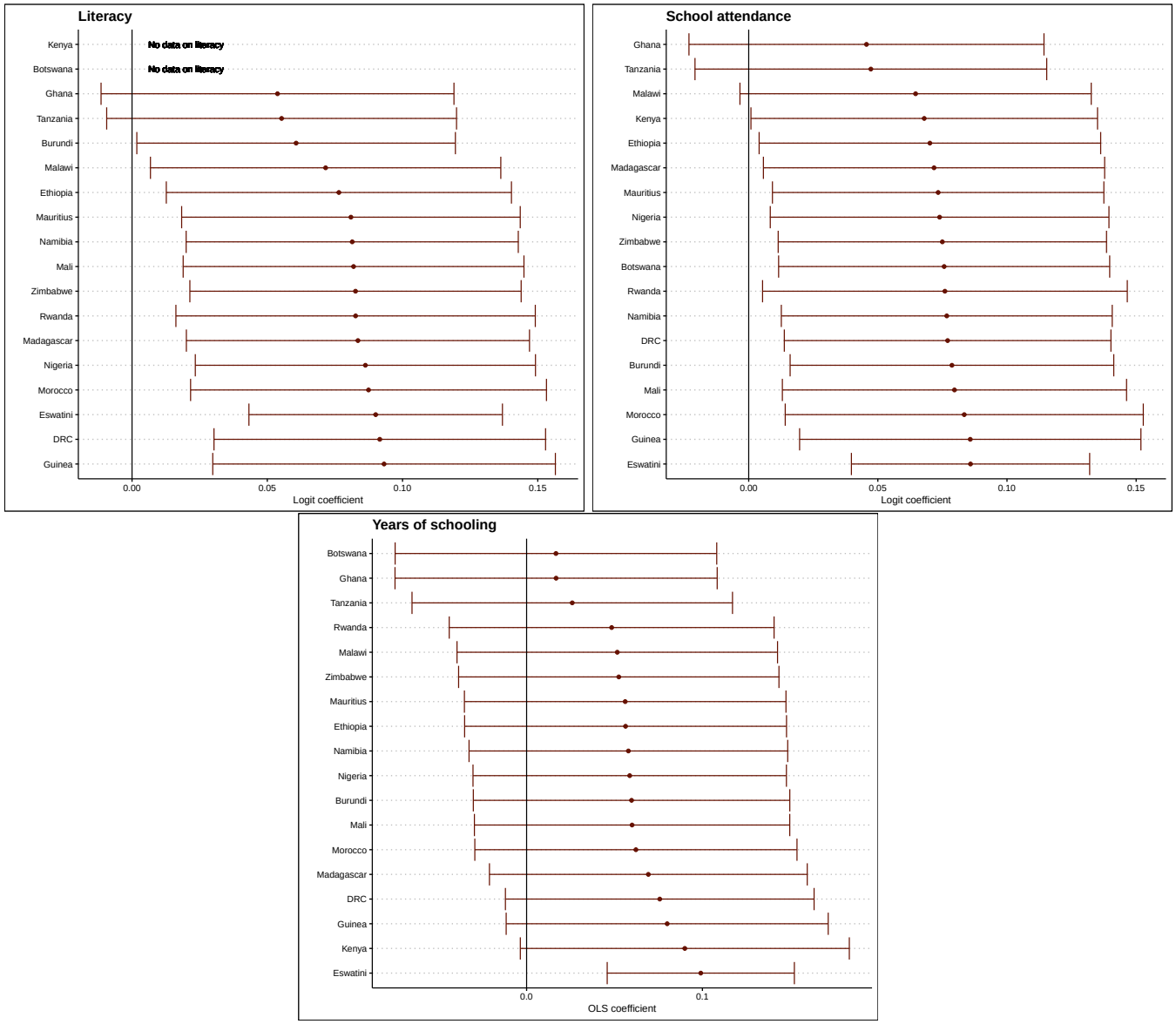

Figure A.3: Descriptive effect of bilingual education by dropping one country at a time

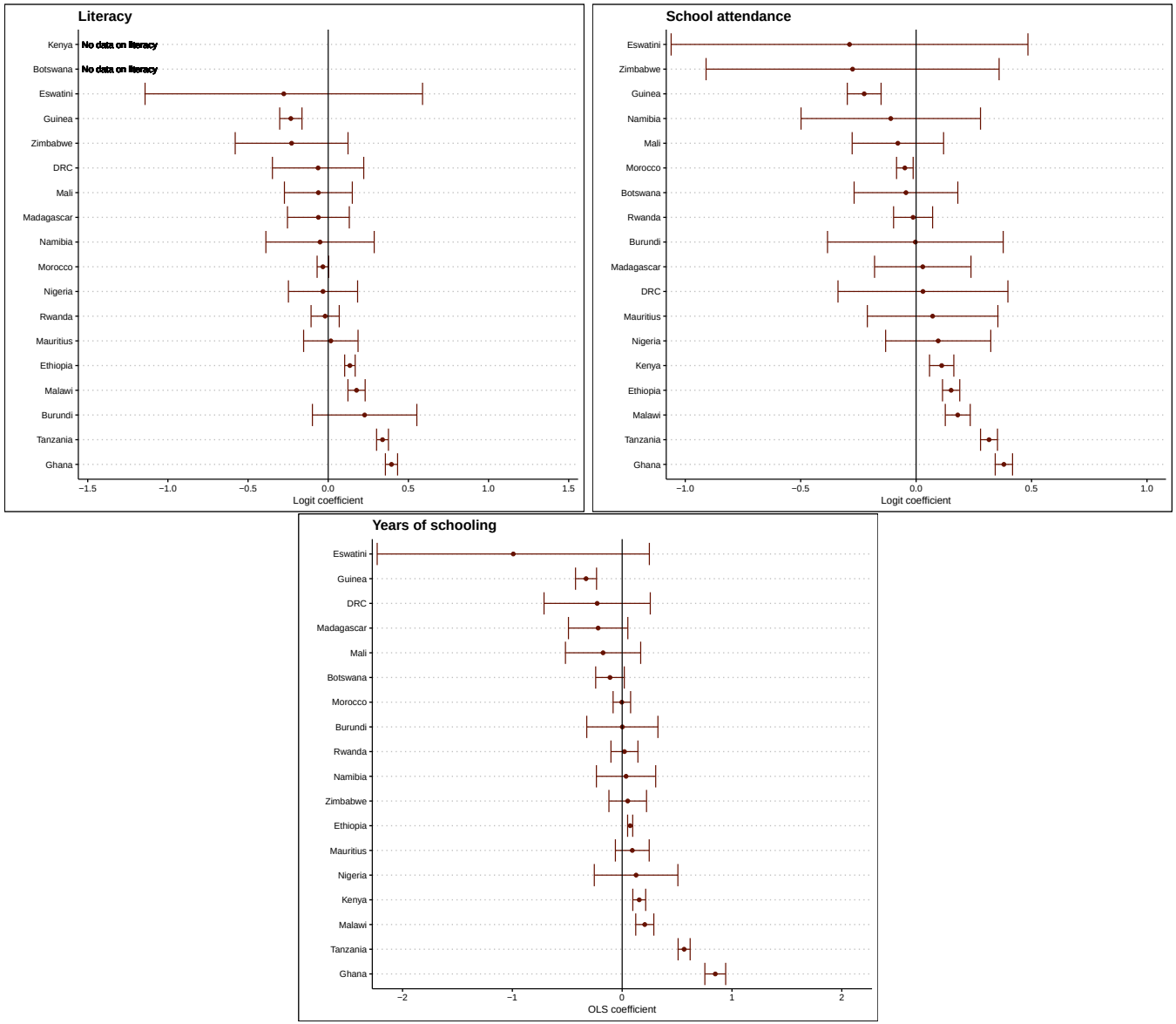

Figure A.4: Descriptive effect of bilingual education disaggregated at the country level

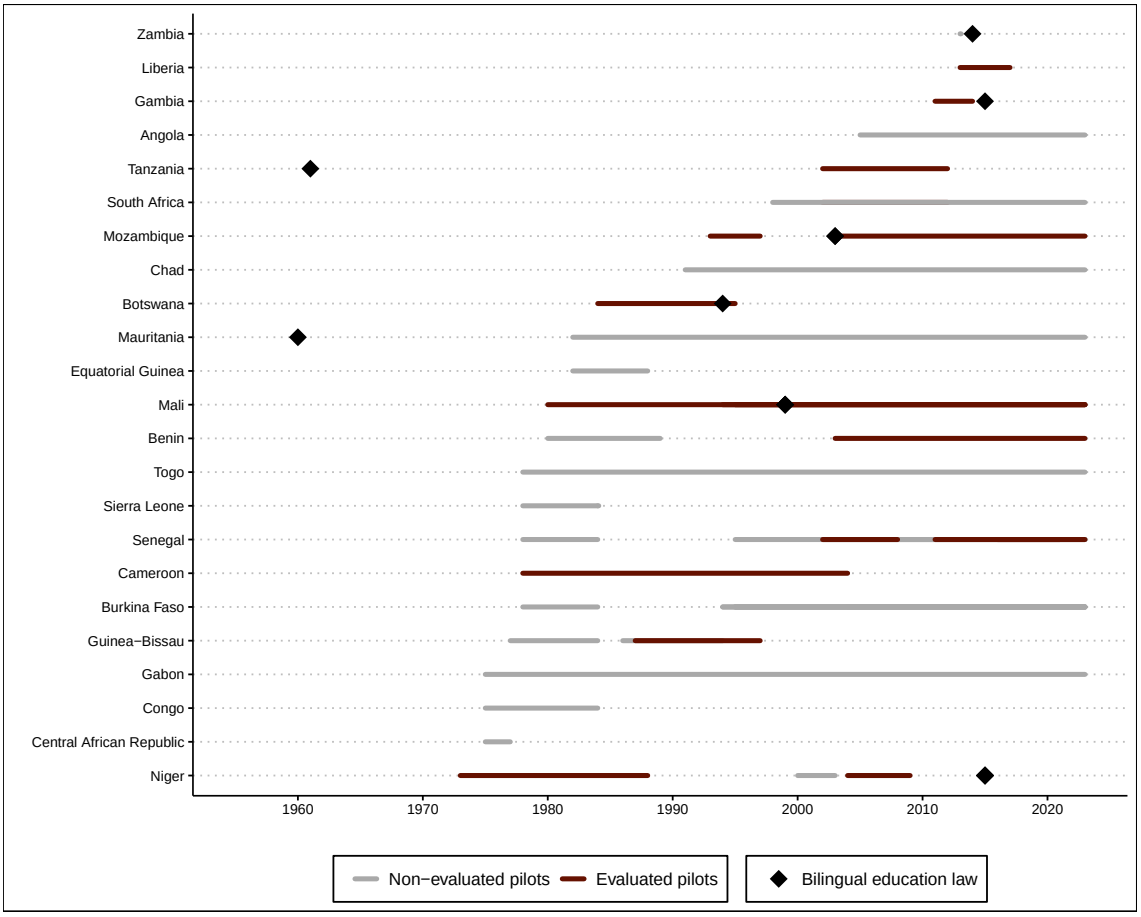

Figure A.5: Bilingual education pilots: time coverage and evaluation

Note: Author's computation. Important recall: time and geographical coverage of the dataset of pilots is not exhaustive.

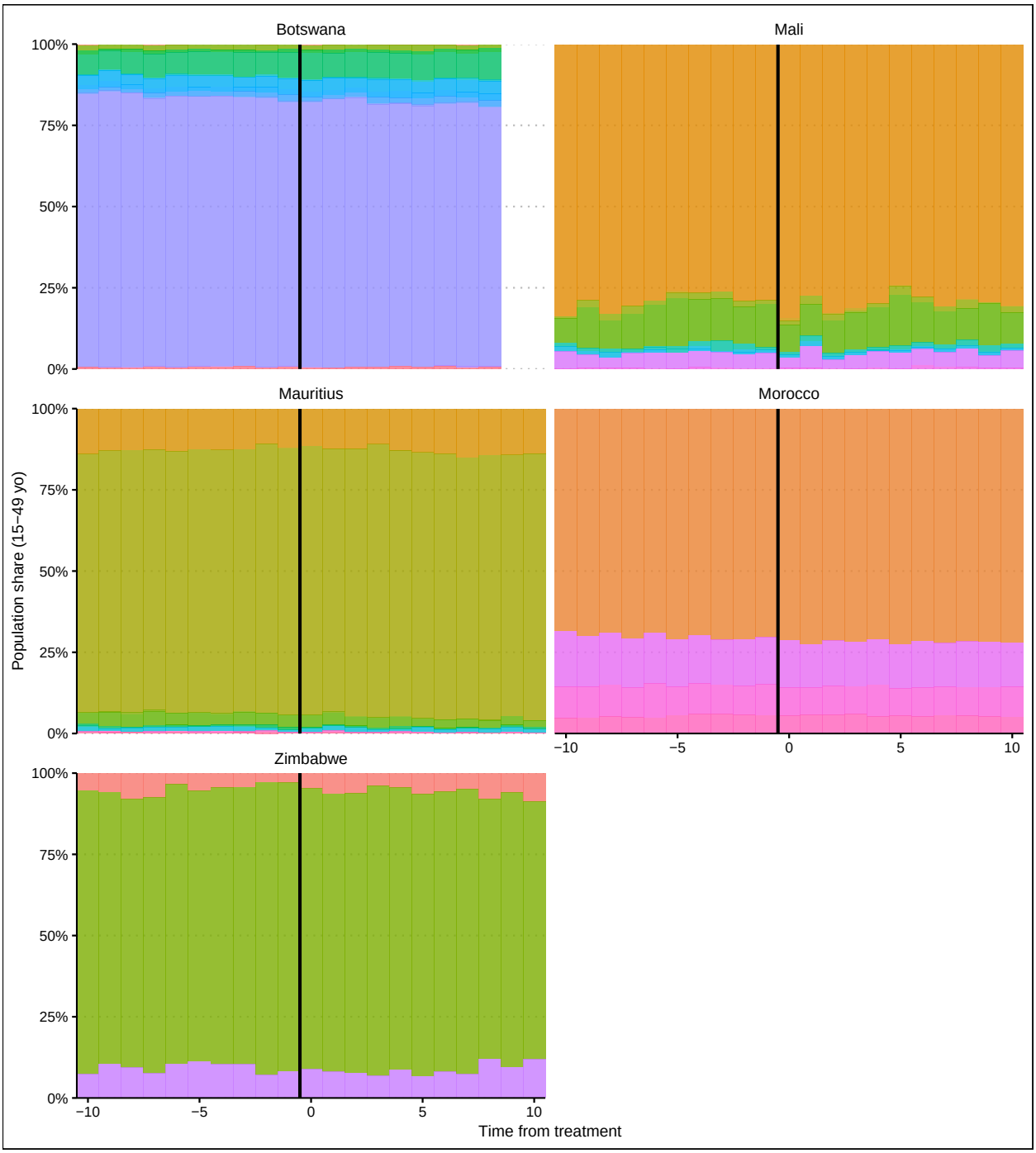

Figure A.6: Linguistic composition of Botswana, Mali, Mauritius, Morocco and Zimbabwe around the bilingual education reforms

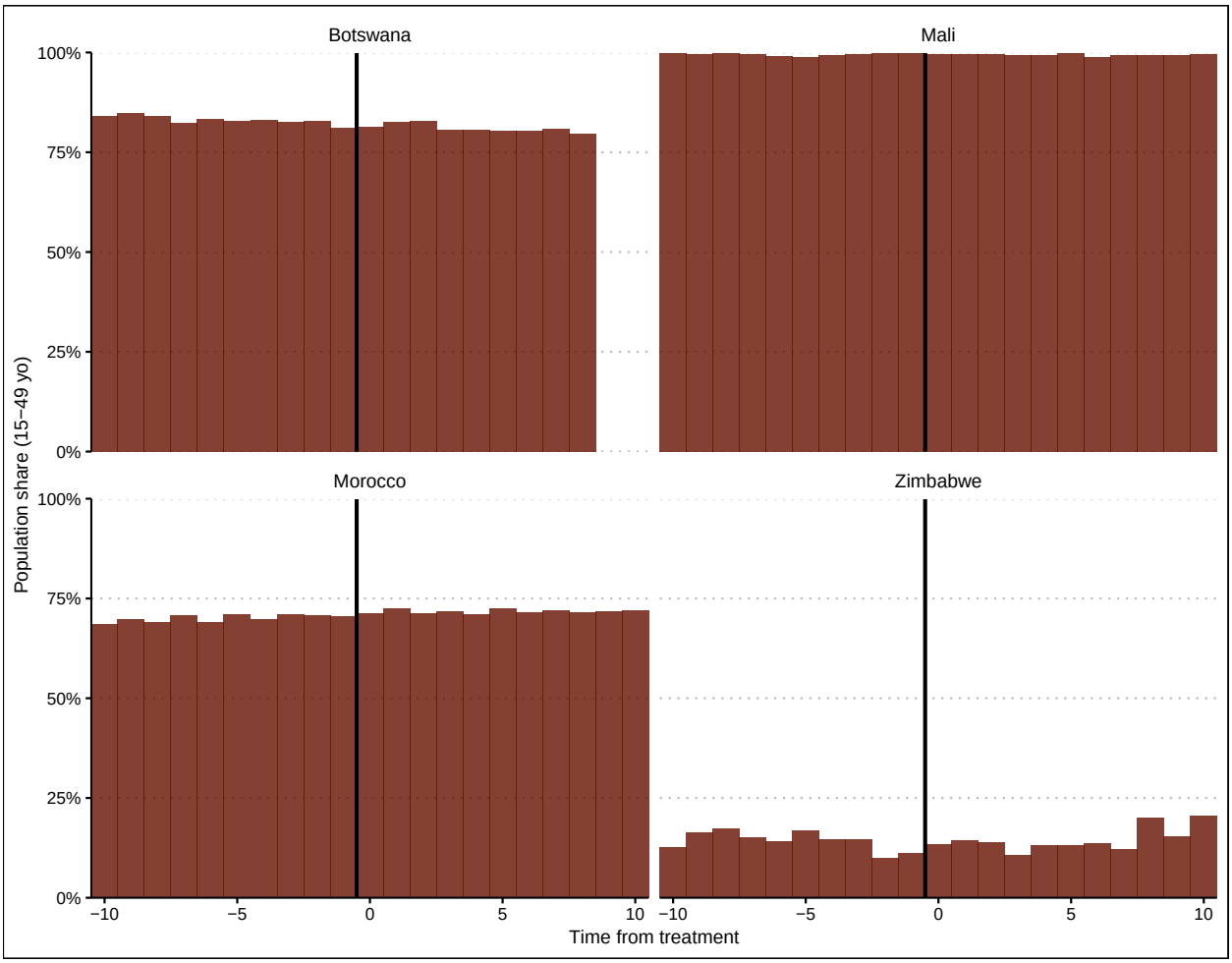

Figure A.7: Population share speaking the language(s) used in primary schools in Botswana, Mali, Morocco and Zimbabwe