

# Language use in education and primary schooling attainment: Evidence from a natural experiment in Ethiopia.

Rajesh Ramachandran\*

October 2012

## Abstract

In 1994 mother tongue instruction in primary schooling was introduced for the largest ethnic group in Ethiopia. Using the fact that the exposure to the language policy change was jointly determined by the date of birth, language group and the region of residence, we estimate the causal effect of the provision of mother tongue instruction on the largest ethnic group in the country. Our results suggest that provision of mother tongue education led to an increase of 0.75 to 1 year of primary schooling and 0.80 to 1.18 years of schooling in the affected cohort. Moreover the entire increase in the years of schooling can be attributed to the intensive margin of education. The language policy change increased the percentage of people completing 6 years or more of schooling by 31%. Applying our findings to a set of African countries shows, even after accounting for the costs of provision, that introduction of mother tongue instruction imply potentially large benefits and increases the percentage of population completing primary schooling by as much as 15% points. These finding have important policy implications at a time when surging enrolment rates and already stretched educational budgets in the African continent imply the need for solutions which can increase the quality of education without requiring huge capital or infrastructural outlays.

JEL: I24, I25, I28, O29, Z18.

Keywords: Education, Language Policy, Difference in Difference Estimator, Institutions.

---

\*International Doctorate in Economic Analysis, Universitat Autònoma de Barcelona, Barcelona-Spain. Contact email: rajesh.ramachandran@uab.cat. I would like to thank Alessandro Tarozzi, Ana Rute Cardoso, Francesco Fasani, Hannes Mueller, Jack Willis, Lidia Farrè, Tommaso Agasisti and the participants at the summer school in development economics, Italy, 2012 and the IIIrd workshop on education economics at the University of Barcelona for their invaluable suggestions, helpful discussions and comments. All errors are mine.

## I Introduction

Cognitive theories suggest teaching in a non-native language at the primary schooling level may have a negative effect on educational outcomes. In Sub-Saharan Africa, with the exception of Tanzania and Ethiopia, no country provides the entire span of primary schooling in a local language let alone the mother tongue. Africa is also characterised by some of the highest repetition and drop out rates in the world.<sup>1</sup> Whether the mismatch between the language of instruction and language used at home can partly help explain the poor educational outcomes observed in the African continent is a vital question with potentially important policy implications, which so far has not been wholly addressed by economists.

The existing literature (Angrist and Lavy 1999, Angrist, Chin and Godoy 2008), analysing effects of policy changes in the language used to teach children, has considered changes occurring only at the secondary schooling level. The language of instruction might have more important implications for cognitive development at the primary rather than the secondary schooling level (Cummins 1978a, 1979, 1981, Verhoeven 1994). Moreover these studies are not primarily concerned with the implications of the change in language policy on educational attainment. Angrist and Lavy (1999) concentrate on the effect of the language policy change on labour market outcomes. They find that replacing French with Arabic as the language of instruction, from the 6th grade onwards, led to a substantial reduction in the returns to schooling in Morocco. In our setting as the medium of instruction for secondary schooling and higher education continues to be English, we have reason to believe that such an effect on labour market outcomes might not be relevant for Ethiopia. Similarly Angrist, Chin and Goody (2008) are interested in the effect of the policy change on English speaking skills and not educational outcomes and moreover find that replacing English with Spanish in secondary schooling did not adversely affect English skills of the individuals affected. A second strand of literature, related to ours, exploits the presence of bilingual programs to estimate the effects of provision of mother tongue education on student achievement (Matsudaira 2005, Chin et.al 2011, Slavin et.al 2011). However, these studies differ from our work in two crucial aspects, the first being that they have been primarily conducted in the context of the United States, where the exposure to the majority language for the language minority students is much higher as compared to the situation in Africa. The difference in exposure to the language at the community level might be an important factor as to why results from bilingual studies based in the United States might not be applicable to the African context. The second reason being that the

outcome of most bilingual programs have been measured after a period of just 3 to 4 years whereas the benefits of bilingual schooling may appear often after as late as the 5th grade.<sup>2</sup> This paper aims to fill this gap by providing the first causal estimates of the impact of provision of mother tongue instruction on long run educational attainment in the context of Sub-Saharan Africa.

We use a change in the medium of instruction policy in primary schooling in Ethiopia to estimate the effects of the policy change on the years of schooling.<sup>3</sup> Since 1962 the language policy was characterised by the exclusive use of Amharic in primary schooling. Amharic is the language of the second largest ethnic group, the Amharas. In 1994, mother tongue instruction in primary schooling was also introduced for the Oromo people, the largest ethnic group in the country. English was and remains the language of instruction in secondary schooling and higher education.

In order to estimate the effect of the language policy change, we use the fact that an individual's exposure to the new language policy differed by the language group, date of birth and region of residence in the country. We use a difference in difference (D-I-D) estimator that controls for systematic variation in education both across language groups and cohorts. As Duflo (2001) notes, D-I-D estimators are especially well suited to evaluate policy changes when the reform comes from a variation in a well defined input; in our case the input being a change in the language used to teach children in primary schooling. The design of the empirical exercise allows us to show that the control and treated groups have similar trends in education attainment before the language policy change. We also show that among the earlier cohorts, who did not benefit from the language policy change as they had already finished primary schooling, the increase in educational attainment from one cohort to the other is not correlated to the language policy change in the country. By showing a clear discontinuity in the educational trend, in the first year in which the policy is implemented, we are also able to rule out other potential explanations such as omitted changes in schooling or regional conditions as otherwise we would expect such changes to also affect other individuals who were still in schools but unaffected by the policy change. Additionally by also comparing the same language group in the treated and untreated regions in the country, we address the concern that factors such as recognition of language rights, increase in group status or differential response of language groups might be confounding our results. The primary focus is on the effect of the language policy change on years of primary schooling and years of schooling. The main result shows that the provision of mother tongue instruction led to an increase of around 0.75 to 1 year of

primary schooling and 0.80 to 1.18 years of schooling in the affected cohort.

The estimates show that the change in language policy did not have any significant effect on the enrolment rates and hence the entire increase in the years of schooling is primarily due to the intensive margin of education.<sup>4</sup> The provision of mother tongue instruction, conditional on enrolment, increased the percentage of the sample completing 6 years or more of schooling by 31%. In fact the analysis shows that the provision of mother tongue instruction had a positive effect at all levels of schooling.

Applying our findings of the effect of provision of mother tongue instruction on primary schooling completion rates to a set of five African countries suggest potentially large benefits. Estimates suggest that the percentage of population, aged 15 to 49, which completes primary schooling could increase by around 10 percentage points from 29% to 39% in Benin, from 59% to 65% in Cameroon and from around 70% to 84% in Ghana. Combining conventional estimates on the rate of return to an additional year of schooling (Psacharopoulos 1994, 2004) with the cost of producing learning materials in local languages suggest that even when these costs are taken into account the potential gains from the introduction of mother tongue instruction remain large. Our findings have important policy implications at a time when increasing budget constraints coupled with surging enrolment rates in Africa imply the need for remedial tools which do not require large capital and infrastructural outlays but at the same time can increase the quality of education provided.

The rest of the paper is organized as follows. In the next section we describe the language policy change and education trends in Ethiopia. Section III outlines the identification strategy and presents the data. Section IV presents the main results and section V provides robustness checks. Section VI identifies the channel through which language policy works. Section VII discusses the implications of providing mother tongue instruction in other African countries and Section VIII concludes.

## II Language and education policy in Ethiopia

Ethiopia is situated in the horn of Africa and with a population of around 80 million it is the second most populous nation in the continent. The population is highly diverse containing more than 80 different ethnic groups. There are more than 90 different languages spoken in Ethiopia and most belong to the Afro-Asiatic language phylum of which three branches are represented, namely, Semitic, Cushtic and Omotic.

The “Oromo” are the largest ethnic group in the country comprising around 33% of the population. Though the most numerous in the country, Oromo can be considered as a minority in terms of political and economic influence they wield as a group both currently and over the course of Ethiopian history.<sup>5</sup> The “Amharas” comprising around 27% of the population are the second largest ethnic group in the country. The other major ethnic groups in the country are the Somali, Tigray, Sidama and Wolaita comprising 6.2%, 6%, 4% and 2.3% of the population, respectively. “Oromigna, the language spoken by the Oromo people belongs to the Cushitic branch and is written using the Latin alphabet. In contrast “Amharic (spoken by the Amhara people) belongs to the Semitic branch and is written using the script Amharic Fidel. The two languages split at the first branch of the Afro-Asiatic language phylum.

## II.1 Language policy

Ethiopia had been a monarchy for most of its modern history and was under the imperial rule of Haile Selassie between 1916 and 1974. The medium of instruction policy under the imperial rule involved the sole use of Amharic in primary schooling followed by the use of English as the medium of instruction for secondary schooling and higher education.<sup>6</sup>

The emperor was replaced by a Soviet-backed Marxist-Leninist military junta, the “Derg” (meaning council in Amharic), which came to power in 1974. The medium of instruction policy during the “Derg” regime, between 1974-1991, was identical to the one in place during the imperial time. Amharic continued to be the sole medium of instruction in primary schooling followed by the use of English from secondary schooling onwards.

Growing discontent against the “Derg” regime led to the establishment of the Tigray People’s Liberation Front (TPLF) in 1975, which merged with other ethnically based opposition parties such as the Oromo Liberation Front (OLF) and the South Ethiopian People’s Democratic Coalition (SEPDC) to form the Ethiopia’s People’s Revolutionary Democratic Front (EPRDF).<sup>7</sup>

The movement against the Junta government reached its peak in May 1991 when the ERPDF forces stormed Addis Ababa and the “Derg” regime was toppled. The vision of political and cultural autonomy for all ethnic groups, the banner under which the coalition forces had come together, meant that the transitional constitution of 1991 and the one voted in 1994 resulted in the creation of a federal republic. The country was

divided into nine regions and two city administrative units along ethnic and linguistic lines. The 9 regions are Afar, Tigray, Oromo, Amhara, Somalia, Benishangul-Gumuz, Southern Nations, Nationalities, and People's Region (SNNPR), Gambella and Harari. The two city administrative units are Addis Ababa and Dire-Dawa, respectively.

The vision of the rights to self-determination led to the introduction of mother tongue instruction in primary schooling for the four major ethnic groups Oromo, Tigray, Sidama and Wolaita starting 1994. Mother tongue instruction in primary schooling for the remaining smaller ethnic groups was to be slowly introduced in the course of the next years. The language policy post 1994 is still characterized by the exclusive use of English as the medium of instruction for secondary schooling and higher education. Pupils from the ethnic group other than the Amharas learn Amharic and English as a subject during the course of primary schooling, while those from the Amhara group only learn English as a subject during primary schooling.

[Insert Table 1]

Table 1 shows the implementation of language policy by the languages introduced as a medium of instruction in the nine regions and two city administrative units in the country. As can be seen in table 1 the Oromo people entering primary schooling after 1994 gained access to mother tongue instruction in the regions of Amhara, Dire-Dawa, Harari and Oromia.

## II.2 Education policy and trends

During the imperial time, formal schooling remained mostly an urban/semi-urban phenomenon with little or no schools in the rural areas. In the period after the imperial rule, between 1975 and 1989, enrolment increased by around 12%. However lack of investment in education meant that the schooling remained out of reach of most rural people.

Post 1994 the education sector was given renewed importance by the EPRDF. The budget for education increased steadily since 1996/97 and in real terms by around 50% in the five year period after 1995/96. The focus on expanding access to education resulted in almost doubling primary schooling enrolment from 4.5 million in 1996/97 to 8.1 million students in 2001/02. In 2001/02 the total education spending stood at US\$ 333 million and was 14% of total public expenditure. Two-thirds was spent on primary and secondary schooling, and the balance on technical and university education.

In per capita terms, however, there has been little increase in education spending. The total expenditure per student increased only by around 5% in the period between 1995/96 to 2001/02. Moreover when total expenditure is broken down by the components of recurrent and capital expenditure, the main picture that emerges in the words of the public expenditure review of Ethiopia by the World Bank in 2004 is the “insufficiency of spending at all levels” (pg. 15). The recurrent expenditure per student at the primary level has decreased by around 20% in real terms over the five year period of 1996/97 to 2001/02. The share of wages and salaries in the recurrent budget was around 97%. Despite this the pupil teacher ratios (PTR) have steadily deteriorated over the period 1995/96 to 2001/02. The PTR have increased from 32:1 and 33:1 in 1995/96 to around 73:1 and 80:1 in 2001/02 in primary and secondary schooling, respectively. Twenty nine percent of the enrolled population drop out by grade 1 and 55% of the enrolled population by grade 3, implying for these children little effective education is taking place.

### **III Identification strategy and data**

#### **III.1 Identification strategy**

As with many policy evaluations, the difficulty in estimating the effect of the provision of mother tongue schooling on education attainment is the lack of a counterfactual. To circumvent this problem the paper employs a D-I-D approach and uses the fact that the date of birth, the language group and the region of residence of an individual jointly determine the exposure to the language policy change.

The main exercise involves comparing the mean years of primary schooling/years of schooling of the Amhara and Oromo group for the older and younger cohorts. For our purpose we define the younger cohort as comprised of individuals who enter primary schooling after the language policy change. The older cohort is defined as comprising of individuals who have finished primary schooling by the time the language policy change was implemented. The older and younger Amhara cohorts had access to mother tongue instruction both before and after 1994, are not directly affected by the policy change, and form our control group of interest. The younger Oromo cohort however gained access to mother tongue instruction after 1994, and comprise our treated individuals. The difference in mean years of primary schooling between the two language groups, for the older cohort is compared with the difference of the younger cohort. This difference in difference can

be interpreted as the causal effect of the policy change on the Oromo people, under the assumption that, in the absence of the change in the medium of instruction policy, there would have been no change in the pattern of similar trends in educational attainment for the two groups.

As Duflo (2001) and Strauss and Thomas (1995) note when a fixed effects estimator is being used to assess the effect of a policy change, the investigator should pay close attention to the validity of the identifying assumption. The estimate may be capturing something due to the pre-existing differential trends for the two groups or the due to the divergence in trends later for reasons unrelated to the policy change. Also if the increase in education of the two groups was negatively correlated with the initial levels, then a treatment effect might be observed even if the program had no real effect.

We try and closely analyse the validity of our identifying assumption. The test of our identifying assumption exploits the presence of multiple groups formed by successive cohorts not exposed to the policy change (Duflo 2001, Heckman and Hotz 1989, Rosenbaum 1987). The individuals belonging to the Oromo group who entered primary schooling before 1994 were not affected by the change in the language policy and hence we should not expect the educational attainment to vary systematically across the Amharic and the Oromo group for the older cohorts. Figure 1 plots the trend in mean years of primary schooling for three older cohorts for the two groups, before the policy change in 1994. The fixed effects estimator allows for the levels across the two comparison groups to be different as long as the trend or the shape of the curve remains the same. We see that the two groups over the 30 year period, of 1964 to 1994, follow parallel trajectories and have very similar trends in primary schooling attainment. In the results section using a difference in difference estimator we formally show that the two groups have similar trends before the policy is implemented .

The policy change was implemented only for individuals who entered primary schooling after 1994, which meant individuals who had entered primary schooling before 1994 but had not yet finished primary schooling were unaffected by the policy change. However as we identify individuals stage of schooling by their age, delayed school entry or repetition would imply that some of the individuals older than the normal primary school entering age could have benefited from the policy change. Hence considering all these people as untreated would provide us with a lower bound of the true effect of the program. For this reason we only consider individuals who were of an age to have finished primary schooling when the language policy change was implemented, as the older cohort.



The data allows for other potential experimental designs to estimate the effect of provision of mother tongue education, though for reasons discussed below we believe the strategy of comparing the Amhara to the Oromo might be the best way to minimize potential estimation bias. The Oromo people gained access to mother tongue education in 4 of the 11 regions in the country. An alternative estimation strategy could involve comparing the older and younger cohorts of the Oromo group in the treated and untreated regions of the country. This would involve comparing the same language group in different regions whereas our main experiment compares different language groups but in the same regions. There exist disparities in access and schooling infrastructure across regions with the south-western and northern parts of the country being the least developed. In order to minimize the concerns that our results are driven by differential access to schooling infrastructure, we prefer our main experiment to the one comparing the Oromo group in the treated and untreated regions of the country. This said in Section V, we do carry out the exercise of comparing the Oromo's in the treated and untreated regions and show that the results obtained are very similar to the ones obtained with our preferred identification strategy.

As noted before, the language policy change involved the introduction of mother tongue instruction for the four major ethnic groups, namely, Oromo, Tigray, Sidama and Wolaita starting 1994. The Tigray people gained access to mother tongue instruction in only the Tigray region of the country. There is from a minimum of 4 to a maximum of 10 observations from the control group Amhara in the Tigray region, implying a difference in difference strategy comparing the Amhara and Tigray groups is not implementable.

The data only allows us to distinguish the population of Ethiopia into four distinct language groups. These are namely the Amhara, Oromo, Tigray and the "Others". The "Others" category includes all the remaining language groups clubbed together. As mother tongue instruction for the language groups besides the Oromo, Tigray, Sidama and Wolaita was introduced over the next years after 1994, most individuals who have finished schooling by 2011 from the language group besides the Sidama and Wolaita in the "Others" category are untreated in our data. As the data does not allow us to uniquely identify these individuals, comparing the "Others" category to the Amhara would provide a lower bound of the real effect of the language policy change. We however create a group called the Non-Amhara comprising of individuals from all other language groups besides the Amhara and carry out a difference in difference strategy comparing the Non-Amhara to the Amhara group, the results of which are presented in the appendix.

### III.2 Data

The data comes from the Demographic and Health Surveys (DHS), which are nationally representative data on health, education and demographic trends in developing countries. The data for Ethiopia are from the year 2011 and include information on a nationally representative sample from the 9 regions and two city administrative areas of Ethiopia. Figure 2 shows a map of Ethiopia, where the regions and city administrative borders are demarcated. The number of observations in our analysis range from a minimum of 3,210 to a maximum of 11,918 observations.

The schooling system in Ethiopia involves 8 years of primary schooling followed by 4 years of secondary schooling. The children in Ethiopia go to primary school normally between 7 and 14 years of age. A child born before 1980 and speaking Oromo as her mother tongue was 14 years old in 1994, had already finished primary schooling, and was unaffected by the change in the medium of instruction policy. The data being from the year 2011 implies that the youngest individual who could finish secondary schooling by 2011 was 2 years old in 1994. Similarly the oldest individual who could have been affected by the policy change was 7 years old in 1994. We hence consider the individuals aged 2 to 7 years in 1994 as the younger cohort and the Oromos aged 2 to 7 years in 1994 as the treated individuals.

The children speaking Oromo as their mother tongue and who were aged 8 to 14 in 1994 were already in primary school before the policy change was implemented. As the policy involved a change in the medium of instruction, the policy in general was implemented only for children who entered primary schooling from 1994 and thereafter. Grade repetition and delayed school entry could lead to some of the children aged 8 to 14 in 1994 to benefit from the program. Considering them as completely untreated would then provide us with a lower bound of the true effect of the program. Thus the cohort aged 13 to 20 in 1994, individuals who were completely unaffected by the policy change, are defined to be the older cohort for our purpose.<sup>8</sup> We however do show that considering the untreated individuals who were aged 8 to 12 years old in 1994, and were in primary schooling when the policy change happened, leaves our results unchanged. We use the information on age of the individual in 1994, along with his language background i.e. the language spoken as the mother tongue and their region of residence to match it with data

on medium of instruction policy implemented in various regions of the country (shown in table 1) to ascertain the impact of provision of mother tongue instruction on the Oromo language group.<sup>9</sup> The descriptive statistics for the younger and the older cohort are shown in table 2.

[Insert Table 2]

## IV Results

### IV.1 Comparison of means

The panel A labelled experiment of interest in table 3 presents the main experiment. There are a total of 5,364 observations with the treated comprising 1,327 observations. Comparing the two groups, for the older cohort aged 13-20 in 1994, shows that the average Amhara and Oromo individual had 2.89 and 1.82 years of primary schooling, respectively. The 1.07 years advantage, for the average Amhara, is due to the Amhara group having had access to mother tongue instruction before 1994 and having been the economic and politically dominant group in the country for the larger part of the 20th century. We observe that the mean years of primary schooling of the younger cohort, aged 2 to 7 in 1994, increases for both the groups due to the spurt in enrolment post 1994. The average years of primary schooling for an Amhara and Oromo individual of the younger cohort stands at 4.16 and 3.81 years, respectively. After gaining access to mother tongue instruction the difference between the average Amhara and Oromo reduces from 1.07 to 0.35 years of primary schooling. The D-I-D or reduction in gap of 0.72 years of primary schooling can be considered as the causal effect of the language policy change. In the next subsection we will check for the statistical significance of the D-I-D calculated.

[Insert Table 3]

The causal interpretation as noted before depends on the identifying assumption of similar trends in primary schooling for the two groups in the absence of the policy change. The panel B in table 3 formally presents our control experiment. We consider two cohorts, aged 13-20 and 21-28 in 1994, not exposed to the policy change from the two groups. The identifying assumption of similar trends should imply that the D-I-D should be equal to zero. The panel B indicates that the D-I-D is equal to -0.07 and very close to zero. In the next subsection we show that the difference in difference calculated is indeed statistically

insignificantly different from zero.

## IV.2 Basic Difference in difference regressions

In order to evaluate the difference in difference estimator we run the following reduced form regression:

$$S_{ijkn} = \delta_0 + \delta_1 * D_j * C_k + \delta_2 D_j + \delta_3 C_k + \delta_4 B_k + \delta_5 R_n + \epsilon_{ijk} \quad (1)$$

$S_{ijkn}$  refers to the years of primary schooling of individual  $i$ , from language group  $j$ , of cohort  $k$  and in region  $n$ .  $D_j$  is a dummy variable taking the value 1 if the individual belongs to the Oromo language group and zero otherwise.  $C_k$  is a dummy variable which takes the value 1 if the individual belongs to the cohort which was aged 2 to 7 in 1994 and zero otherwise.  $B_k$  is a vector of year of birth dummies for the individuals aged 2 to 7 and 13 to 20 in 1994 for each year of birth and  $R_n$  is a vector of region dummies.

The results of the main experiment are shown in panel A of table 4. Column (1) does not control for year of birth or region dummies. The language group dummy which captures the difference in level between the two groups is negative and significant at the 1% level. It captures the fact that the average Amhara has 1.07 more years of primary schooling as compared to the average Oromo. The cohort dummy captures the time trend of increasing years of primary schooling in the country and is positive and significant at the 1% level. The coefficient shows that on an average the younger cohort has 1.27 years more of primary schooling due to the increased enrolment post 1994. The main coefficient of interest is  $\delta_1$ , the one associated with the interaction term between the language and the cohort dummy, and it captures the effect of provision of mother tongue instruction. The calculated D-I-D of 0.72 years is statistically significant at the 1 % level. Column (2) additionally controls for year of birth and region dummies. Additionally controlling for these does not change the significance and in fact increases the value of the point estimate of  $\delta_1$  and shows that the provision of mother tongue instruction increased years of primary schooling by 0.75 years in the affected cohort. In column (3) the dependent variable considered is the mean years of schooling instead of mean years of primary schooling, and includes as controls the region and year of birth dummies. The coefficient capturing the effect of the provision of mother tongue instruction is positive and significant and shows that provision of mother tongue instruction increased average years of schooling by around

0.80 years.

[Insert Table 4]

The panel B of table 4 presents the results of our control experiment or the placebo test. Column (1) does not control for year of birth or region dummies and shows that the D-I-D of -0.07 years is indeed insignificantly difference from zero. Column (2) controls for the year of birth and region dummies. Additionally controlling for these reduces the size of the point estimate to -0.009 and it remains statistically insignificant. The results of panel B in table 4 provide evidence in support of the assumption that in the absence of the policy change there would have been no divergence in the trend for years of schooling for the two groups.

Comparing the size of the coefficients of the two interactions terms, in panel A and B of table 4, shows that the coefficient in the experiment of interest is about 100 times the size of the coefficient in panel B. Moreover the associated standard errors are nearly identical implying that the insignificant coefficients found in the control experiment are not due to the issue of lower precision in the estimation of the placebo test.

### IV.3 Estimating the effect of the policy change for every cohort

The analysis carried out in the previous section compares the cohort aged 13 to 20 in 1994, who are completely untreated, to the cohort aged 2 to 7 in 1994 and who are completely treated. In this subsection we extend our identification strategy to a generalized interaction term analysis to take into account the effect of the policy change for each cohort aged 2 to 21 in 1994.

The relationship between the education ( $S_{ijkn}$ ) of an individual  $i$ , from language group  $j$ , in year  $k$ , of region  $n$  and their exposure to the language policy change can be expressed as follows:

$$S_{ijkn} = \delta_0 + \sum_{l=2}^{l=21} (D_j * d_{il})\delta_{1l} + \delta_2 D_j + \delta_3 B_k + \delta_4 R_n + \epsilon_{ijk}, \quad (2)$$

where  $d_{il}$  is a dummy that indicates whether individual  $i$  is of age  $l$  in 1994. The other variables have the same interpretation as in equation (1). The omitted dummy category is the individual aged 21 years old in 1994. Each coefficient  $\delta_{1l}$  can be interpreted as the effect of the language policy change on a given cohort of the Oromo language group. As

all children who had finished or entered primary school before 1994 did not benefit from the introduction of mother tongue instruction imply that  $\delta_{1l}$  should be equal to 0 for  $l > 7$  and positive for  $l < 8$ .<sup>10</sup>

In panel A of figure 3 we plot the estimated coefficients,  $\hat{\delta}_{1l}$ . Each dot on the solid line corresponds to the coefficient of interaction between the dummy for whether individual  $i$  is of age  $l$  in 1994 with the Oromo language dummy (the 95% confidence interval is plotted in broken lines). Each dot thus summarizes the effect of the language policy change on the cohort aged  $l$  in 1994 and belonging to the Oromo language group in the treated regions. These reduced form estimates for each year of birth allows us to verify whether  $\hat{\delta}_{1l}$  follows the pattern implied by the assumption underlying the identification strategy. As can be seen, these coefficients fluctuate around zero and are statistically insignificantly different from zero for all ages between 20 and 8 and start increasing for ages below 8. Panel A in figure 3 shows a discontinuity for the coefficient,  $\hat{\delta}_{1l}$ , of the cohort aged 7 in 1994. It is seen that the first cohort to benefit from the language policy change has a break from the prevailing trend and is the first coefficient which is statistically different from zero. The above shows that the policy change did not have any effect on education of cohorts not exposed to it and had a positive effect on the education of all younger cohorts.<sup>11</sup> The pattern observed helps us address the concerns that our estimate may be capturing something due to the omitted changes in schooling and regional conditions. If other omitted changes in schooling conditions or other policy variables were driving our estimate, we would expect it to affect also individuals who were already in primary school or also maybe in secondary schooling. The discontinuity demonstrated in the graph seems to suggest that the effect being captured by the interaction term is the effect of the language policy change.

In the next section to address the concern that it is not the differential response of a particular language group to the general emphasis given to education sector post 1994 or the recognition of language rights or increase in group status that is driving our results, we compare the same language group in the treated and untreated regions of the country. We also additionally explore other potential explanations that could be put forth to explain our findings to argue that the effect being found is indeed the result of the language policy change.

## V Robustness tests

### V.1 Comparing the Oromo's in the treated and untreated regions

Post 1994 the Oromo people gained access to mother tongue instruction in 4 of the 11 regions in the country. As a first robustness test we compare the Oromo's in the treated and untreated regions of the country. This exercise involves comparing the same language group in different regions whereas the main experiment involved comparing different language groups but in the same regions. There are a total of 3,210 observations with the treated comprising 1,327 observations.

Table 5 compares the mean years of primary schooling of the Oromos for the older and the younger cohorts in the treated and untreated regions of the country. The older cohort in the untreated region have on an average 1.22 more years of primary schooling as compared to the older cohort in the treated regions. This is due to the untreated regions including the capital Addis Ababa; the most developed part of the country. After gaining access to mother tongue education the gap between the average individual in the untreated and treated regions reduces to 0.47 years of primary schooling. The D-I-D suggests that the provision of mother tongue instruction increased average years of primary schooling by 0.75 years in the affected cohort, very similar to the increase of 0.72 years found in table 3. Looking at the Oromo untreated individuals, we see that the younger cohort gains around 1.24 years of primary education. The increase for the non-treated individuals from the Oromo group is very similar to the increase of 1.27 years observed for the untreated younger Amhara cohort in table 3, although we are looking at two different groups in different regions of the country.

[Insert Table 5]

Table 6 shows the results of formally evaluating the D-I-D estimator. Column (1) does not control for year of birth or region dummies and shows that the difference in difference of 0.75 years found through comparison of means is statistically significant at the 1% level. Column (2) additionally controls for year of birth and region dummies. Additionally controlling for these increases the size of the point estimate to 1.00 and it remains statistically significant. In column (3) the dependent variable considered is the mean years of schooling instead of mean years of primary schooling, and includes as controls the region and year of birth dummies. The estimate shows that the provision of mother tongue instruction increased average years of schooling by around 1.18 years.

[Insert Table 6]

We now as in the Section IV.3 estimate the effect of the policy change for each cohort aged 2 to 21 in 1994. The results are presented in panel B of figure 3. Each dot on the solid line in panel B of Figure 3 plots the coefficient of the interaction between a dummy for being a given age in 1994 and the regional dummy, which takes the value one for the regions in which the Oromo's were treated. Each dot tells us the effect of the language policy change for a particular cohort for the Oromo people in the treated regions of the country. We again observe a very similar pattern. All the coefficients are very close to zero up until the cohort aged 9 in 1994.<sup>12</sup> The graph has a break in trend for the cohort aged 7 in 1994, the first cohort to be exposed to the language policy change in the country.

The observed break in trend again helps address the concern that it is not the effect of other changes in schooling or regional conditions which is confounding our estimate. Moreover comparing the same language group also helps us rule out the concern that it different groups reacting differently to the general expansionary trend in education that was taking place in Ethiopia or it is the effect of political changes such as recognition of language rights or increase in group status that is being captured by our estimate.

In the appendix in table A and B are shown the results of comparing the Non-Amhara to the Amharas. These shows that the results are robust to considering other language groups and regions of the country.

## V.2 Other potential confounding factors

A potential explanation driving the results could be that it is not the change in the language used to instruct children but the change in the composition of teachers. The existing literature has however found that race, gender and ethnic composition of teachers are likely to matter more for subjective evaluation rather than objective performance of students (Ehrenberg, Goldhaber and Brewer 1995). In our context, due to lack of data, we only provide some suggestive evidence as to why change in the composition of teachers is not likely to affect our result. As we noted in section II, the primary school enrolment rose from around 4.5 million to around 8.1 million and at the same time the pupil-teacher ratio (PTR) increased from around 32:1 to around 73:1. This seems to suggest that as enrolment doubled so did PTR implying that no new teachers were hired. It is interesting to note in this regard that teachers in public schools in Ethiopia are government employees, making hiring and firing decisions quite rigid. This potentially seems to suggest that there were



no major changes in the composition of public school teachers.

As we noted before the enrolment rates in the country increased sharply post 1994. Another potential confounding factor could be the differential access to schooling infrastructure. In this regard it should be noted that in the main experiment we are looking at the treated and untreated individuals in the same regions so the question of differential access to infrastructure is partly addressed by this approach. Moreover our results are robust to restricting our sample to any combination of the four treated regions. Moreover in the next section we show that there are no systematic differences in enrolment rates across the control and the treated group, after the policy change, and the increase in schooling primarily comes about due to the higher completion and lower dropout rates.

The identification strategy and the accompanying robustness tests suggests that our identifying assumptions and the causal interpretation provided are reasonable and plausible.

## **VI Identifying the Channel of Increase in Educational Attainment**

The estimates from the previous exercise shows that the provision of mother tongue instruction led to an increase of about 0.75 to 1 year of primary schooling in the affected population. The question that we try to address in this section is how much of the increase is due to the extensive margin and the intensive margin of education, respectively. For our purpose, we define the change in years of schooling associated with increased enrolment due to the change in the institution of language policy as the extensive margin. The intensive margin is the change in schooling associated with the people who would enrol irrespective of the choice of medium of instruction but choose different levels of schooling under the alternative scenarios.

The data allows us to identify whether the individual was ever enrolled in schooling or not. Using the same design as the main experiment in table 3 and 4, we continue to compare the Amhara and the Oromos in the regions where the Oromos gained access to mother tongue instruction. We create a dummy variable enrolment equal to 1 in case the individual was ever enrolled in schooling and zero otherwise. Table 7 compares the two groups for the cohorts aged 2 to 7 and 13 to 20 in 1994, where the dependent variable is the proportion of individuals from each group ever enrolled in schooling.

[Insert Table 7]

Comparing the cohorts aged 13 to 20 in 1994, we see that the proportion of individuals ever enrolled are 43% and 48% for the Amharas and the Oromos, respectively. The level of enrolment, as discussed in section II, increases sharply for the younger cohort and stands at 70.9% and 71.1% for the Amharas and the Oromos, respectively. Calculating the D-I-D suggests that language policy resulted in increasing enrolment by around 5% points. We formally estimate the D-I-D estimator by using a Probit model to implement equation (1), where however now the dependent variable is the dummy variable enrolment. The results are shown in column (1) of table 8.

The interaction term capturing the increased probability of enrolment, due to provision of mother tongue instruction, is very close to zero and insignificant. The language policy seems to have had no significant effect on the probability of enrolment. The benefit of mother tongue instruction on cognitive development have not been well understood in most policy circles and by stakeholders such as parents, as can be seen with the widespread preference and continuing practice of using former colonial languages as a medium of instruction in most African countries. Moreover the knowledge of the former colonial language or the dominant language of the country are often seen as a prestigious mark of education and modernity in many countries in Africa, and is cultivated by many as a means to acquire status.<sup>13</sup> Given the above two factors it is not surprising that the provision of mother tongue instruction did not have any significant effect on the probability of enrolment of the Oromos.

The above exercise suggests that the entire increase in years of primary schooling can be attributed to the intensive margin. The literature on language, bilingualism and child development has highlighted the role of mother tongue instruction on cognitive development of children (Cummins 1978b, 1979, 1981, 1984). Although measuring cognitive skills has been a tricky issue, the literature has been able to demonstrate that various measures of cognitive skills are positively correlated with schooling, wages and labour market outcomes (Heckman et.al 2006). The availability of mother tongue instruction resulted in assisting the cognitive development of children, which reduced the cost of obtaining education and hence increased the completion rates. The fact that the language policy works through the intensive and not the extensive margin can be reconciled by applying a theoretical framework of sequential schooling choice made under uncertainty (Altonji 1993, Zamarro 2004). The pupils have probability distributions defined over the cost of effort.

Introduction of mother tongue instruction does not alter the probability distribution of the cost of effort, as it is a new technology over which individuals hold no beliefs. Hence the same set of individuals enrol under the two language policy settings. However once enrolled they find that the cost of effort while studying in the mother tongue reduces, leading to updating of probability distribution functions, which leads to pupils choosing more years of schooling. We should hence expect that in the future as people become cognizant of the benefits of mother tongue instruction, the policy would have an effect on the extensive margin also.

[Insert Table 8]

The cumulative distribution function (CDF) showing the proportion of total students dropping out at each grade conditional on enrolment, for grades 1 to 8, is shown in figure 5. Panel A shows the CDF for the two cohorts, aged 13 to 20 and 2 to 7 in 1994, from the Oromo group. We see that that the CDF of the younger cohort stochastically dominates the one of the older cohort. Comparing the younger with the older Oromo cohort shows that conditional on enrolment 58% of the population now finishes 6 years or more of schooling as compared to only 37% before. The panel B depicts the CDF for the two cohorts, aged 13 to 20 and 2 to 7 in 1994, from the Amhara group. We see that up until grade 5 the two curves almost overlap and after grade 5 in fact the CDF for the older cohort dominates the one of the younger cohort. This decrease in completion, conditional on enrolment, for the Amhara group can be attributed to the reduction in per capita recurrent expenditure per student. The panel C in figure 5 depicts the D-I-D of the CDFs. The dot on the 5th year of education, for instance, indicates that provision of mother tongue instruction induced 20% of the sample to complete 6 years of schooling or more as compared to 5 years or less. The curve indicates that the policy change had a positive effect at all levels of primary schooling.<sup>14</sup>

In order to formally estimate the effect of language policy on completion rates, we construct a dummy called “completion”. The variable takes value 1 if the individual completed 6 years or more of schooling and zero otherwise.<sup>15</sup> We estimate the regression given by equation (1), where now the dependent variable is the dummy completion. This regression is estimated both for the entire sample and only for the individuals who were ever enrolled in schooling. The results of the restricted and the entire sample are shown in column (b) and (c) of table 8, respectively. The coefficient on the interaction term capturing the effect of provision of mother tongue instruction on probability of completing

6 years or more of schooling as compared to 5 years or less is positive and significant. The estimate suggest that the change in language policy, conditional on enrolment, resulted in inducing 31% of the sample to complete 6 years of schooling or more as compared to 5 years or less whereas the estimate for the entire sample is 24%.

The increase in completion rates is similar to the one found by Jackson (2000). He finds that the use of French in the first two years of primary schooling instead of Kirundi in Burundi led to an increase in the dropout rate from around 28% to 40%. Similarly Patrianos and Psacharopoulos (1995) based on a household survey in 1990 in Paraguay indicate that language strongly influences school attainment and performance. They find that language was the single best predictor of repetition and the cost of being a Guarani-only speaker is about one year of schooling attainment.

The World Bank (2004) study notes that one of the big problems facing Ethiopia is the large number of students dropping out before finishing grade 3. Dropping out at such early stages imply that for these children little or no effective education is taking place. The problem of high dropout remains prevalent throughout the African continent implying many resources are spent on educating people who never achieve effective education. The use of foreign languages as a medium of instruction might be an important factor driving such high dropout rates observed on the African continent.

The finding that the use of mother tongue as a medium of instruction primarily works through the intensive margin also has other important policy implications. In Ethiopia and other African countries the surge in enrolment rates has meant that per capita expenditure on students, especially recurrent education expenditure, has been declining. In such a context means of improving quality of education which do not require massive infrastructural or capital investments are crucial policy tools to ensure effective education for all. As we noted in section II in the years between 1995/96 and 2001/02, the recurrent education expenditure per student in Ethiopia declined by around 20% in real terms. Comparing the younger and the older Oromo cohort we see that the average years of education obtained, conditional on enrolment, increased from 4.27 years to 5.37 years. In the case of the Amharas the average years of education, conditional on enrolment, actually decreased marginally from 5.97 to 5.86 years. This small reduction as noted before can be attributed to the reduction in per capita recurrent expenditure per student. The fact the the Oromo people increased their educational attainment despite per capita recurrent expenditure going down highlights the fact that language policy might be an important policy tool to increase quality and years of schooling in countries with scarce resources.

## VII Applications to the African Continent

In this section we intend to explore how provision of mother tongue instruction could affect the proportion of population completing primary schooling in the African context and the net benefits arising from an additional year of schooling. It should be noted at the outset that the task associated with extrapolating results found in one specific context to other countries comes with its many associated pitfalls and problems. We do not seek to claim that our estimates do not suffer from these problems but look at the exercise as a way to shed some light and provide some benchmark estimates about potential benefits of mother tongue education. First, using data from the DHS, for other African countries, we calculate how the provision of mother tongue instruction to all ethnic groups which comprise 10% of the population or more would change the percentage of population completing primary schooling.

In the previous section we saw that the provision of mother tongue instruction induced 31% of the enrolled sample to complete 6 years or more of education as compared to 5 years or less. This is the key figure from the previous results that we will employ to generate the required counterfactuals. The procedure utilized to calculate the change in the proportion of population completing primary schooling is explained in the appendix. The same procedure is applied to a set of five African countries and the results are shown in table 9.

[Insert Table 9]

The results suggest that in Benin, where there exist three language groups with population shares greater than 10%, provision of mother tongue instruction to these groups could increase the percentage of population completing primary schooling for the people aged 15 to 49 from around 29% to 38.5%, an increase of nearly 10% points.

In the case of Burkina Faso where there is only language group with a population share of greater than 10%, the “Mossi”, who comprise 56% of the population. The estimate suggests that the provision of mother tongue instruction could increase the percentage of population aged 15 to 49 completing primary schooling from 16% to 23% points for the Mossi and from 15% to 20% for the country as a whole. In the case of Ghana, provision of mother tongue instruction increases the percentage of population aged 15 to 49 completing primary schooling from 70% to 84% and finally in the case of Gabon and Cameroon the increase is from 66% to 80% and 59% to 65%, respectively. This application at the face of it

suggests potentially large benefits where the percentage of population completing primary schooling increases by as much as 15% points. In what follows we calculate the net present value arising from the gain of one additional year of primary schooling due to the provision of mother tongue instruction.

In order to calculate the net present value of an additional year of schooling, the associated costs of introducing mother tongue instruction have to be taken into account. The task of estimating the costs of producing learning materials through standardised methodology is highly problematic because this cost depends on a variety of factors such as the state of development of languages to be used, population sizes, attitudes towards use of local languages etc. In what follows we present some estimates of this cost based on existing studies. The main objective of the exercise is to show that even when these costs are taken into account the potential gains from the introduction of mother tongue instruction remain large.

Patrinos and Vadwa (1995) analyse the production costs of introducing local language material in the context of Guatemala and Senegal. The estimates for Guatemala are based on 500,000 textbooks developed by Direccion General de Educacion Bilingue Intercultural (DIGBI) for the four majority Mayan languages. The authors estimate that the introduction of Mayan curriculum increased the unit cost of primary education by 9 percent over the cost of Spanish-only curriculum. This however overestimates costs for the future years as this includes the curriculum development costs, accounting for 37% of the total cost, which would not have to be borne in the later years. In the case of Senegal the estimates suggest that whereas the cost of producing a French textbook is US\$ 0.35, this increases to US\$ 0.84 in the case of textbooks in Wolof. An important point to be noted is the estimates for cost per textbook for French is based on producing around 150,000 books whereas for Wolof the number of books produced were only 4,140. The authors point out that the per unit cost would decrease significantly as the number of books produced increase, as the associated fixed cost per unit would decrease. They estimate that economies of scale in production can be achieved by printing around 10,000 books and in such a scenario there would be no difference in the cost of a French or a Wolof textbook. Using the above estimates we assume that in the first year there is an increase of 10% in per capita spending per pupil and from the year onwards there is no difference in the cost of provision of local or foreign language instruction.

In order to calculate the return to education, we use the latest estimates of the coefficient on years of schooling in the Mincerian wage equation from the work of

Psacharopoulos and Patrinos (2004).<sup>16</sup> The figures on gross domestic product (GDP) per capita and expenditure per student as percentage of GDP per capita are taken from the World Bank indicators for the latest available year. The GDP per capita are measured in constant 2000 US\$.

The net present value of one additional year of education, assuming an individual works for 30 years, is given by:

$$NPV_{ij} = \sum_{t=1}^{30} \frac{(m_j(GDP_{PCj}))}{(1+r)^t} - (0.10)(p_jGDP_{PCj}) - (p_jGDP_{PCj}) \quad (3)$$

where  $NPV_{ij}$  refers to the net present value from an additional year of education for individual  $i$  in country  $j$ .  $m_j$  refers to the coefficient on years of schooling from the Mincerian wage equation for country  $j$ ,  $GDP_{PCj}$  refers to the GDP per capita in country  $j$  and  $p_j$  is the percentage of GDP per capita spent per student in country  $j$ .  $r$  refers to the discount rate and is assumed to be equal to 10%. Here note that we assume that the entire increase in cost and the cost of an additional year of schooling are borne by the individual herself in the current period.

As an illustrative exercise we calculate the net present value for an individual from Burkina Faso, Ethiopia and Ghana, respectively. The results are shown in table 10.

[Insert Table 10]

The calculation suggests that the net present value of such an investment is equal to about 60% of the yearly per capita income in these countries. These moreover assume that the GDP per capita remains constant over the 30 year horizon and does not take into account any endogenous effects of increase in human capital on the growth rate or other externalities arising from an educated workforce. For instance Appleton (2000) estimates that a 1 year rise in the average primary schooling of neighbouring farmers is associated with a 4.3% rise in output compared with a 2.8% effect of own farmer primary education in Uganda. The above evidence seems to suggest that African educationalists and policy makers should reassess current language use in education policies and introduction of mother tongue instruction might involve sizeable gains and benefits for the majority of the population.

## VIII Conclusion

The paper studies the role of the institution of language use in education on educational attainment. We analyse how provision of mother tongue instruction in Ethiopia to the ethnic group Oromo affected their years of schooling. Our estimates suggest that the policy change had a sizeable positive impact and increased mean years of primary schooling by around 0.75 to 1 year and mean years of schooling by 0.80 to 1.18 years in the affected cohorts. The analysis shows that the language policy works primarily through the intensive margin of schooling. The estimate suggest that the change in language policy, conditional on enrolment, resulted in inducing 31% of the sample to complete 6 years or more of education as compared to 5 years or less. The finding that language policy works through the intensive and not the extensive margin can be rationalized by a theoretical framework of sequential schooling choice under uncertainty.

The importance of education to growth and development of nation states imply that the African continent, which is characterised by the extensive use of the former colonial language in primary schooling, could have potentially large benefits from rethinking its language use in education policy. Applying our findings to a set of African countries show that provision of mother tongue instruction could increase the percentage of population completing primary schooling by as much as 15% points. Even accounting for the costs of provision show there are still sizeable benefits from the introduction of mother tongue education. The implications of language choices in society extend beyond its effect on educational attainment. The institution of language policy have important implications on health, political participation and division of power in society. Today most post colonial countries like India, Cameroon, Ghana, South Africa, to name a few, are marked by socioeconomic inequality along linguistic lines. The wider socioeconomic impacts of the institution of language use in education in particular and language choices in society in general remain relatively unexplored and an important area for future research.

## References

- [1] Altonji, J. (1993), "The demand for and return to education when education outcomes are uncertain," *Journal of Labor Economics*, 11(1), 48-83.
- [2] Angrist, J and Chin, A and Godoy, R. (2008), "Is Spanish-only Schooling Responsible for the Puerto Rican Language Gap?" *Journal of Development Economics*, 85(1-2):



105-128.

- [3] Angrist, J. and Lavy, V. (1997), “The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco,” *Journal of Labor Economics*, 15 S48-S76.
- [4] Appleton, S. (2000), “Education and health at the household level in sub-Saharan Africa,” Working Paper No. 33, Harvard University Center for International Development.
- [5] Bamgbose, A. (2005), “Mother-tongue Education, Lessons from the Yoruba Experience,” in: Brock-Utne, B. & Hopson, R.K (2005), *Languages of Instruction for African Emancipation: Focus on Postcolonial Contexts and Considerations*, Cape Town: CASAS.
- [6] Bertrand, M. and Duflo, E. and Mullainathan, S. (2004), “How Much Should We Trust Differences-in-Differences Estimates?” *The Quarterly Journal of Economics*, 119(1), 249-275.
- [7] Card, D (1999), “The Causal Effect of Education on Earnings,” in Orley Ashenfelter and David Card, eds., *Handbook of labor economics*, North Holland, 1802-63.
- [8] Card, D. and Krueger, A. (1992), “Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States,” *Journal of Political Economy*, 100(1), 140.
- [9] Chin, A. and Daysal, N. and Imberman, S. (2012), “Impact of Bilingual Education Programs on Limited English Proficient Students and Their Peers: Regression Discontinuity Evidence from Texas,” NBER Working Paper No. 18197.
- [10] Church, J. and King, I. (1993), “Bilingualism and Network Externalities,” *Canadian J. of Econ*, 26,2, 337- 45.
- [11] Cummins, J. (1978a), “The cognitive development of children in immersion programs,” *The Canadian Modern Language Review*, 34, 855-983.
- [12] Cummins, J. (1978b), “Bilingualism and the development of metalinguistic awareness,” *Journal of Cross-Cultural Psychology*, 9: 131-149.

- [13] Cummins, J. (1979), "Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters," *Working Papers on Bilingualism*, 19: 197-205.
- [14] Cummins, J. (1981), "The role of primary language development in promoting educational success for language minority students," in California State Department of Education (Ed.), *Schooling and language minority students: a theoretical framework*, 3-49.
- [15] Cummins, J. (1991), "Interdependence of first- and second-language proficiency in bilingual children," in Bialystok, E. (ed.), *Language Processing in Bilingual Children*, Cambridge University Press.
- [16] Duflo, E. (2001), "Schooling and Labor. Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91(4): 795-813.
- [17] Ehrenberg, G. and Goldhaber, D. and Brewer, J. (1995), "Do Teachers Race, Gender, and Ethnicity Matter? Evidence from NELS:88," *Industrial and Labor Relations Review*, 48(3): 547-561.
- [18] Ethiopian Demographic and Health Surveys (EDHS) (2011).
- [19] Heckman, J. and Hotz, J. (1989), "Choosing Among Alternative Non Experimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84(408), 86274.
- [20] Heckman, J. and Stixrud, J. and Urzua, S. (2006), "The Effects of Cognitive and Noncognitive Abilities on Labor Market," *Journal of Labor Economics*, 24(3):411-82.
- [21] Heugh, K. and Benson, C. and Bogale, B and Gebre Yohannis, M.A. (2007), "Final Report: Study on Medium of Instruction in Primary Schools in Ethiopia," Commissioned by the Ministry of Education, Ethiopia.
- [22] Hovens, M. (2002), "Bilingual Education in West Africa: Does It Work?" *International Journal of Bilingual education and Bilingualism*, 5(5): 249-266.
- [23] *International Encyclopedia of Education* (Third Edition), Elsevier.

- [24] International Labour Organisation (2010), “Global Wage Report 2010/11: Wage policies in times of crisis.”
- [25] Jackson, T. (2000), “Equal Access to Education: A Peace Imperative for Burundi,” London: International Alert.
- [26] Klaus, D. (2003), “The Use of Indigenous Languages in Early Basic Education in Papua New Guinea: A Model for Elsewhere?” *Language and Education*, 17 (2):105-111.
- [27] Krueger, A. and Lindahl, M. (1998), “Education for Growth: Why and For Whom?” mimeo, Princeton University.
- [28] Lazear E. (1999), “Culture and Language,” *Journal of Political Economy*, Supplement, 95-125.
- [29] Laitin, D. (1994), “The Tower of Babel as a Coordination Game: Political Linguistics in Ghana,” *American Political Science Review*, 88 (3): 622-34.
- [30] *Language and National Identity in Africa*, (2008), Oxford University Press.
- [31] Lewis, M. and Lockheed, M (2006), “Inexcusable Absence: Why 60 Million Girls Still Aren’t in School and What to do About it,” Washington, D.C., Center for Global Development.
- [32] MacKinnon, J. G., and White, H. (1985), “Some heteroscedasticity-consistent covariance matrix estimators with improved finite sample properties,” *Journal of Econometrics*, 29, 53-57.
- [33] Marschak, J. (1965), “Economics of language,” *Behavioural Science*, 10, 135-40
- [34] Ministry of Education, Ethiopia (2010), “Adequacy and Effectiveness of Public Education Spending in Ethiopia.”
- [35] Patrinos, H. and Psacharopoulos, G. (1995), “Educational performance and child labor in Paraguay,” *International Journal of Educational Development*, 15(1): 47-60.
- [36] Patrinos, H. and Psacharopoulos, G., (2004), “Returns to investment in education: a further update,” *Education Economics*, 12(2), 111-134.

- [37] Patrinos, H. and Velez, E. (2009), “Costs and benefits of bilingual education in Guatemala: A partial analysis,” *International Journal of Educational Development*, 29 (6): 594-598.
- [38] Rosenbaum, P. (1987), “The role of a second control group in an observational study,” *Statistical Science*, 2, 292-316.
- [39] Strauss, J. and Thomas, D. (1995), Human Resources: Empirical Modeling of Household and Family Decisions, in Jere Behrman and T. N. Srinivasan, eds., *Handbook of development economics*, North Holland, 3A(9), 1885-2023.
- [40] Vawda, A. Y. and H. A. Patrinos (1999), “Producing Educational Materials in Local Languages: Costs from Guatemala and Senegal,” *International Journal of Educational Development*, 19: 287-299.
- [41] Verhoeven, L. (1994), “Transfer in bilingual development: The linguistic interdependency hypothesis revisited,” *Language Learning*, 44: 381- 415.
- [42] World Bank. (2005), “Education in Ethiopia : Strengthening the Foundation for Sustainable Progress.”
- [43] World Bank. (2005), “In their Own Language... Education for All, Education Notes.”
- [44] World Bank. (2004), “Ethiopia Public Expenditure Review: The Emerging Challenge Volume I and II: Public Spending in the Social Sectors 2000-2020.”
- [45] Zamarro, G. (2010), “Accounting for Heterogeneous Returns in Sequential Schooling Decisions,” *Journal of Econometrics*, 156: 260-276.

## Notes

<sup>1</sup>Refer to Sabates, R., Akyeampong, K., Westbrook, J., Hunt, F. (2010) for a recent review and international comparison of dropout and repetition rates.

<sup>2</sup>Refer to Thomas and Collier (2002) for the effect of provision of bilingual schooling on long-term academic achievement. They find that minority language students learning only in English start to show decreases in achievement by the beginning of the 5th grade.

<sup>3</sup>We use the terms language use in education, language policy and medium of instruction policy interchangeably in the paper. For our purpose they should be understood as the language used to teach children in schools.

<sup>4</sup>The enrolment rate in our data stands at around 70% implying there is much scope for increasing years of schooling also through the extensive margin.

<sup>5</sup>Refer to to year 2000 report by the Minority Rights Group International titled Ethiopia: A New Start?

<sup>6</sup>Haile Selassie was born from parents of three Ethiopian ethnicities, the Oromo and Amhara and the Gurage. He decided to adopt his Amharian heritage as the banner under which the centralization of Ethiopia was undertaken. The policy of centralization resulted in the choice of Amharic as the official language.

<sup>7</sup>The “Derg” period was characterized by strong state control. This period has sometimes been referred to as the Red terror campaign as wide scale human rights abuses were carried out by the establishment in power against any protesting voices.

<sup>8</sup>The results are not sensitive to the choice of ages of the younger and older cohorts. Expanding the definition of the younger and older cohort to range from -2 to 7 and 8 to 36, respectively, leaves our results essentially unchanged. We additionally in the results section estimate the effect of the language policy for each age in 1994 from 2 to 21 years old.

<sup>9</sup>A potential problem is that the data only lets us identify the current region of residence and not the actual region where education was obtained. In case the current region of residence is one of the four regions of the country where mother tongue instruction was introduced but the actual region of education was different from the current region of residence, then our estimates would provide a lower bound of the true effect as we wrongly consider untreated individuals as treated. The 2011 DHS data provides no information on childhood place of residence or for how long the individual has resided in the current region. Data on how long the individual has resided in the current region was however collected for the 2005 round of the DHS. Looking at the 2005 DHS round, we see that around 85% of the Oromo men have always lived in the current region of residence and around 70% of the women have always lived in the current region of residence. Internal migration moreover does not bias our estimates if it takes place within the four treated regions of the country. Given that nearly 90% of the Oromo population does indeed live within these four regions further reduces the possibility of potential bias due to internal migration.

<sup>10</sup>Delayed entry and grade repetition could mean that also some individuals who were aged 8 years in 1994 could also benefit from the policy change.

<sup>11</sup>The coefficient for the cohort aged 2 and 3 in 1994 are positive but just slightly below conventional significance level.

<sup>12</sup>The coefficient on the cohort aged 8 in 1994 is just significantly different from zero but as pointed out before this could be due to delayed entry and grade repetition.

<sup>13</sup>Language and National Identity in Africa, Oxford University Press 2008.

<sup>14</sup>The policy in fact has a positive effect at all levels of schooling and not just all levels of primary schooling. Results available on request.

<sup>15</sup>The choice of 6 years of education as the cutoff point is chosen as this is the number of years of required to finish primary schooling in most countries and completion of primary schooling by all by 2015 is an important objective in most African states. Using any other year as the cutoff does not change the essence of the result and as mentioned before the language policy changes has a positive effect at all levels of schooling.

<sup>16</sup>We are aware of the problems associated with using Mincerian wage regression such as the endogeneity of post-schooling human capital accumulation, the fact that schooling and training are treated symmetrically in calculating the rate of return to schooling and that the general equilibrium affects are not accounted for. This said most instrumental variable estimates are found to be larger than the ordinary least squares estimates suggesting if anything we are underestimating the benefits of the policy change.

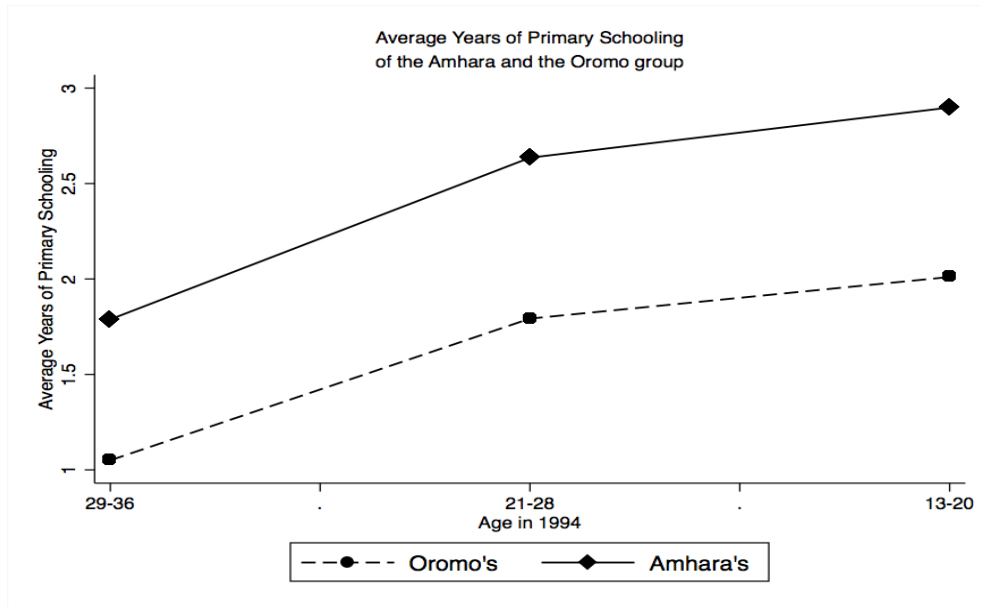


Fig. 1: Trends Pre-1994 for the Amhara and the Oromo Language Group

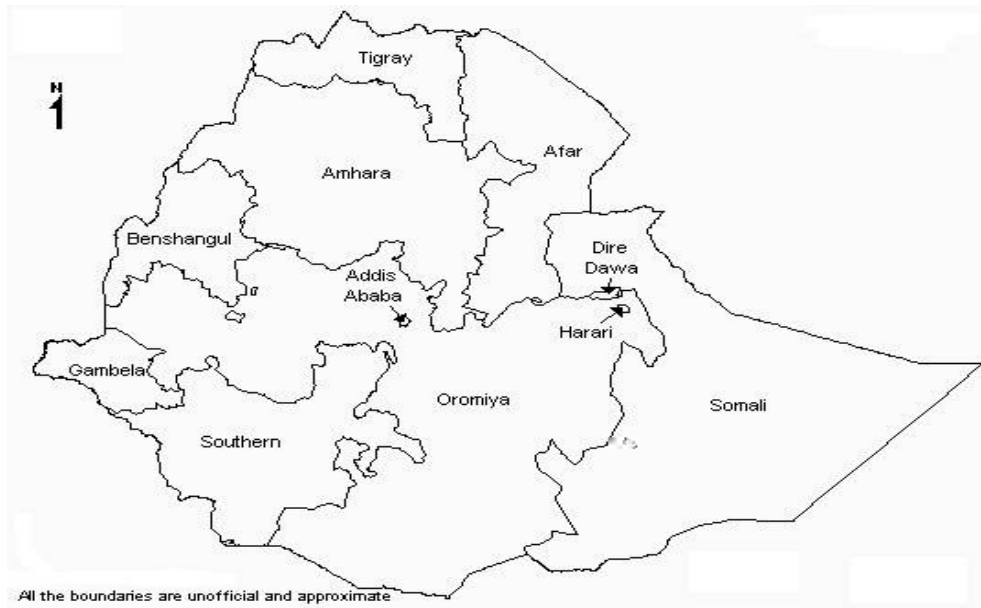
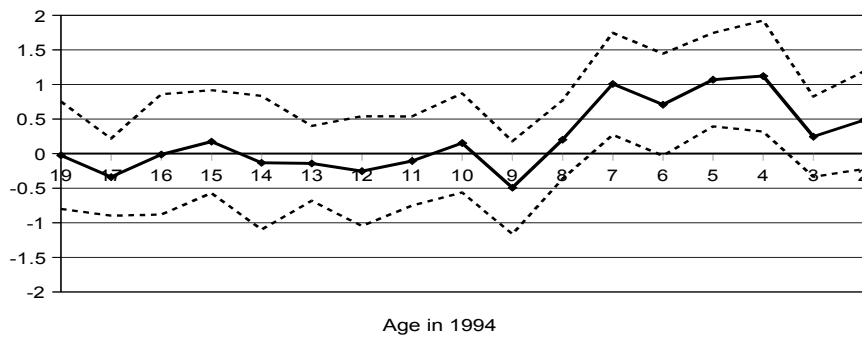


Fig. 2: Map showing the regions and city administrative units of Ethiopia

Panel A  
Coefficient of interaction age in 1994 with the Oromo language group dummy



Panel B  
Coefficient of interaction age in 1994 with the regions where the Oromo are treated

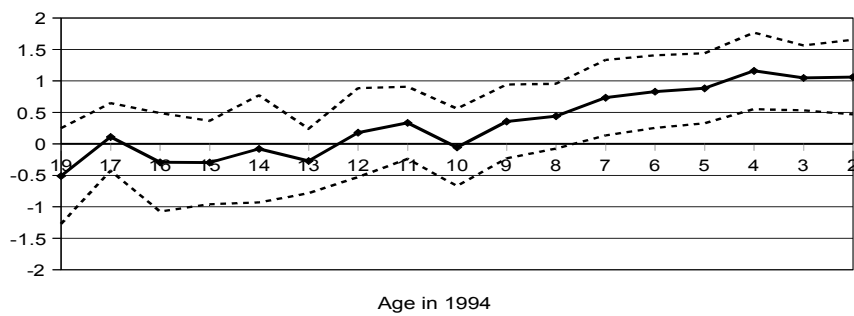
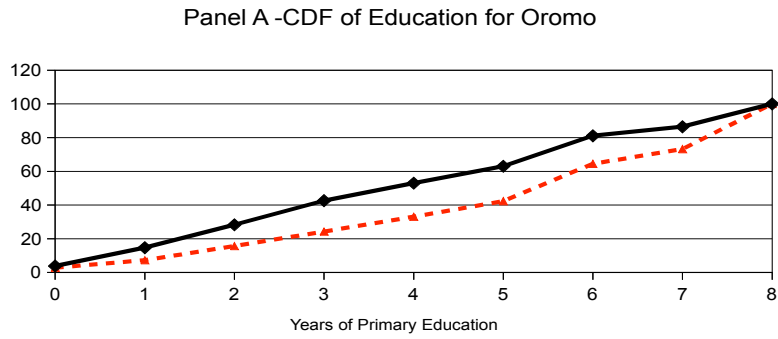
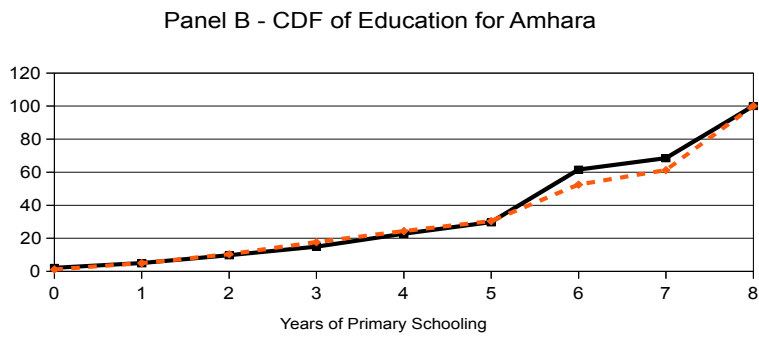


Fig. 3: Effect the effect of the policy change for each cohort

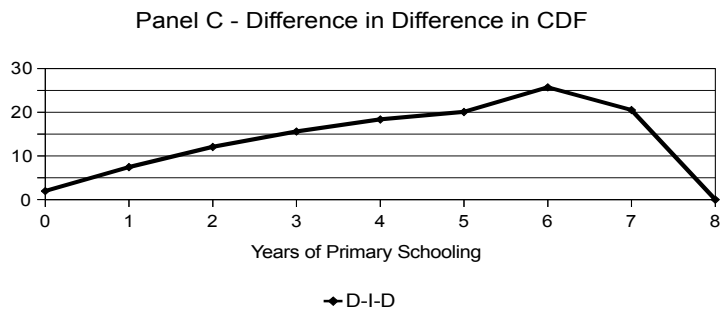




— Oromo Cohort Aged 2 to 7 in 1994    — Oromo Cohort Aged 13 to 20 in 1994



— Amhara Cohort aged 2 to 7 in 1994    — Amhara cohort aged 13 to 20 in 1994



— D-I-D

Fig. 4:

Table 1: Medium of Instruction (MOI) Policy in Ethiopia by Regions.

Region Name	Languages implemented as MOI in primary schooling
ADDIS ABABA	Amharic
DIRE DAWA	Amharic, Oromigna, Somali.
AFAR	Amharic and Afar.
AMHARA	Amharic, Awingi, Hamittlena and Oromigna.
BENISHANGUL GUMUZ	Amharic
GAMBELLA	Nuek, Anguak and Meshenger.
HARARI	Amharic, Harari and Oromigna.
OROMO	Amharic and Oromigna
SNNPR	Amharic, Dawro, Gamo, Gedeo, Gofa, Hadiya, Kembata, Kafinono, Kotigna, Sidama and Wolaita.
SOMALI	Amharic and Somali.
TIGARY	Tigrinya.

a. The medium of instruction in primary schooling was Amharic in all the regions prior to 1994.

b. Source: Heugh, K. and Benson, C. and Bogale, B and Gebre Yohannis, M.A. (2007). Final Report: Study on Medium of Instruction in Primary Schools in Ethiopia. Commissioned by the Ministry of Education, Ethiopia. September to December 2006.

Table 2: Descriptive Statistics for Individuals aged 2 to 7 or 13 to 20 in 1994

Variable	Country Average	Amhara Group	Oromo	Non- Amhara Group
Mean Years of Schooling	4.35	6.27	3.76	3.39
Mean Years of Primary Schooling	3.04	4.31	2.98	2.76
% of Households classified as Middle Class and Below	53	30	47.29	57.43
Age of Household Head	40.44	41.16	39.21	39.72
% of Households with Male Head	80	73	84	79.9
% of Households with Bank Accounts	13	29	9	9.4

Table 3: **Average years of schooling by Language Group and Cohort**

	Years of primary schooling of the Oromo Language Group	Years of primary schooling of the Amhara Language Group	Difference
<b>Panel A: Experiment of Interest</b>			
	(1)	(2)	(3)
2-7 Years old in 1994	3.81 (1327, 3.13)	4.16 (1430, 3.19)	-0.35
13-20 Years old in 1994	1.82 (1310, 2.62)	2.89 (1297, 3.35)	-1.07
Difference	1.99	1.27	0.72
<b>Panel B: Control Experiment</b>			
13-20 Years old in 1994	1.82 (1310, 2.62)	2.89 (1297, 3.35)	-1.07
21 - 28 Years olds in 1994	1.65 (108, 2.63)	2.65 (368, 1.99)	-1.00
Difference	0.17	0.24	-0.07

a. Number of observations and standard errors in parentheses.

b. The observations are from the regions where the Oromo younger cohort is treated, namely Amhara, Oromia, Harari and Dire-Dawa.

Table 4: Impact of provision of mother tongue instruction on Oromo people

	(1)	(2)	(3)
<b>Panel A: Experiment of Interest: Individuals Aged 2 to 7 or 13 to 20 in 1994.</b>			
<b>(Youngest cohort aged 2 to 7 in 1994)</b>			
Cohort Dummy*Oromo Language Group Dummy	0.721*** (0.169)	0.745*** (0.154)	0.797*** (0.206)
Cohort Dummy	1.270*** (0.126)	1.580*** (0.265)	2.342*** (0.353)
Oromo Language Group Dummy	-1.075*** (0.118)	-2.868*** (0.129)	-4.394*** (0.179)
Other Controls	No	Yes	Yes
Observations	5,364	5,364	5,364
R-squared	0.080	0.240	0.424

	(1)	(2)
<b>Panel B: Control Experiment: Individuals Aged 13 to 20 or 21 to 28 in 1994.</b>		
<b>(Youngest cohort aged 13 to 20 in 1994)</b>		
Cohort Dummy*Oromo Language Group Dummy	-0.095 (0.180)	-0.009 (0.153)
Cohort Dummy	0.263* (0.140)	0.495* (0.261)
Oromo Language Group Dummy	-0.979*** (0.136)	-3.373*** (0.146)
Other Controls	No	Yes
Observations	4,448	4,448
R-squared	0.030	0.309

a. The dependent variable in column (1) and (2) is years of primary schooling and in column (3) is years of schooling.

b. Other controls include year of birth and region dummies.

c.  $HC_2$  standard errors are in parentheses.

d. \*, \*\* and \*\*\* significant at 10, 5 and 1 % significance level respectively.

e. The Amhara are the control groups the regions are where the Oromo group is treated, namely, Amhara, Oromo, Harari and Dire-Dawa.

Table 5: **Average years of Primary Schooling by Language Group and Cohort**

	Years of primary schooling of the Oromo people in Treated Areas	Years of primary schooling of the Oromo people in the Untreated Areas	Difference
Experiment of Interest	(1)	(2)	(3)
2-7 Years old in 1994	3.54 (1327, 3.13)	4.28 (305, 3.62)	-0.47
13-20 Years old in 1994	1.82 (1310, 2.62)	3.04 (268, 3.25)	-1.22
Difference	1.99	1.24	0.75

a. Number of observations and standard errors in parentheses

b. The Oromo Language group are treated in the regions of Amhara, Oromia, Harari and Dire-Dawa and untreated in the regions of Afar, Benishangul Gumuz, Gambella, SNNPR, Somali and Tigray.

Table 6: **Comparing the Oromo in the treated and untreated regions**

	(1)	(2)	(3)
Cohort Dummy*Region Dummy	0.751*** (0.287)	1.00*** (0.269)	1.178*** (0.371)
Cohort Dummy	1.240*** (0.264)	1.233*** (0.429)	1.877*** (0.581)
Regional Dummy	-1.22*** (0.212)	-1.37 (1.65)	-2.553 (2.472)
Other Controls	No	Yes	Yes
Observations	3,210	3,210	3,210
R-squared	0.104	0.154	0.160

a. The dependent variable in column (1) and (2) is years of primary schooling and in column (3) is years of schooling.

b. Other controls include year of birth and region dummies.

c.  $HC_2$  standard errors are in parentheses.

d. \*, \*\* and \*\*\* significant at 10, 5 and 1 % significance level respectively.

**Table 7: Proportion Enrolled by Language Group and Cohort**

	Proportion Enrolled of the Oromo Language Group	Proportion Enrolled of the Amhara Language Group	Difference
	(1)	(2)	(3)
2-7 Years old in 1994	0.709 (1327, 0.45)	0.711 (1430, 0.45)	-0.002
13-20 Years old in 1994	0.426 (1310, 0.49)	0.484 (1297, 0.49)	-0.058
Difference	0.283	0.227	0.056

a. Number of observations and standard errors in parentheses.

b. The observations are from the regions where the Oromo younger cohort is treated, namely Amhara, Oromia, Harari and Dire-Dawa.

Table 8: PROBIT regression: Impact of provision of mother tongue instruction on Oromo people on enrolment and completion

	(1) Dependent variable Dummy - Enrolled	(2) Only Enrolled Sample Dependent variable Dummy -6 years of schooling or more	(3) Full Sample Dependent Variable Dummy -6 years of schooling or more
Experiment of Interest: Individuals Aged 2 to 7 or 13 to 20 in 1994.			
(Youngest cohort aged 2 to 7 in 1994)			
Cohort Dummy*Oromo Language Group Dummy	0.0466 (0.074)	0.318*** (0.100)	0.249*** (0.0780)
Cohort Dummy	0.375*** (0.125)	0.327* (0.172)	0.402*** (0.132)
Oromo Language Group Dummy	-0.951*** (0.065)	-1.090*** (0.0856)	-1.284*** (0.0664)
Observations	5,364	3,147	5,364

a. All specifications include year of birth and region dummies.

b. \*, \*\* and \*\*\* significant at 10, 5 and 1 % significance level respectively.

c. The Amhara are the control groups the regions are where the Oromo group is treated, namely, Amhara, Oromia, Harari and Dire-Dawa.



Table 9: Impact of Provision of Mother Tongue Instruction for African Countries on Primary Schooling Attainment.

(1) Country	(2) Current % of pop. completing Primary Schooling	(3) % of pop. completing Primary Schooling with mother tongue provision	(4) Ethnic groups with pop. shares of 10% and more	(5) Share of the Ethnic groups completing Primary Schooling currently	(6) Share of the Ethnic groups completing Primary Schooling with mother tongue provision
<b>BENIN</b>	29%	38.5%	Adja (11.96%), Bariba (9.67%) and Fon (42.66%)	Adja (25.83%), Bariba (11.14%) and Fon (26.98%)	Adja (41.07%), Bariba (17.95%) and Fon (43.18%)
<b>BURKINA FASO</b>	15%	20%	Mossi (56%)	Mossi (16%)	Mossi (23%)
<b>CAMEROON</b>	59.8%	65%	Bamblike (20%) and Beti (9.01%)	Bamblike (81%) and Beti (80%)	Bamblike (100%) and Beti (100%)
<b>GABON</b>	66%	80%	Fang (25%), Nzabi-duma (10.2%) and Shira-pun (21%)	Fang (82%), Nzabi-duma (58%) and Shira-pun (69%)	Fang (100%), Nzabi-duma (89%) and Shira-pun (99%)
<b>GHANA</b>	70%	84%	Akan (47%), Ewe (13.7%) and Mole-dagiban (16%)	Akan (83.5%), Ewe (76.5%) and Mole-dagiban (40%)	Akan (100%), Ewe (100%) and Mole-dagiban (57.75%)

a. In column (4) in parenthesis are the population share of each group.

b. In column (5) in parenthesis are the % of population from each group completing primary schooling currently.

c. In column (6) in parenthesis are the % of population from each group completing primary schooling after provision of mother tongue instruction.

Table 10: Net Present Value arising for an Individual from Provision of Mother Tongue Instruction.

Country	Discount rate	GDP per capita	Coefficient on years of schooling	% of GDP per capita spent per student	NPV
BURKINA FASO	10%	212	9.6	30%	121
ETHIOPIA	10%	178	8.0	18%	99
GHANA	10%	360	7.1	12%	193

## IX Appendix

### IX.1 Comparing the Amharas to the Non-Amharas

Table A: Average years of schooling by Language Group and Cohort

	Years of primary schooling of the Non-Amhara Language Group	Years of primary schooling of the Amhara Language Group	Difference
<b>Panel A: Experiment of Interest</b>			
	(1)	(2)	(3)
2-7 Years old in 1994	3.54 (4538, 2.66)	4.31 (2001, 3.15)	-0.77
13-20 Years old in 1994	1.74 (4794, 2.66)	3.03 (1756, 3.32)	-1.29
Difference	1.80	1.28	0.52
<b>Panel B: Control Experiment</b>			
13-20 Years old in 1994	1.74 (4794, 2.66)	3.03 (1756, 3.32)	-1.29
21 - 28 Years olds in 1994	1.45 (3284, 2.37)	2.65 (1604, 2.84)	-1.20
Difference	0.29	0.38	-0.09

a. Number of observations and standard errors in parentheses

b. The sample includes all regions of the country where the Non-Amharic language group is treated, namely Afar, Amhara, Benishangul Gumuz, Dire-Dawa, Gambella, Harari, Oromia, SNNPR, Somali and Tigray.

**Table B: Impact of provision of mother tongue instruction on non-Amharic language people: Coefficient of interaction between cohort dummy and non-Amharic Language Group Dummy**

Dependent variable - Number of years of Primary Schooling		
	(1)	(2)
<b>Panel A: Experiment of Interest: Individuals Aged 2 to 7 or 13 to 20 in 1994.</b>		
<i>(Youngest cohort aged 2 to 7 in 1994)</i>		
Cohort Dummy*Non-Amharic Language Group Dummy	0.577*** (0.130)	0.553*** (0.116)
Cohort Dummy	1.210*** (0.114)	1.578*** (0.191)
Non-Amharic Language Group Dummy	-1.31*** (0.094)	-2.40*** (0.102)
Other Controls	No	Yes
Observations	11,918	11,918
R-squared	0.091	0.207
Dependent variable - Number of years of Primary Schooling		
	(1)	(2)
<b>Panel B: Control Experiment: Individuals Aged 13 to 20 or 21 to 28 in 1994.</b>		
<i>(Youngest cohort aged 13 to 20 in 1994)</i>		
Cohort Dummy*Non-Amharic Language Group Dummy	-0.055 (0.143)	0.035 (0.117)
Cohort Dummy	0.358*** (0.130)	0.672*** (0.193)
Non-Amharic Language Group Dummy	-1.259*** (0.107)	-2.741*** (0.113)
Other Controls	No	Yes
Observations	10,281	10,281
R-squared	0.041	0.198

a. Other controls include year of birth and region dummies.

b.  $HC_2$  standard errors are in parentheses.

c. The sample includes all regions of the country where the Non-Amharic language group is treated, namely Afar, Amhara, Benishangul Gumuz, Dire-Dawa, Gambella, Harari, Oromo, SNNPR, Somali and Tigray.

d. \*, \*\* and \*\*\* significant at 10, 5 and 1 % significance level respectively.

## IX.2 Methodology applied to calculate effect of mother tongue instruction on a set of African countries

The procedure utilized to calculate the change in the proportion of population completing primary schooling, due to provision of mother tongue instruction, is best explained with the help of an example. The same procedure is then applied to a set of five African countries and the results are shown in table 11.

Benin is used as an example to outline the mechanics underlying the procedure. We consider the population aged 15 to 49 from the DHS data of 2006. The DHS data provides us with information on both the education attainment<sup>17</sup>, the language of the respondent and whether the individual was ever enrolled in schooling. Approximately 29% of the population aged 15 to 49 in our sample completes primary schooling or more. Looking at the population shares of the various language groups, there are three groups with population shares of 10% or more. The “Adja ” people comprise about 11.96%, the “Bariba” people about 9.67%<sup>18</sup> and the “Fon” about 42.6% of the population. The percentage of population completing primary schooling for the three groups are 25.83%, 11.14% and 26.98%, respectively. We have information on the CDF of education for the three groups both conditional on enrolment and for the whole sample i.e. both individuals who enrolled and did not enrol. The estimate from the previous section suggests that the provision of primary schooling induced 31% of the enrolled sample to complete 6 years of schooling or more as compared to 5 years or less. The sample size of the enrolled population, for instance for the ethnic group Fon, is 7609 individuals. The calculated estimate implies that now 31% or 2419 additional individuals complete primary schooling as compared to before. Looking at the entire sample of individuals aged 15 to 49 from the Fon group, which includes 14935 individuals, we see that 10905 individuals were dropping out with 5 years of education or less. The provision of mother tongue, given our estimate, would reduce the number of individuals dropping out with 5 years of education or less from 10905 individuals to about 8485 individuals (i.e. 10905 minus 2419). This would imply that the proportion of population which now completes primary schooling or more increases from around 26.98% to 43.18%. Doing a similar exercise for the Adja and the Bariba people suggests that the proportion of people completing primary schooling or more would increase from 25.83% and 11.14% to 41.07% and 17.95% respectively. The percentage of population completing primary schooling for the other language group remains unchanged. We hence using the new values of the percentage of population completing primary schooling for the three

groups along with their population shares calculate the overall change in the percentage of population completing primary schooling or more. The estimate suggests that the provision of mother tongue instruction to these three groups, which comprise about 64% of the population, would increase the percentage of population completing primary schooling, for the people aged 15 to 49, in the country from around 29% to 38.5%, an increase of nearly 10% points.