Andrén, Daniela; Levin, Jörgen; Vimefall, Elin

Working Paper

Ethnolinguistic Background and Enrollment in Primary Education: Evidence from Kenya

Working Paper, No. 7/2015

Provided in Cooperation with:
Örebro University School of Business

Suggested Citation: Andrén, Daniela; Levin, Jörgen; Vimefall, Elin (2015) : Ethnolinguistic Background and Enrollment in Primary Education: Evidence from Kenya, Working Paper, No. 7/2015, Örebro University School of Business, Örebro

This Version is available at:
https://hdl.handle.net/10419/244511

WORKING PAPER

7/2015

Ethnolinguistic Background and Enrollment in Primary Education: Evidence from Kenya

Elin Vimefall, Daniela Andrén and Jörgen Levin

Economics

ISSN 1403-0586

http://www.oru.se/Institutioner/Handelshogskolan-vid-Orebro-universitet/Forskning/Publikationer/Working-papers/
Örebro University School of Business
701 82 Örebro
SWEDEN

Ethnolinguistic Background and Enrollment in Primary Education: Evidence from Kenya

Elin Vimefall,[1] Daniela Andrén and Jörgen Levin

Abstract
In Kenya, educational enrollment rates increased significantly for both girls and boys after 2003, when primary education became free of charge. Unfortunately, approximately one million school-aged children are still not enrolled in school. Earlier literature provides empirical evidence that educational opportunities differ among children, due to poverty, gender, rural area of residence and disability. Our paper contributes to the literature by providing empirical evidence of the importance of children's ethnolinguistic background for their probability of being in school. Estimates from a three-level random intercept probit model using data from the Kenya Integrated Household Budget Survey 2005/06 reveal that Somali and Maasai children are least likely to be in school. A separate analysis by child's gender shows that compared to Kikuyu children both girls and boys from the Somali and Maasai groups, but also Mijikenda and Swahili girls, have a lower probability to be in school. This might be an indication that gender norms are stronger in these groups.

Keywords: school-aged children, school enrollment, free primary education, ethnolinguistic background, Kenya, three-level random intercept model.

1 Corresponding author: elin.vimefall@oru.se

We are grateful for comments from Fredrik Sjöholm, Henrik Jaldell and Therese Nilsson. We also want to thank participants at the Nordic Conference in Development Economics 2013, the SWEGPEC Ph.D. workshop and seminar participants at Örebro University, the University of Gothenburg and the University of Nairobi.

1 Introduction

A better-educated population is expected to lead to reduced poverty, stronger economic growth and higher individual wellbeing. Since the 1950s, several international agreements have resulted in measurable gains with respect to children's access to education and completion of primary school;[2] as of 2012, however, approximately 58 million primary-school-aged children around the world were still not enrolled in school (UNDP 2014). The previous literature has reported that children's educational opportunities differ because of poverty, gender, rural area of residence and disability (Global Partnership for Education 2012). Our paper contributes to the existing literature by examining the relationship between ethnolinguistic background and the child's probability of being in school. We expect that several attributes related to the household's ethnolinguistic background, including culture, norms and language, affect the household's decision to invest in children's human capital by letting them attend school.

Our paper focuses on Kenya, one of the world's most diversified countries, in terms of both ethnicity and language (Alesina et al. 2003). There are approximately 70 languages spoken in Kenya (Lewis et al. 2013) and ethnicity has played a central role in both political mobilization and resource allocation in the country (Kimenyi 1997). Although primary education has been free in Kenya since 2003, in 2011 approximately one million school-aged children were not enrolled in school (UNESCO 2013).

We use data from the Kenya Integrated Household Budget Survey (KIHBS) 2005/06. The novelty of our paper lies in the way in which we control for the ethnolinguistic background of the child. We build on data evidencing that many respondents in rural areas chose to answer the questionnaire in one of the eleven local languages that were allowed to be used as an alternative to either English or Swahili, the two official languages of Kenya. Given that the data collection process was designed in terms of clusters of ten households living in a small community, it is likely that rural households that answered in the same local language share the same culture and norms. Therefore, we decided to focus only on the rural areas and argue that the language in which a respondent answered the questionnaire is a good proxy for the household's ethnolinguistic background. To our knowledge, this study is the first to control for the importance of ethnolinguistic background when studying the decision over whether to send a child to school, in addition to traditional variables related to child-, household- and region-specific characteristics.

To analyze the importance of the household's ethnolinguistic background on the child's probability of being in school, we use a three-level random intercept probit model that controls for unobserved characteristics at both the community and household level. Regardless of the model specification, our results show that Somali and Maasai children have a lower probability of being in school than children from all other language groups. A separate analysis by child's gender shows that compared to Kikuyu children both girls and boys from the Somali and Maasai groups, but also Mijikenda and Swahili girls, have a lower probability to be in school. This might be an indication that gender norms are stronger in these groups.

The remainder of this paper is organized into six sections. Section 2 reviews earlier studies of educational decisions and formulates the hypotheses that are tested herein. Section 3 presents relevant institutional settings and data. Section 4 discusses the empirical strategy. Section 5 discusses the results, and Section 6 concludes.

2 Earlier studies and our paper's hypotheses

In the classical human capital framework (Schultz 1960, 1963, Mincer 1958, 1974, Becker 1975), the optimal level of schooling is determined by the marginal cost and marginal benefit of education, both of which are affected by preferences and beliefs that can be held by individuals and/or groups (Fernández and Fogli 2005). We assume that a household's expectations regarding the costs and benefits of education vary across groups with different ethnolinguistic backgrounds, i.e., are differentiated according to culture, norms and language. Our study's main hypothesis is that ethnolinguistic background is an important determinant of the decision to be in school or not.

2 For example, in the 1950s, the Universal Declaration of Human Rights proclaimed, "Everyone has the right to education". See UNESCO's Education for All and the Millennium Development Goals (MDGs), particularly MDG 2 on universal primary education and MDG 3 on gender equality in education by 2015.

The previous economics literature addressing the determinants of primary schooling in Kenya has found that school enrollment is related to the characteristics of the children, the level of education of the parents, the welfare and composition of the family and household, the cost of sending a child to school and the area of residence (Kabubo-Mariara and Mwabu 2007, Nyokabi 2009). The direct costs of schooling include books, uniforms, transportation and tuition, among others, whereas the opportunity cost of education consists of the child's time that could have been used in household production or in the labor market.

Even though primary education in Kenya is tuition free, there are both direct and indirect costs that not all households can afford, and therefore, for some children, particularly in rural areas, working on a household farm or in a household business may be the alternative to spending time in school. Studies have shown that achieving universal primary education will require government funding that does more than merely remove infrastructure fees; at the very least, activity and examination fees must also be eliminated (e.g., Mukudi 2004, Omwami and Omwami 2010). Therefore, we expect that a household's income and/or poverty level are important for the household's decision regarding the child's school attendance.

Being able to understand and speak the language of instruction is obviously important for learning (Trudell 2005). Earlier literature highlights the importance of language as a possible explanation for achievement gaps between indigenous and non-indigenous children in some countries (Patrinos and Psacharopoulos 1996 for Bolivia and Guatemala, McEwan and Trowbridge 2007 for Guatemala). African educationists argue that the major learning problem for African children is linguistic in nature (Obanya 1980), and African linguists emphasize the advantages of using the mother tongue as the language of instruction (Brock-Utne 2001). Children who speak a language other than the language of instruction experience barriers to learning, and these barriers might be essential to the decision over whether to remain in school or drop out (Lockheed and Verspoor 1991). Therefore, we expect that children who belong to ethnolinguistic groups that speak a language other than the neighborhood school's language of instruction will be less likely to be in school.

Moreover, ethnolinguistic background might also incorporate culture and norms. Households from the same ethnolinguistic group sometimes share the same way of living: some groups might be characterized by nomadic traditions, others by being traders, etc. When Buchmann (2000) analyzed how household structure and parental perception influence the child's probability to be in school in Kenya, she included ethnicity as an explanatory variable. Buchmann (2000)'s study builds on the hypothesis that the link between ethnicity and education has deep historical roots; some ethnic groups (e.g., Kikuyu, Embu and Meru) are likely to value education more due to their earlier exposure to Western culture and Christianity, whereas other ethnic groups (e.g., the Mijikenda, Maasai and Somali) were typically exposed to Western-style education much later. Using a relatively limited data set (n = 643), she tested whether Kikuyu and Mijikenda children have different probabilities of being in school, using all other groups for comparison. However, the empirical results did not support the hypothesis that ethnicity affects school enrollment. Having richer data (n=11389) and a different data design that focuses on ethnolinguistic background instead of ethnicity, we test whether such background has an impact on children's probability of being in school.

Gender-based differences in education are larger among ethnic minorities, and ethnicity-based differences in education are stronger among women than men (Tas et al. 2014), and approximately 70 percent of the world's out-of-school girls come from excluded groups (Lewis and Lockhead 2007). This suggests that the probability of being in school might be affected by gender norms, and that these norms vary among ethnolinguistic groups. If girls are expected to marry relatively earlier and leave the family after marriage, while boys are expected to inherit familial property, then educating sons will yield a higher return for the parents. Furthermore, if girls are expected to become homemakers and not enter the labor market, their expected return on education might be lower. Buchmann (2000) reported that, if parents expect the labor market to be worse for women than for men, then their daughters are less likely to be enrolled in school.[3] Therefore, we expect that, if gender norms are part of the ethnolinguistic background of the household, then our language dummies will have different impacts on children's probability of being in school for boys and girls.

[3] Based on an earlier literature review of culture and economic outcomes, Fernandez (2010) reported that culture was important for explaining economic behavior.

Another hypothesis about the differences in school attendance across ethnic groups comes from political economy's approach, which argues that political leaders in Sub-Saharan Africa are known to favor their own ethnic group (Franck and Rainer 2012). Using data from Kenya, Kramon and Posner (2012) reported empirical evidence that having a president from one's own ethnic group during the years when one was of primary education age increased one's years of education by about 4 percent. They explained that this was due to both targeting from elites, who increased the inputs to schooling and health, and to a change in parents' expectations, who anticipated a higher return from education when they had a co-ethnic as president. Given that the president of Kenya during 2005/06 was an ethical Kikuyu, Mwai Kibaki, we expect that this would have increased the probability of Kikuyu children being in school.[4]

3 Institutional settings and data

The Kenyan education system is highly centralized. Formal education is managed by the Ministry of Education, Science, and Technology (MoEST), which is responsible for policy and planning. Together with the Teachers Service Commission (TSC), the Kenya Institute of Education, and the Kenya National Examination Council, MoEST sets the curricula, administers the public examinations and pays the salaries of the teachers in public schools. The TSC is responsible for the recruitment and allocation of teachers. Private schools do not receive any public funds (World Bank 2004).

Since 1985, Kenya's educational system has been designed to provide eight years of primary education, four years of secondary education, and four years of tertiary education. There are 16 local languages approved for instruction, and until the fourth year the language of instruction should be the predominant language of the school's catchment area. Thereafter, English should be used as the language of instruction (Cleghorn et al. 1989). The goal of the language policy is to make primary education locally accessible to linguistic/ethnic minorities, but it is difficult to implement mother-tongue teaching in regions in which several ethnolinguistic groups co-exist (Ogle et al. 2010). Therefore, English or Swahili may be used as the initial medium of instruction in linguistically mixed schools. Glewwe et al. (2009) argued that the Kenyan educational system had developed to serve the strongest students, while being ill-suited to the typical student. As an example of this, they found that providing textbooks did not raise the average test score, but had a positive impact on the test scores of the best students. One explanation given was that the textbooks are written in English, which is the third language of most students.

In 2003, Kenya implemented a free primary education (FPE) policy aiming to give every child access to education. Consequently, enrollment rates increased by approximately 14 percent between 2002 and 2003 (World Bank 2004). The increase in enrollment after the introduction of FPE has varied across districts, and poorer districts have experienced the largest increase (Bold et al. 2010). However, other school related inputs have not kept pace with the increase in the number of students, leading to a potential decrease in the quality of education. For example, the student-to-teacher ratios increased from 33:1 to 39:1 between 2000 and 2003 (World Bank 2004). Some households have reacted by transferring their children to private schools (Bold et al. 2011). Nishimura and Yamano (2013) showed that this reaction was more common among wealthier households and that girls were less likely to be transferred than boys. As a result of this, as Lucas et al. (2012) showed, there has been an increase in the concentration of students from low socioeconomic groups in public schools.

Our data were extracted from the KIHBS 2005/06, which was conducted by the Kenya Bureau of Statistics over a period of 12 months beginning on May 16, 2005. Its respondents included 66725 individuals belonging to 13430 randomly selected households from 861 rural and 482 urban clusters that were, in turn, randomly selected from all the districts in Kenya. The sample is representative not only at the national level but also at the urban/rural, provincial and district levels. The survey covered all household members (usual residents).

According to the KIHBS 2005/06 data, approximately 93 percent of children aged 6-17 had attended school for at least one term. In the case of children who had never attended school, lack of money for school expenses (20 percent) or ill health (9.9 percent) were stated as the main reasons.

[4] Mwai Kibaki was president during 2002-2013.

We selected the sub-sample of all rural children aged 6-14. As mentioned in the introduction, our reason for focusing on the rural households is that a high proportion of households from the rural areas answered the questionnaire in a local language. After cleaning the data of non-responses, we ended up with a final sample of 11138 children. Table A1 in the appendix presents a short description of all variables used in the empirical analysis, and Table A2 presents descriptive statistics for all the children in our sample, by school enrolment and by gender.

About 90 percent of the children in the final sample were enrolled in school, and there was no significant difference between boys and girls. Of the children who were in school, about 91 percent were in public schools, 8 percent in private schools and the remaining 1 percent in community schools.

There were significant differences between the children who were in school and the ones that were not. Children living in households defined as "hard core poor" make up 28 percent of the children in our final sample and 52 percent of those children who were not in school.[5] This indicates that, although primary education is free, poverty continues to keep children from being in school. The enrollment level varied among districts: although almost all the children in Central Province (98 percent) were in school, the corresponding rate for North Eastern Province was only 54 percent.

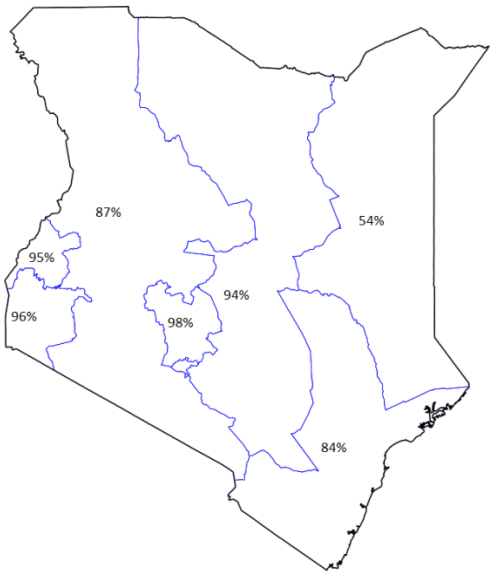


Figure 1: *Percentage of children aged 6-14 living in rural areas who were in school at the time of the survey*

As already mentioned, given that, in Kenya, there is a strong connection between ethnicity and language, we considered the language in which the household members answered the questionnaire to be a good proxy for the household's ethnolinguistic background. There were 13 alternatives that could be used to answer: Swahili and English, the two official languages of Kenya, and eleven local languages (Embu, Kalenjin, Kamba, Kikuyu, Kisii, Luhya, Luo, Maasai, Meru, Mijikenda, and Somali). According to Kenya Census (2009), approximately 89 percent of the population belongs to one of the ethnic groups connected to these eleven languages (Kenya National Bureau of Statistics 2010). The Swahili- and English-language dummies might be contaminated by the fact that respondents who spoke a language other than one of the local languages offered by the survey chose to answer in one of the two official languages. However, ethnic Swahili speakers primarily live in Coast Province. Therefore, we constructed a Swahili-coast dummy that took the value of one for children living in households whose members were born in and lived in Coast Province and answered the survey in Swahili (5.6 percent), and zero otherwise. All other children who

5 A household was defined as hard core poor if it could not meet its basic food needs even if it only consumed food. The food poverty line was based on the cost of consuming 2250 kilocalories per adult equivalent per day. In monthly adult equivalent terms, the food poverty line was 988 Kenyan shillings for the rural areas (Kenya National Bureau of Statistics 2007).

belonged to a household that answered in Swahili were put into the language group "Swahili-rest" (22.4 percent). Only 2.9 percent of the children lived in a household that answered the survey in English; most of those children (94 percent) lived in the Moyale district, which is located between Kenya and Ethiopia. Notably, all the respondents from this district answered in English.

The enrollment rate in primary school varied widely among the ethnolinguistic groups. Whereas nearly all Kikuyu and Kamba primary school-aged children were in school (98 percent), only 55 percent of the Somali children were. Educational traditions, measured as the education of the head of the household, also varied among the ethnolinguistic groups. Whereas 34 percent of the Kisii children lived in a household in which the head had at least some secondary education, the same held true for only 1 percent of the Somali group. Moreover, the Somali and Mijikenda children lived in the poorest households.

Although our data do not explicitly include information about the language of instruction, we are aware that this factor might be an important component of explanations of the differences in schooling between ethnolinguistic groups.

The traditional variables used in the empirical analysis can be divided into three groups: child characteristics, household characteristics and community/district characteristics. Child characteristics include the age and gender of the child and a variable indicating whether the child has had a chronic illness. Household characteristics include the education of the head of household (no education, primary education or some secondary education), household expenditure, whether or not the household owns land that is used for agricultural purposes, a variable indicating whether the household is defined as hard core poor, and a set of dummy variables for the household's religion (Catholic, Protestant, Muslim, other or no religion).

We considered several characteristics of the small community and/or district. Where possible, the mean values were calculated at the small community level. However, some variables were only available at the district level (quality of education and distance to school). The cost of education (mean expenditure on education for a student in a government primary school divided by 1000) was almost the same (around 0.56) for all observations, with relatively smaller numbers for those children who were not in school. The quality of education (percentage of students in the second class that could write) was higher among the sample of enrolled children, which might suggest that some parents do not let their children go to school because of low expectations. The distance to school (percentage of households located more than 5 kilometers from the nearest public primary school) also differed between the enrolled and unenrolled children; for children who were not in school, the percentage of households in their area with more than 5 kilometers to travel to the nearest public school was higher, suggesting that the child's probability of being in school is dependent on the distance to school. The rate of working children, measured as the percentage of children aged 6-14 in a community for whom work was given as their main activity, was much higher (almost double) for the unenrolled than enrolled children. This descriptive also suggests that there might be some kind of peer effect motivating parents to either let their children go to school or not. The rate of homemakers, measured as the percentage of women in the community who stated "being a homemaker" as their main activity, was higher among the households with children who were not enrolled in school, indicating that gender norms might be important.

4 Empirical strategy

Our data have a multilevel structure: each child belongs to a household and each household belongs to a small community. Children might share several common observed and unobserved attributes at both levels, and it is reasonable to assume that these shared characteristics affect the probability of being in school. An appropriate model with which to analyze the decision to send a child to school in this setting is a random intercept model that accommodates three levels: children nested within a household, and households within a small community. We considered a random intercept model because it allowed us to model grouped cross-sectional data for which the responses from the same group, i.e household and small community, could not be assumed independent after conditioning on the observed exogenous variables.

For example, at the household level, parents have some (unobserved) expectations about the expected return on education, which will influence the school decision in the same way for all their children. At the small community level, children live in households within the same area that shares the same (unobserved)

access to schools with the same (unobserved) characteristics. To model this, we used a random intercept model with random intercepts at the household and community levels.

Each child i (i = 1,…, n_{jk}) belongs to a household j (j= 1,…, n_k) nested within a small community k (k = 1,…, K). The equation for the decision to send the child to school can be stated in terms of a latent response model as follows:

$$y^*_{ijk} = x_{ijk}\beta + \mu_{jk} + u_k + \varepsilon_{ijk}$$

where

$y_{ijk} = 1$ if $y^*_{ijk} > 0$

$y_{ijk} = 0$ otherwise

where y^*_{ijk} is the expected net benefit from the education of child i, who lives in household j in small community k. If the expected net benefit is positive, the child is sent to school, and we observe y_{ijk} = 1; otherwise y_{ijk} = 0. The observed characteristics of child i who lives in household j in small community k are denoted by x_{ijk}. The household-level (second-level) random intercept is $\mu_{jk}^{(2)}$ and the small community (third-level) random intercept is $\mu_k^{(3)}$. The individual-level error terms, ε_{ijk}, have a standard normal distribution with mean 0 and variance 1 and are independent of u_{jk} and μ_k. The random intercepts are assumed to be normally distributed with mean 0 and variance σ_2^2 and σ_3^2. The random intercepts are not directly estimated as model parameters but instead are summarized according to their variances (StataCorp 2013).

Parameter estimation is based on full-information maximum likelihood estimation using a mean-variance adaptive Gauss-Hermite quadrature (using seven integrating points) to approximate multiple random effects.

We also re-estimated the parameters using (1) a probit model without controlling for the correlation of the error terms, (2) a probit model clustering the error terms at the household level, and (3) a probit model clustering the error terms at the community level.

5 Results

We estimated several model specifications based on the hypotheses presented in Section 2. The estimation was performed for all children and then separately for boys and girls.

5.1 Traditional variables

Table 1 presents the estimated parameters for the traditional variables for child, household and small community characteristics only (Column 1), and their values when we extend the model specification to include ethnolinguistic background measured as the poverty rate, the share of homemakers, the quality of education and the cost of education by language group (Column 2) or dummies for language groups (Column 3). The last column (Column 4) presents the estimates for a probit model with the same variable specification as that reported in Column 3 but ignoring the multilevel structure of the data. We report the estimates for the language dummies in Table 3 and for the mean values by language group in Table 4. The direction of the impact (i.e., the sign of the estimated parameters) is the same across all four columns, but the size of the parameters and the standard errors vary for most variables included in the analysis. The differences between the standard errors of the probit model and the three-level random intercept models suggest that there is some dependence of the school decision within households and at the household level within small communities. This finding is further supported by the fact that both the variance from the household and that from the small community are significant. A likelihood-ratio test reveals that there is enough variability among households and communities to favor a random intercept model over a traditional probit model.

Table 1 Probit and three-level random intercept probit model estimates (standard errors in parentheses), n=11138

	Three level random intercept probit model			Probit model
	(1)	(2)	(3)	(4)
Constant	-2.799 ***	-1.999 ***	-2.754 ***	-2.376 ***
	(0.435)	(0.496)	(0.468)	(0.349)
Age	0.889 ***	0.895 ***	0.893 ***	0.707 ***
	(0.080)	(0.080)	(0.080)	(0.064)
Age squared	-0.041 ***	-0.041 ***	-0.041 ***	-0.032 ***
	(0.004)	(0.004)	(0.004)	(0.003)
Boy	0.126 ***	0.121 **	0.120 **	0.071 *
	(0.048)	(0.048)	(0.048)	(0.038)
Chronically ill	-0.437 ***	-0.477 ***	-0.483 ***	-0.359 ***
	(0.132)	(0.132)	(0.131)	(0.107)
Head of household's education (Ref: no education)				
Primary education	0.505 ***	0.415 ***	0.414 ***	0.437 ***
	(0.072)	(0.072)	(0.071)	(0.048)
Secondary education or better	0.889 ***	0.778 ***	0.780 ***	0.749 ***
	(0.113)	(0.112)	(0.111)	(0.079)
Hard core poor	-0.263 ***	-0.265 ***	-0.262 ***	-0.180 ***
	(0.076)	(0.076)	(0.076)	(0.054)
Expenditure	0.884 **	0.920 **	0.938 **	0.776 ***
	(0.423)	(0.426)	(0.423)	(0.306)
Expenditure squared	-0.279	-0.276	-0.284	-0.233 *
	(0.185)	(0.187)	(0.186)	(0.130)
Owns land	0.289 ***	0.243 ***	0.211 ***	0.205 ***
	(0.072)	(0.071)	(0.071)	(0.046)
Head of household's religion (Ref: no religion)				
Catholic	0.368 ***	-0.027	0.361 ***	0.330 ***
	(0.125)	(0.021)	(0.124)	(0.080)
Protestant	0.320 ***	0.342 ***	0.317 ***	0.318 ***
	(0.120)	(0.123)	(0.119)	(0.078)
Other religion	0.276 **	0.295 **	0.269 **	0.225 **
	(0.137)	(0.118)	(0.137)	(0.090)
Muslim	0.004	0.229 *	0.337 *	0.292 ***
	(0.152)	(0.136)	(0.180)	(0.110)
Number of children aged 6-14	-0.036 *	0.208	-0.027	-0.026 *
	(0.021)	(0.168)	(0.021)	(0.014)
Mean values at district/community level				
Cost of education	-0.132 **	-0.031	-0.024	-0.031
	(0.062)	(0.068)	(0.067)	(0.039)
Distance to school	-0.393 ***	-0.329 **	-0.235	-0.103
	(0.144)	(0.144)	(0.160)	(0.088)
Poverty rate	-0.763 ***	-0.679 ***	-0.671 ***	-0.462 ***
	(0.166)	(0.166)	(0.165)	(0.088)
Child work rate	-1.200 ***	-1.262 ***	-1.328 ***	-1.328 ***
	(0.257)	(0.248)	(0.246)	(0.246)
Rate of homemakers	-0.807 ***	-0.302	-0.363 **	-0.363 **
	(0.176)	(0.185)	(0.182)	(0.182)
Quality of education	0.317 *	0.224	0.254	0.254
	(0.179)	(0.218)	(0.215)	(0.215)
Ethnolinguistic background				
Mean values by language group	No	Yes[a]	No	No
Language-group dummies	No	No	Yes[b]	Yes
Random effect variance				
Small community	0.412 ***	0.350 ***	0.326 ***	
	(0.062)	(0.056)	(0.053)	
Household (within community)	0.377 ***	0.369 ***	0.360 ***	
	(0.073)	(0.072)	(0.071)	
Log-likelihood	-2566	-2540	-2531	-2660

*** 1% significance level, ** 5% significance level, and *10% significance level; [a] the estimates are reported in Table 4; [b] the estimates are reported in Table 3.

All statistically significant estimates of traditional variables are consistent with the findings reported in earlier studies: the probability of being in school increases at a decreasing rate with the age of the child; boys have a higher probability of being in school than girls, and the probability that a child is in school decreases if he/she suffers from a chronic illness. The probability that the child is in school increases with the educational level of the head of the household, the household's ownership of land used for agriculture, and the household's income. Additionally, children living in religious households have a higher probability of being in school.

Children's probability of being in school decreases with the share of working children in the small community in which they are living, with the number of school-aged siblings, and when the household in which they are living is hard core poor. This last result support the hypothesis that, even when primary education is tuition free, budget constraints still hinder the poorest children from being in school.

5.2 Traditional variables and gender

Table 2 presents the estimates for the traditional variables, for girls and boys separately, when controlling for mean values by language group measured as the poverty rate, the share of homemakers, the quality of education and the cost of education by language group (1) and when controlling for language-group dummies (2). Most of the estimates show the same pattern as those reported in Table 1, but there are certain interesting exceptions. The magnitude of the estimated parameters for age, all religion dummies, and poverty at the district level are higher for boys than for girls. The positive estimate for the impact of household expenditure is only statistically significant for girls, suggesting that the household income level is more important for girls.

The parameter for the number of school-aged children is statistically significant and negative only for boys, suggesting a decrease in boys' probability of being in school when there are other school-aged children in the household.

The impact of the percentage of women in the district who are homemakers is statistically significant and negative for both boys and girls, which suggests that when the mother is at home the child has a lower probability of being in school. The impact is stronger for girls, supporting our hypothesis that gender norms are important.

Table 2 Three-level random intercept probit model estimates, by gender (standard error in parentheses)

	(1) Mean values by language group		(2) Language-group dummies	
	Girls n= 5517	Boys n = 5621	Girls n= 5517	Boys n = 5621
Constant	-0.888	-2.744 ***	-1.953 ***	-3.310 ***
	(0.703)	(0.660)	(0.672)	(0.638)
Age	0.725 ***	1.005 ***	0.723 ***	1.002 ***
	(0.119)	(0.120)	(0.119)	(0.120)
Age squared	-0.034 ***	-0.045 ***	-0.034 ***	-0.045 ***
	(0.006)	(0.006)	(0.006)	(0.006)
Chronically ill	-0.440 **	-0.491 **	-0.422 **	-0.509 ***
	(0.188)	(0.193)	(0.188)	(0.192)
Head of household's education (Ref: no education)				
Primary education	0.457 ***	0.450 ***	0.466 ***	0.452 ***
	(0.099)	(0.097)	(0.098)	(0.096)
Secondary or better	0.825 ***	0.819 ***	0.834 ***	0.829 ***
	(0.162)	(0.150)	(0.161)	(0.149)
Hard core poor	-0.245 **	-0.209 **	-0.231 **	-0.211 **
	(0.113)	(0.103)	(0.112)	(0.102)
Expenditure	2.499 **	0.423	2.510 ***	0.410
	(0.986)	(0.561)	(0.967)	(0.556)
Expenditure squared	-1.962 *	-0.127	-1.948 *	-0.128
	(1.094)	(0.233)	(1.063)	(0.234)
Owns land	0.316 ***	0.225 **	0.293 ***	0.174 *
	(0.094)	(0.093)	(0.094)	(0.093)
Number of children aged 6-14	0.007	-0.056 **	0.007	-0.056 **
	(0.028)	(0.028)	(0.028)	(0.028)
Head of household's religion (Ref: no religion)				
Catholic	0.270	0.409 ***	0.258	0.446 ***
	(0.169)	(0.155)	(0.172)	(0.157)
Protestant	0.277 *	0.336 **	0.276	0.370 **
	(0.165)	(0.148)	(0.168)	(0.150)
Other religion	0.123	0.279	0.140	0.339 *
	(0.186)	(0.174)	(0.189)	(0.176)
Muslim	0.175	0.253	0.309	0.418 *
	(0.218)	(0.209)	(0.235)	(0.227)
Mean values at the district level				
Cost of education	-0.048	-0.039	-0.038	-0.034
	(0.091)	(0.078)	(0.090)	(0.076)
Distance to school	-0.430 **	-0.184	-0.291	-0.171
	(0.178)	(0.169)	(0.199)	(0.189)
Poverty rate	-0.504 **	-0.802 ***	-0.519 ***	-0.757 ***
	(0.202)	(0.200)	(0.200)	(0.198)
Child work rate	-1.703 ***	-1.446 ***	-1.817 ***	-1.445 ***
	(0.300)	(0.286)	(0.298)	(0.285)
Rate of homemakers	-0.430 *	-0.318	-0.484 **	-0.381 *
	(0.221)	(0.215)	(0.217)	(0.214)
Quality of education	0.277	0.182	0.299	0.215
	(0.265)	(0.255)	(0.261)	(0.253)
Ethnolinguistic background				
Means by language group	Yes [a]	Yes [a]	No	No
Language-group dummies	No	No	Yes [b]	Yes [b]
Random effect variance				
Small community	0.305 ***	0.264 ***	0.278 ***	0.246 ***
	(0.075)	(0.070)	(0.071)	(0.067)
Household within community	0.346 **	0.454 ***	0.324 **	0.434 ***
	(0.139)	(0.135)	(0.136)	(0.132)
Log-likelihood	-1229	-1342	-1222	-1334

Notes: *** 1% significance level, ** 5% significance level, and * 10% significance level; [a] the estimates are reported in Table 4; [b] the estimates are reported in Table 3.

5.3 Ethnolinguistic background

When the traditional variables specification is extended using ethnolinguistic background, measured as the poverty rate, the share of homemakers, the quality of education and the cost of education by language group (Column 2 in Table 1) or measured using the language dummies (Column 3 in Table 1), some estimates of the impacts of the traditional variables change in magnitude. Most affected are the estimates for the small community/district characteristics, which suggests that the effects are picked up by the mean values of these variables across the language groups. The coefficients for the cost and quality of education, and for the rate of homemakers become insignificant (Column 2 in Table 1), but the coefficients are statistically significant for the coefficients of the mean values of these variables by language group (Column 1 in Table 4). Interestingly, the quality of education at the district level increases the child's probability of being school, but the average level of the quality within the language group decreases the child's probability of being in school. This might suggest that children from particular ethnolinguistic groups might live in districts with good schools, but their households might still face other constraints (for example, the language of instruction) that decrease the child's probability of being in school.

Compared with the model that includes only the traditional variables, both model specifications that control for ethnolinguistic background (Columns 2 and 3 in Table 1) lead to a decrease in the random intercepts' variance at the small community and household levels, which suggests the importance of ethnolinguistic background in explaining the child's probability of being in school.

In the model specification that includes the language dummies (Column 3 in Table 1), we used Kikuyu, the largest ethnolinguistic group in Kenya, as the reference group, which allowed us to identify potentially disadvantaged groups. Estimates for the impacts of the language groups' dummies, for all children and for girls and boys separately, are presented in Table 3. We also tested how the definition of our dependent variable influenced our results, by changing it to an indicator showing whether the child had ever attended school.

Compared to Kikuyu children, children from the Maasai and Somali groups show a lower probability of being in school. The estimated coefficient (-1.347) shows the impact to be stronger than, for example, the impact of having a head of household with some secondary education (0.819). These results hold for boys and girls separately, with higher values for girls. Moreover, compared with Kikuyu boys, Kamba boys have a higher probability of being in school. For girls, more estimates are statistically significant, suggesting that there might be gender norms that are picked up in the estimates. Girls belonging to the Maasai, Meru, Mijikenda, Somali, Swahili-coast and English language groups have a lower probability of being in school than girls from the Kikuyu group. This finding is in line with Tas et al. (2014), who show that ethnicity-based differences are more pronounced for females than males. Re-estimating the same model specification with different reference groups, we find that Maasai and Somali children have a lower probability of being in school than children from all other ethnolinguistic groups (appendix, Table A3).

Changing the dependent variable to an indicator for whether a child has ever attended school shows ethnolinguistic background to be even more important in determining whether or not a child gets any education at all (Table3, column 4). The estimated coefficients for Kalenjin and Mijikenda become statistically significant (negative) and for all other groups the estimated coefficients increase in magnitude.

The estimates for the impacts of the two official languages, Swahili and English, are negative and statistically significant. The estimate for the impact of English might be attributable to the girls in that group. Because most children who lived in a household that answered in English lived in the Moyale district, it appears that girls may be particularly disadvantaged in that location. As a robustness check, we re-estimated our model both by combining the English and Swahili groups into one group and by excluding these groups. This did not change the statistical significance or the direction of the estimated coefficients for the other language groups.

Table 3 Estimates for ethnolinguistic background not reported in Tables 1 and 2 (standard errors in parentheses)

	Enrolled in school 2005/2006						Enrolled in school 2005/2006 and/or earlier	
	Girls		Boys		All children		All children	
	n = 5517		n = 5621		n = 11138		n = 11138	
Language-group dummies (Ref: Kikuyu)								
Embu	-0.030		0.060		-0.032		-0.365	
	(0.400)		(0.388)		(0.299)		(0.351)	
Kamba	0.163		0.461	*	0.315		0.146	
	(0.281)		(0.261)		(0.214)		(0.271)	
Kalenjin	-0.280		-0.113		-0.185		-0.473	**
	(0.240)		(0.217)		(0.180)		(0.228)	
Kisii	0.157		-0.144		-0.063		-0.474	
	(0.373)		(0.283)		(0.250)		(0.293)	
Luhya	-0.190		0.145		-0.072		-0.328	
	(0.319)		(0.310)		(0.242)		(0.292)	
Luo	-0.248		0.236		-0.038		-0.313	
	(0.254)		(0.247)		(0.194)		(0.243)	
Maasai	-1.148	***	-0.804	***	-1.074	***	-1.536	***
	(0.291)		(0.276)		(0.233)		(0.282)	
Meru	-0.587	*	-0.318		-0.427	*	-0.807	***
	(0.302)		(0.296)		(0.234)		(0.288)	
Mijikenda	-0.726	**	0.060		-0.355		-0.689	**
	(0.337)		(0.349)		(0.280)		(0.322)	
Somali	-1.687	***	-0.943	***	-1.347	***	-1.842	***
	(0.336)		(0.311)		(0.263)		(0.307)	
Swahili-coast	-0.873	***	-0.410		-0.655	***	-1.049	***
	(0.284)		(0.264)		(0.218)		(0.263)	
Swahili	-0.303		-0.230		-0.293	*	-0.601	***
	(0.225)		(0.201)		(0.166)		(0.214)	
English	-0.620	*	-0.315		-0.505	*	-0.902	***
	(0.345)		(0.331)		(0.280)		(0.321)	
Characteristics by level								
Child	Yes		Yes		Yes		Yes	
Household	Yes		Yes		Yes		Yes	
Community or district	Yes		Yes		Yes		Yes	

Notes: *** 1% significance level, ** 5% significance level, and * 10% significance level. All specifications includes controls for characteristics at Child, Household and Community/district levels.

Even though children are supposed to start school when they are six year old, this might vary among groups. To examine how our choice of age range for our sample might have influenced the results, we ran the model using several different age groups. Furthermore, the choice of age interval might affect the probability of being in school differently depending on whether the child was of school age before and/or after 2003 when primary school became free of charge. We controlled for this using a dummy variable showing whether the child had turned six after the reform (appendix, Table A4). Using different age samples, we found that the impact for the Swahili and English groups seemed to be due to the lower age cohorts, and did not hold when these cohorts were dropped. When using a sample of children aged 9-14, (or with a higher lower limit), we found that even children from Kiisi and Luhya had a higher probability of being in school than children from the Kikuyu group. Regardless of which age sample we used, Maasai and Somali children always had a lower probability of being in school than children from the Kikuyu group. Including the dummy for being of school age after the reform did not have any significant impact on the main results.

To investigate which characteristics of the language groups that are important for the child's probability to be in school we change our model specification using the rate of homemakers, the poverty level and the cost of education for each language group instead of language dummies. Their estimates are reported separately in Table 4. The child's probability of being in school decreases with the rate of homemakers, the poverty level and the cost of education within the language group. The estimated impacts of the rate

of homemakers and the poverty level are stronger for girls than for boys, suggesting that gender norms and poverty might lead households to withdraw girls from school before boys.

Table 4 Estimates for ethnolinguistic background, mean values (standard errors in parentheses)

	Three-level, random intercept probit model		
	All children n = 11138	Girls n = 5517	Boys n =5621
Mean value by language group			
Poverty rate	-1.367 *	-2.076 **	-1.142
	(0.707)	(0.872)	(0.841)
Rate of homemakers	-1.370 ***	-1.542 ***	-0.852 *
	(0.423)	(0.524)	(0.510)
Quality of education	0.319	0.125	0.483
	(0.394)	(0.500)	(0.480)
Cost of education	-0.957 ***	-1.062 ***	-0.968 ***
	(0.268)	(0.331)	(0.328)
Characteristics by level			
Child	Yes	Yes	Yes
Household	Yes	Yes	Yes
Community or district	Yes	Yes	Yes

Notes: *** 1% significance level, ** 5% significance level, and * 10% significance level.

To further understand what explains the differences among the ethnolinguistic groups, Table 5 shows how the magnitude of the language dummies (Column 1) change when controlling for additional variables: traditional variables for children (Column 2), households (Column 3), small community and district characteristics (Column 4), interactions between language group and gender (Column 5), interactions between gender and religion (Column 6), and interactions between gender and rate of homemakers (Column 7).

Our estimates suggest that differences among language groups are explained to some extent by traditional variables. The estimated coefficients for most of the language groups (i.e., Kalenjin, Luhya, Maasai, Meru, Mijikenda, Somali, Swahili-coast, Swahili and English) are both negative and significant, indicating that children not belonging to these groups are less likely to be in school than a Kikuyu child (Column 1). However, the magnitude varies across language groups, with the largest values found in the Somali and Maasai groups. When adding characteristics of the children (Column 2), the magnitude of the impact of ethnolinguistic background becomes slightly stronger, except in the Luhya group, where the level of significance decreases. Adding household characteristics (Column 3) decreases the differences across language groups, and the estimate for the Luhya group becomes insignificant, suggesting that at least part of the differences in the probability of being in school is explained by differences in household characteristics across language groups. For example, the poverty level in the Luhya group is significantly higher than that in the Kikuyu group. In the next step, we extend the model specification by adding neighborhood and district characteristics (Column 4), which affects some of the estimates of the language groups, which decrease in magnitude, and the estimate for Kalenjin becomes insignificant. This finding suggests that the characteristics of the neighborhood and district in which language groups live affect the probability that a child will be in school. However, estimates for the other language-group dummies remain statistically significant, which implies that they continue to contain information that influences a child's probability of going to school. Therefore, in the next step (Column 5), we explore the existence of gender norms by adding interaction variables between the language-group dummies and the gender of the child (being a boy). All the coefficients for the language groups remain negative, but they increase in magnitude. Adding interaction with religion (Column 6) and the rate of homemakers (Column 7) has only a minor impact on the estimates of the language dummies, suggesting that there are some common characteristics within language groups, other than religion and gender norms, that drive a household's decision over whether to send a child to school.

Table 5 Three-level, random intercept probit model (standard errors in parentheses), n=11138

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Language group							
(Comparison group: Kikuyu)							
Embu	-0.333	-0.329	-0.158	-0.032	-0.126	-0.098	-0.093
	(0.314)	(0.341)	(0.311)	(0.299)	(0.398)	(0.400)	(0.401)
Kamba	-0.004	0.015	0.227	0.315	0.213	0.210	0.214
	(0.224)	(0.242)	(0.221)	(0.214)	(0.287)	(0.288)	(0.288)
Kalenjin	-0.572 ***	-0.618 ***	-0.355 *	-0.185	-0.308	-0.301	-0.296
	(0.184)	(0.199)	(0.182)	(0.180)	(0.239)	(0.241)	(0.241)
Kisii	-0.288	-0.264	-0.166	-0.063	0.120	0.135	0.134
	(0.269)	(0.291)	(0.262)	(0.250)	(0.371)	(0.374)	(0.374)
Luhya	-0.495 **	-0.499 *	-0.253	-0.072	-0.294	-0.295	-0.292
	(0.249)	(0.271)	(0.252)	(0.242)	(0.320)	(0.321)	(0.321)
Luo	-0.281	-0.278	-0.168	-0.038	-0.286	-0.286	-0.281
	(0.200)	(0.216)	(0.196)	(0.194)	(0.253)	(0.253)	(0.254)
Maasai	-1.847 ***	-1.999 ***	-1.451 ***	-1.074 ***	-1.318 ***	-1.313 ***	-1.296 ***
	(0.233)	(0.252)	(0.225)	(0.233)	(0.281)	(0.282)	(0.287)
Meru	-0.442 *	-0.490 *	-0.453 *	-0.427 *	-0.534 *	-0.530 *	-0.527 *
	(0.246)	(0.265)	(0.239)	(0.234)	(0.303)	(0.304)	(0.304)
Mijikenda	-1.203 ***	-1.254 ***	-0.510 *	-0.355	-0.686 **	-0.641 *	-0.642 *
	(0.294)	(0.320)	(0.295)	(0.280)	(0.331)	(0.340)	(0.340)
Somali	-2.494 ***	-2.691 ***	-1.635 ***	-1.347 ***	-1.837 ***	-1.741 ***	-1.730 ***
	(0.230)	(0.250)	(0.270)	(0.263)	(0.306)	(0.328)	(0.330)
Swahili-coast	-1.389 ***	-1.494 ***	-0.870 ***	-0.655 ***	-1.024 ***	-0.958 ***	-0.947 ***
	(0.211)	(0.228)	(0.224)	(0.218)	(0.269)	(0.282)	(0.284)
Swahili	-0.766 ***	-0.819 ***	-0.536 ***	-0.293 *	-0.398 *	-0.378 *	-0.373 *
	(0.168)	(0.181)	(0.166)	(0.166)	(0.222)	(0.223)	(0.224)
English	-1.380 ***	-1.543 ***	-0.880 ***	-0.505 *	-0.841 **	-0.750 **	-0.738 **
	(0.265)	(0.287)	(0.292)	(0.280)	(0.328)	(0.345)	(0.347)
Characteristics, by level							
Child	No	Yes	Yes	Yes	Yes	Yes	Yes
Household	No	No	Yes	Yes	Yes	Yes	Yes
Community or district	No	No	No	Yes	Yes	Yes	Yes
Gender interactions							
Boy*Language	No	No	No	No	Yes	Yes	Yes
Boy*Religion	No	No	No	No	No	Yes	Yes
Boy*Home-maker	No	No	No	No	No	No	Yes
Random effect variance							
Community	0.626 ***	0.726 ***	0.457 ***	0.326 ***	0.331 ***	0.333 ***	0.333 ***
	(0.076)	(0.090)	(0.065)	(0.053)	(0.055)	(0.055)	(0.055)
Household	0.283 ***	0.387 ***	0.371 ***	0.360 ***	0.384 ***	0.387 ***	0.386 ***
	(0.060)	(0.073)	(0.073)	(0.071)	(0.074)	(0.075)	(0.075)
Log-likelihood	-2767	-2653	-2559	-2531	-2513	-2511	-2511

Notes: *** 1% significance level, ** 5% significance level, and * 10% significance level.

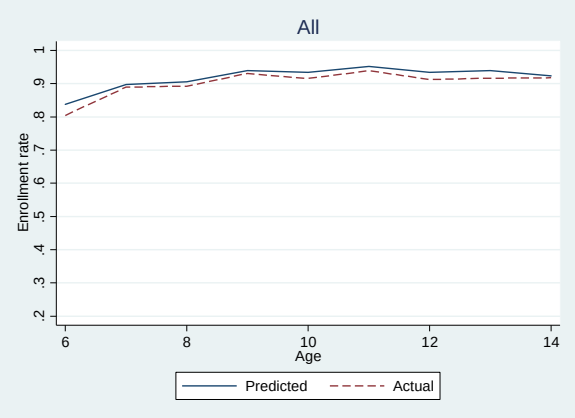
Table 6 further investigates the interaction between gender and the language groups. We begin with a model specification that includes only interactions between gender and the language group (Column 1). Then we add the traditional variables for the child, household, and small community and district characteristics (Column 2) and the interaction with religion (3). When we add the interaction between gender and the language group, the coefficient for being a boy changes its sign and becomes insignificant. When all the traditional controls are included (Column 2), only the interaction with Maasai and Somali are statistically significant for boys. The coefficients for the interactions with these groups, together with Meru, Mijikenda, Swahili-coast, Swahili and English, are also negative and statistically significant for the interaction with being a girl. The coefficients for Maasai and Somali are statistically significantly stronger for the interaction with being a girl, which suggests that the girls in these groups are less likely to be in school. Thus, gender norms might vary not only among language groups but also among religions. To control for this possibility, we include interactions for gender and religion (Column 3), but those interactions do not change our results for the language groups.

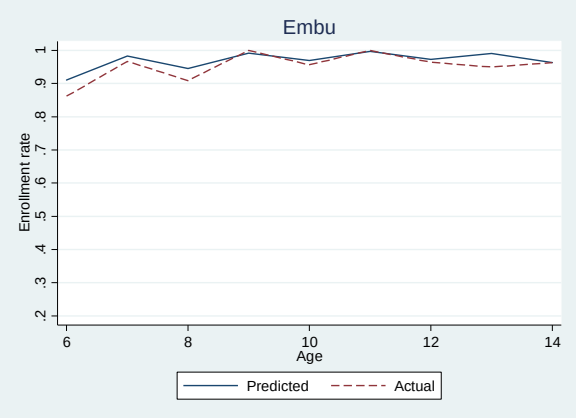
Table 6 Three-level, random intercept probit model estimates (standard errors in parentheses), ender interactions

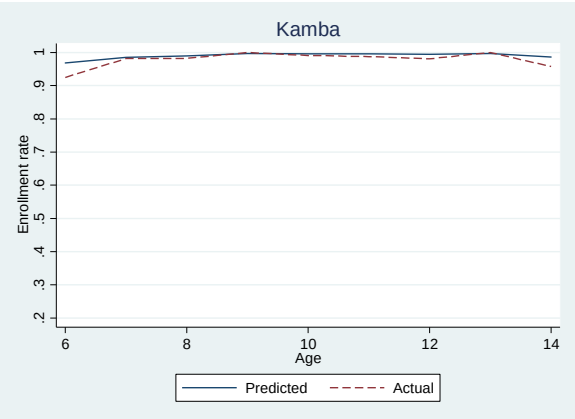
	(1)		(2)		(3)	
	Boys	Girls	Boys	Girls	Boys	Girls
Language-gender interactions						
Embu	-0.244	-0.462	0.050	-0.126	0.034	-0.098
	(0.400)	(0.397)	(0.385)	(0.398)	(0.386)	(0.400)
Kalenjin	-0.504 **	-0.683 ***	-0.088	-0.308	-0.093	-0.301
	(0.216)	(0.241)	(0.216)	(0.239)	(0.217)	(0.241)
Kamba	0.090	-0.130	0.406	0.213	0.405	0.210
	(0.267)	(0.291)	(0.260)	(0.287)	(0.261)	(0.288)
Kisii	-0.337	-0.132	-0.092	0.120	-0.104	0.135
	(0.303)	(0.374)	(0.289)	(0.371)	(0.289)	(0.374)
Luhya	-0.256	-0.783 **	0.129	-0.294	0.130	-0.295
	(0.310)	(0.320)	(0.307)	(0.320)	(0.308)	(0.321)
Luo	-0.071	-0.524 **	0.203	-0.286	0.203	-0.286
	(0.245)	(0.258)	(0.243)	(0.253)	(0.244)	(0.253)
Maasai	-1.665 ***	-2.107 ***	-0.852 ***	-1.318 ***	-0.872 ***	-1.313 ***
	(0.265)	(0.283)	(0.267)	(0.281)	(0.268)	(0.282)
Meru	-0.374	-0.561 *	-0.361	-0.534 *	-0.360	-0.530 *
	(0.296)	(0.312)	(0.291)	(0.303)	(0.291)	(0.304)
Mijikenda	-0.916 ***	-1.515 ***	-0.026	-0.686 **	-0.062	-0.641 *
	(0.341)	(0.345)	(0.337)	(0.331)	(0.342)	(0.340)
Somali	-2.078 ***	-3.016 ***	-0.921 ***	-1.837 ***	-1.009 ***	-1.741 ***
	(0.256)	(0.280)	(0.290)	(0.306)	(0.310)	(0.328)
Swahili	-0.690 ***	-0.890 ***	-0.203	-0.398 *	-0.220	-0.378 *
	(0.197)	(0.224)	(0.198)	(0.222)	(0.199)	(0.223)
Swahili-coast	-1.066 ***	-1.774 ***	-0.318	-1.024 ***	-0.371	-0.958 ***
	(0.245)	(0.264)	(0.253)	(0.269)	(0.263)	(0.282)
English	-1.101 ***	-1.715 ***	-0.193	-0.841 **	-0.277	-0.750 **
	(0.302)	(0.316)	(0.319)	(0.328)	(0.335)	(0.345)
Male	-0.315		-0.297		-0.280	
	(0.225)		(0.231)		(0.297)	
Other characteristics						
Child	No		Yes		Yes	
Household	No		Yes		Yes	
Community/district	No		Yes		Yes	
Boys x religion	No		No		Yes	
Girls x religion	No		No		Yes	
Random effect variance						
Community	0.644 ***		0.331 ***		0.333 ***	
	(0.079)		(0.055)		(0.055)	
Household	0.314 ***		0.384 ***		0.387 ***	
	(0.064)		(0.074)		(0.075)	
Log-likelihood	-2744		-2513		-2511	

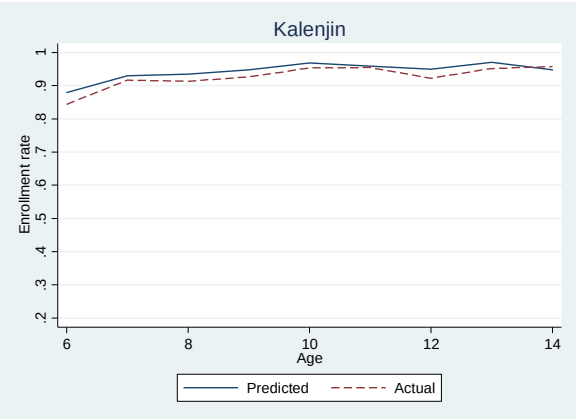
Notes: *** 1% significance level, ** 5% significance level, and * 10% significance level.

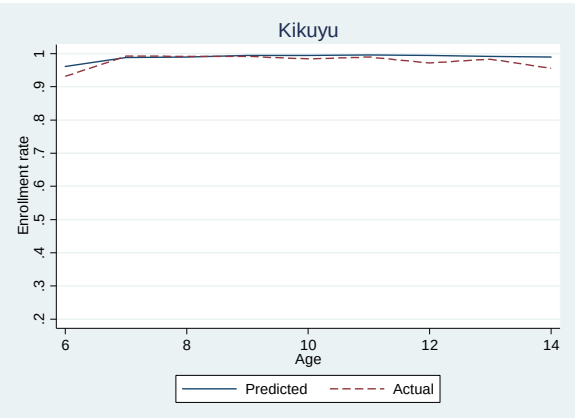
Figure 2 shows the model's power in predicting the rate of children who are in school across language groups, and displays the predicted and actual rates from specification 3 (Table 1) by age and language group. Generally, our model appears to accurately predict whether a child will be in school or not. The predicted probability that a six-year-old Somali child will be in school is approximately 30 percent, whereas the probability for a six-year-old Kikuyu child is over 90 percent.

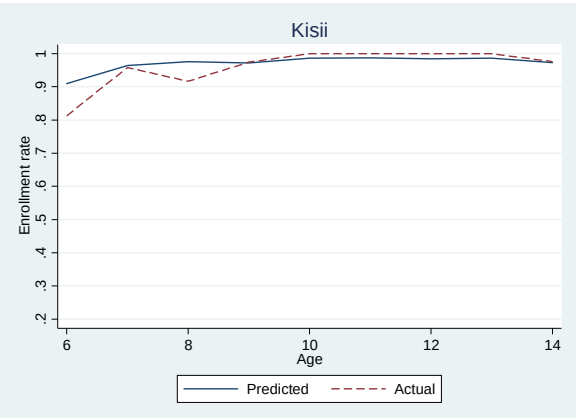

All
Enrollment rate
.2 .3 .4 .5 .6 .7 .8 .9 1
6 8 10 12 14
Age
Predicted
Actual

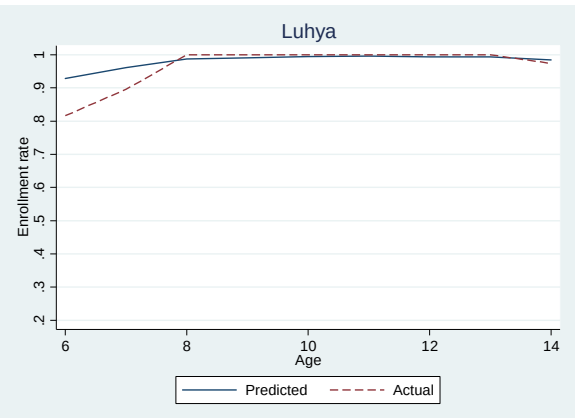

Embu
Enrollment rate
.2 .3 .4 .5 .6 .7 .8 .9 1
6 8 10 12 14
Age
Predicted
Actual

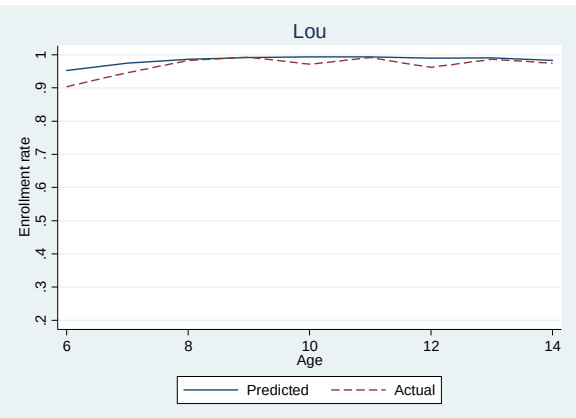

Kamba
Enrollment rate
.2 .3 .4 .5 .6 .7 .8 .9 1
6 8 10 12 14
Age
Predicted
Actual


Kalenjin
Enrollment rate
.2 .3 .4 .5 .6 .7 .8 .9 1
6 8 10 12 14
Age
Predicted
Actual


Kikuyu
Enrollment rate
.2 .3 .4 .5 .6 .7 .8 .9 1
6 8 10 12 14
Age
Predicted
Actual


Kisii
Enrollment rate
.2 .3 .4 .5 .6 .7 .8 .9 1
6 8 10 12 14
Age
Predicted
Actual


Luhya
Enrollment rate
.2 .3 .4 .5 .6 .7 .8 .9 1
6 8 10 12 14
Age
Predicted
Actual


Lou
Enrollment rate
.2 .3 .4 .5 .6 .7 .8 .9 1
6 8 10 12 14
Age
Predicted
Actual

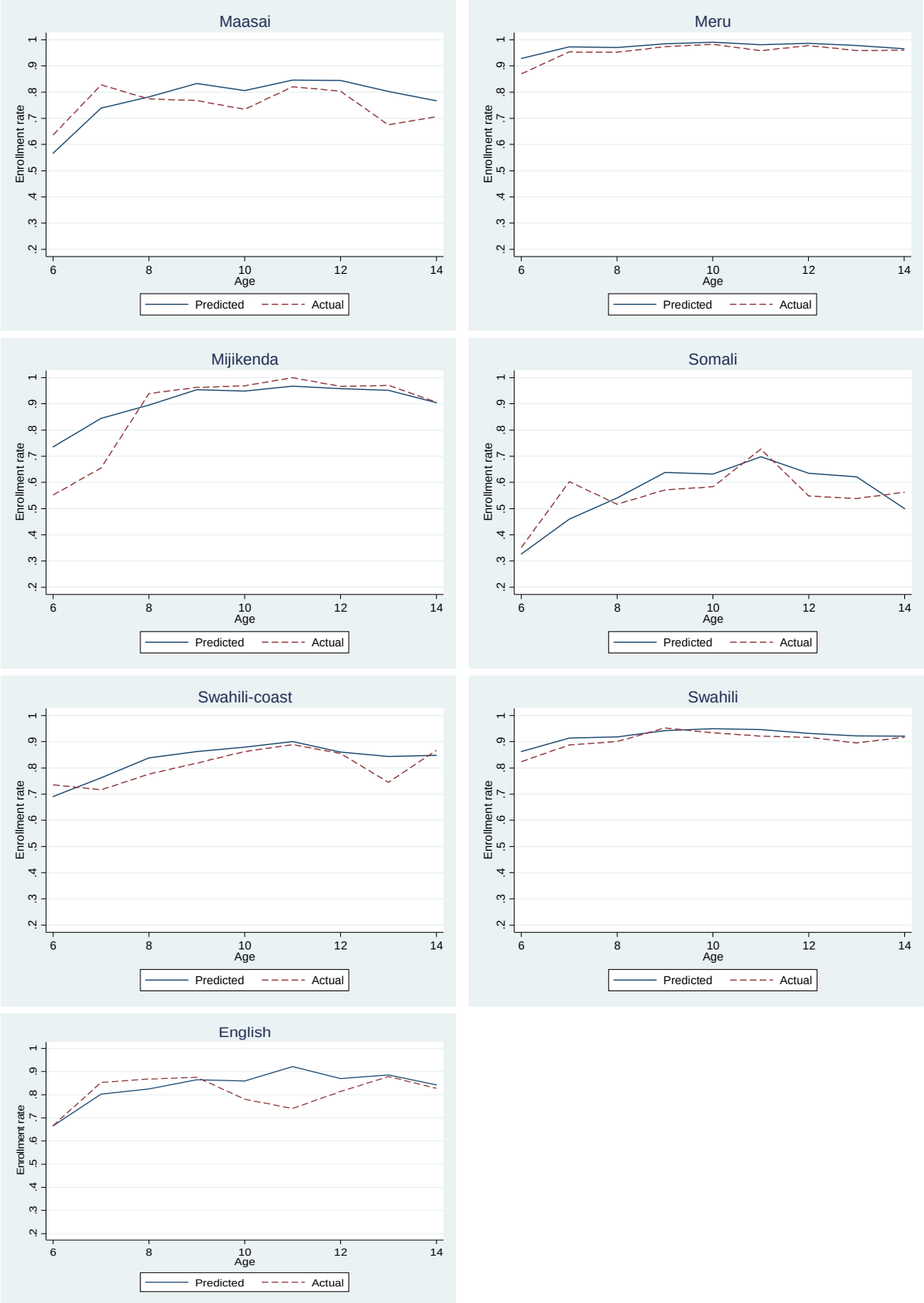


Figure 2 *Predicted enrollment, entire sample, and by language group*

5.4 Other robustness checks

To examine the importance of the statistical model, we also estimate the parameters by using (1) a probit model without controlling for the correlation of the error terms, (2) a probit model clustering the error term at the household level, and (3) a probit model clustering the error term at the community level. Table A5 in the appendix reports the estimates for the impacts of ethnolinguistic background on the child's probability of being in school, for these models. The advantage of the standard probit model is that it requires fewer assumptions about the distribution of the error term. However, it comes at the cost of less efficient estimates for the coefficients. Even though all statistical models give the same overall conclusions (the same variables are statistically significant and they have the same signs), the magnitudes of the coefficients change. To give a feeling for the sizes of the impacts, we calculate the marginal effects using the probit model, clustering the standard error at the community level. With the warning that there could be some bias here, we can summarize some of these results as follows: compared to Kikuyu children, the probability of being in school is about 19 percentage points lower for Somali children and 13 percentage points lower for Maasai children.

6 Conclusions

This paper contributes to the literature of children's enrollment in primary education by providing empirical evidence of the importance of ethnolinguistic background for the probability of being in school. We identified the language groups using the Kenya Integrated Household Budget Survey, which was designed to allow the respondent to answer in one of the eleven local languages, or Swahili or English, the latter two being Kenya's official languages. We expect that this variable contains several attributes, such as culture, norms and language that are essential to the decision over whether to send a child to school. Answering in a local language implies that the household uses that language at home, which makes it a good proxy for ethnolinguistic background.

Our results show that ethnolinguistic background influences a child's probability of being in school. Even after controlling for child, household, community and district characteristics, we find that ethnolinguistic background has a statistically significant impact, which supports our hypothesis that differences in culture and norms among language groups influence the expected costs and benefits of education.

However, we are aware that other factors could explain the differences among language groups. For example, the expected benefits of an education will be influenced by the different probability of children from different language groups being able to obtain an education in their mother tongue. Although we do not have data regarding the language of instruction at a household's nearest school, the estimates of the language group dummies leave some room for interpretation that children who do not speak the language of instruction may be less likely to be in school. This issue requires more and higher-quality data, which might support earlier policy recommendations that children should initially be instructed in either English or Swahili in before starting primary education.

Even though our data do not allow us to explicitly determine which factors are behind the differences among ethnolinguistic groups, we do find some evidence of potential mechanisms. For example, in line with previous research, we find that differences among ethnolinguistic groups are most pronounced for girls, indicating that the differences among groups can to some extent be explained by differences in gender norms. This finding is further supported by the result that, in language groups where a large proportion of the women stated that they took care of the household as their main activity, the child's probability of being in school was lower.

Somali and Maasai children are less likely to be in school than are children from all other groups. Both the Somali and Maasai groups have nomadic traditions, which might be a factor that influences the perceived costs and benefits of education. A separate analysis by child's gender shows that compared to Kikuyu children both girls and boys from the Somali and Maasai group, but also Mijikenda and Swahili girls, have a lower probability to be in school. This might be an indication that gender norms are stronger in these groups.

Therefore, one policy recommendation would be to focus on girls in these groups by for example conditional cash transfers in the areas where these groups live.

References

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. (2003). Fractionalization. Journal of Economic Growth, 8, 155-194.

Becker, G. S. (1975). Human capital: a theoretical and empirical analysis with special reference to education (2nd ed). National Bureau of Economic Research, New York.

Bold, T., Sandefur, J., Mwabu, G. and Kimenyi, M. (2010). Determinants of Acivement in Kenya since the Introduction of FPE. In Adam, C., Collier, P. and Ndung'u, N. Kenya: Policies for Propensity. Oxford University Press.

Bold, T., Kimenyi, M., Mwabu, G. and Sandefur, J. (2011). Why Did Abolishing Fees Not Increase Public School Enrolment in Kenya? Working Paper 271, Center for Global Development.

Brock-Utne, B. (2001). Education for All - in Whose Language? Oxford Review of Education, 27(March, 1), 115-134.

Buchmann, Claudia (2000). Family structure, parental perception, and child labor in Kenya: What factors determine who is enrolled in school? Social Forces, 78(4), 1349-1378.

Cleghorn, A., Merritt, M. and Abagi, J. O. (1989). Language policy and science instruction in Kenyan primary schools. Comparative Education Review, 33(1), 21-39.

Fernández and Fogli (2005). Culture: An Empirical Investigation of Beliefs, Work, and Fertility. NBER Working Paper No 11268.

Fernández, Raquel (2010). Does Culture Matter? Draft for the Handbook of Social Economics.

Franck, R. and Rainer, I. (2012). Does the leader's ethnicity matter? Ethnic favoritism, education and health in Sub-Saharan Africa. Working paper, Bar-Ilan University, No. 2012-06.

Glewwe, P., Kremer, M. and Mouline, S. (2009). Many Children Left Behind? Textbooks and Test Scores in Kenya. American Economic Journal: Applied Economics, 1(1), 112-135.

Global Partnership for Education (2012). Results for Learning Report 2012: Fostering Evidence-Based Dialogue to Monitor Access and Quality in Education. (November). Washington, DC: GPE.

Kabubo-Mariara, J. and Mwabu, D. (2007). Determinants of School Enrollment and Education Attainment: Empirical Evidence from Kenya. South African Journal of Economics, 75(3), 572-593.

Kenya National Bureau of Statistics (2010). The 2009 Kenya Population and Housing Census, Vol 2. Population and Household Distribution by Socio-Economic Characteristics.

Kenya National Bureau of Statistics (2007). Basic Report on Well-being – KIHBS 2005/2006.

Kimenyi, M. S. (1997). Ethnic Diversity, Liberty and the State: The African Dilemma. Edward Elgar Press, Cheltenham.

Kramon, E. J. and Posner, D. N. (2012). Ethnic Favoritism in Primary Education in Kenya. Unpublished manuscript.

Lewis, M. and Lockhead, M. (2007). Inexcusable absence: Why 60 million girls still aren't in school and what to do about it. Centre for Global Development, Brookings Institute Press.

Lewis, P., Simons, G. and Fenning, C. (2013). Ethnologue: Languages of the World (17th ed). SIL International, Dallas, Texas. Online version: http://www.ethnologue.com.

Lockheed, M. E. and Verspoor, A. M. (1991). Improving primary education in developing countries. New York: Oxford University Press.

Lucas, A. M . and Mbiti, I. (2012). Access, Sorting, and Achivement: The short run Effect of Primary Education in Kenya. American Economic Journal: Applied Economics, 4(4), 226-253.

McEwan, P. and Trowbridge, M. (2007). The achievement of indigenous students in Guatemala primary schools. International Journal of Educational Development, 27, 61-76.

Mincer, J. (1958). Investment, human capital and personal income distribution. Journal of Political Economy, 66(1), 281-321.

Mincer, J. (1974). Schooling Experience and Earnings. National Bureau of Economic Research, New York.

Mukudi, E. (2004). Education for all: a framework for addressing the persisting illusion for the Kenyan context. Internation Journal of Educational Development, 24(3), 231-240.

Nishimura, M. and Yamano, T. (2013). Emerging Private Education in Africa: Determinants of School Choice in Rural Kenya. World Development, 43, 266-275.

Nyokabi, J. (2009). Determinants of primary schooling in Kenya. KIPPRA discussion paper 104.

Obanya, Pai (1980). General methods of teaching progress at their own pace. Macmillan Nigeria Publishers Ltd, Lagos.

Ogle, Mukhtar, Wasanga, Paul and Wambua, Richard (2010). The report on monitoring learner achivenment study for class 3 in literacy and numeracy. The National Assessment Centre.

Omwami, M. E. and Omwami, R. K. (2010). Public investment and the goal of providing universal access to primary education by 2015 in Kenya. International Journal of Educational Development, 30(3), 243-253.

Patrinos, H. and Psacharopoulos, G. (1996). Socioeconomic and ethnic determinants of age-grade distortion in Bolivian and Guatemalan primary schools. International Journal of Educational Development, 16(1), 3-14.

Schultz, T. W. (1960). Capital Formation by Education. Journal of Political Economy, 68(12), 571-583.

Schultz, T.W. (1963). The Economic Value of Education. Columbia University Press, New York.

StataCorp. (2013). Stata: Release 13. Statistical Software. StataCorp LP, College Station, Texas.

Tas, E., Reimão, M. and Orlando, M. (2014). Gender, Ethnicity, and Cumulative Disadvantage in Educational Outcomes. World Development, 64, 538-553.

Trudell, B. (2005). Language choice, education and community identity. International Journal of Educational Development, 25, 237-251.

UNDP (2014). The Millennium Development Goals report 2014. United nations, New York.

UNESCO (2013). Institute for statistics database.

World Bank (2004). Kenya Strengthening the Foundation of Education and Training in Kenya: Opportunities and Challenges in Primary and General Secondary Education, Report No. 28064-KE.

Appendix

Table A1 Definition of variables

Variable	Definition
Language group dummmies	
Embu	Language of interview was Embu
Kamba	Language of interview was Kamba
Kalenjin	Language of interview was Kalenjin
Kikuyu	Language of interview was Kikuyu (ref)
Kisii	Language of interview was Kisii
Luhya	Language of interview was Luhya
Luo	Language of interview was Luo
Maasai	Language of interview was Maasai
Meru	Language of interview was Meru
Mijikenda	Language of interview was Mijikenda
Somali	Language of interview was Somali
Swahili-coast	Language of interview was Swahili, the individual lives in the coast province and all individuals in the household were born in the province.
Swahili	Language of interview was Swahili but not defined as Swahili-coast
English	Language of interview was English
Mean values by language group	
Poverty rate	Percentage of households defined as hard core poor (excluding household i). Calculated by language group.
Rate of homemaker	Percentage of women aged 18+ who stated being a homemaker as their main activity. Calculated by language group.
Quality of education	Quality of school measured as the percentage of the students in the second class who can write, calculated by language group.
Cost of education	Mean education cost in Kenyan Shillings for students in primary government school, divided by 1000. Calculated by language group.
Child characteristics	
Age	Age (in years)
Boy	The child is a boy
Chronically ill	Child suffers from chronic illness
Household characteristics	
Head no education	The head of household has no education (ref)
Head primary education	The head of household has primary education (and no more)
Head secondary educ.	The head of household has secondary education or more
Expenditure	Monthly per adult equivalent total household expenditure (divided by 10000) in Kenyan Shillings, in current prices year 2005/2006
Owns land	The household owns land used for agricultural purposes
Hard core poor	Household is defined as hard core poor if it could not meet its basic food needs even if it only consumed food
Catholic	The head of household is Catholic
Protestant	The head of household is Protestant
Other religion	The head of household is not Catholic, Protestant or Muslim
Muslim	The head of household is Muslim
No religion	The head of household does not have a religion (ref)

Community and/or district characteristics	
Quality of education	Quality of school measured as the percentage of the students in the second class who can write, calculated by district
Distance to school	Percentage of households that are more than 5 km from the nearest public primary school. Calculated by district.
Child work rate	Percentage of children aged 6-14 for whom work was stated as their main activity (excluding child i). Calculated by community.
Cost of education	Mean education cost in Kenyan Shillings for students in primary government school, divided by 1000. Calculated by community.
Poverty rate	Percentage of households defined as hard core poor (excluding household i). Calculated by community.
Rate of homemakers	Percentage of women aged 18+ who stated being a homemaker as their main activity. Calculated by community.

Table A2 Selected variables, mean and standard deviation (in parentheses), all children and by gender

	All Children			Girls			Boys		
	All n= 11138	Enrolled n= 10022	Not enrolled n= 1116	All n= 5517	Enrolled n= 4941	Not enrolled n= 576	All n= 5621	Enrolled n= 5081	Not enrolled n= 540
Enrolled	0.900	1.000	0.000	0.896	1.000	0.000	0.904	1.000	0.000
	(0.300)	(0.000)	(0.000)	(0.306)	(0.000)	(0.000)	(0.295)	(0.000)	(0.000)
Age	9.855	9.931	9.168	9.898	9.956	9.399	9.813	9.907	8.920
	(2.594)	(2.566)	(2.744)	(2.587)	(2.563)	(2.738)	(2.600)	(2.568)	(2.732)
Boy	0.505	0.507	0.484	0.000	0.000	0.000	1.000	1.000	1.000
	(0.500)	(0.500)	(0.500)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Chronically ill	0.030	0.030	0.032	0.033	0.033	0.033	0.028	0.027	0.031
	(0.171)	(0.170)	(0.177)	(0.178)	(0.178)	(0.179)	(0.164)	(0.163)	(0.175)
Head of household's education									
Primary education	0.476	0.503	0.234	0.486	0.518	0.208	0.467	0.489	0.261
	(0.499)	(0.500)	(0.423)	(0.500)	(0.500)	(0.406)	(0.499)	(0.500)	(0.440)
Secondary education	0.209	0.228	0.038	0.210	0.231	0.030	0.207	0.224	0.046
	(0.406)	(0.419)	(0.190)	(0.408)	(0.422)	(0.169)	(0.405)	(0.417)	(0.210)
Hard core poor	0.281	0.255	0.517	0.280	0.249	0.545	0.281	0.260	0.487
	(0.449)	(0.436)	(0.500)	(0.449)	(0.433)	(0.498)	(0.450)	(0.438)	(0.500)
Expenditure	0.173	0.179	0.120	0.172	0.179	0.113	0.173	0.178	0.127
	(0.139)	(0.141)	(0.100)	(0.127)	(0.128)	(0.092)	(0.150)	(0.153)	(0.107)
Own land	0.668	0.702	0.363	0.670	0.710	0.326	0.667	0.696	0.402
	(0.471)	(0.457)	(0.481)	(0.470)	(0.454)	(0.469)	(0.471)	(0.460)	(0.491)
Religion									
Catholic	0.258	0.265	0.189	0.261	0.270	0.184	0.255	0.261	0.194
	(0.437)	(0.442)	(0.392)	(0.439)	(0.444)	(0.388)	(0.436)	(0.439)	(0.396)
Protestant	0.420	0.443	0.211	0.420	0.448	0.184	0.419	0.438	0.239
	(0.494)	(0.497)	(0.408)	(0.494)	(0.497)	(0.388)	(0.493)	(0.496)	(0.427)
Other religion	0.123	0.124	0.109	0.126	0.128	0.108	0.119	0.120	0.111
	(0.328)	(0.330)	(0.312)	(0.332)	(0.334)	(0.310)	(0.324)	(0.325)	(0.315)
Muslim	0.132	0.104	0.391	0.130	0.094	0.439	0.135	0.113	0.339
	(0.339)	(0.305)	(0.488)	(0.336)	(0.292)	(0.497)	(0.341)	(0.317)	(0.474)
Number of children ages 6-14	2.785	2.751	3.094	2.802	2.765	3.113	2.769	2.736	3.074
	(1.333)	(1.309)	(1.493)	(1.347)	(1.317)	(1.549)	(1.319)	(1.302)	(1.432)
Characteristics at district/community level									
Child work rate	0.059	0.053	0.113	0.060	0.053	0.116	0.059	0.053	0.109
	(0.143)	(0.135)	(0.188)	(0.144)	(0.136)	(0.192)	(0.141)	(0.135)	(0.185)
Homemakers	0.260	0.239	0.443	0.258	0.235	0.461	0.261	0.244	0.425
	(0.256)	(0.240)	(0.310)	(0.252)	(0.235)	(0.304)	(0.259)	(0.246)	(0.317)
Quality of education	0.703	0.710	0.648	0.705	0.711	0.648	0.702	0.708	0.648
	(0.222)	(0.222)	(0.216)	(0.222)	(0.223)	(0.210)	(0.221)	(0.220)	(0.222)
Cost of education	0.569	0.572	0.542	0.562	0.566	0.527	0.576	0.578	0.557
	(0.591)	(0.601)	(0.481)	(0.565)	(0.577)	(0.444)	(0.615)	(0.624)	(0.517)
Distance to school	0.562	0.555	0.618	0.560	0.553	0.618	0.564	0.558	0.619
	(0.284)	(0.287)	(0.252)	(0.286)	(0.290)	(0.248)	(0.282)	(0.284)	(0.257)
Poverty rate	0.234	0.216	0.392	0.231	0.210	0.411	0.237	0.222	0.372
	(0.237)	(0.219)	(0.322)	(0.236)	(0.215)	(0.321)	(0.238)	(0.223)	(0.322)

Language group									
Embu	0.021	0.022	0.010	0.023	0.024	0.012	0.019	0.020	0.007
	(0.143)	(0.148)	(0.099)	(0.151)	(0.155)	(0.110)	(0.136)	(0.140)	(0.086)
Kamba	0.089	0.097	0.020	0.088	0.097	0.017	0.089	0.097	0.022
	(0.285)	(0.296)	(0.139)	(0.284)	(0.296)	(0.131)	(0.285)	(0.295)	(0.148)
Kalenjin	0.148	0.152	0.110	0.149	0.155	0.095	0.146	0.149	0.126
	(0.355)	(0.359)	(0.313)	(0.356)	(0.362)	(0.294)	(0.354)	(0.356)	(0.332)
Kisii	0.037	0.039	0.018	0.035	0.038	0.009	0.039	0.040	0.028
	(0.188)	(0.193)	(0.133)	(0.183)	(0.190)	(0.093)	(0.193)	(0.196)	(0.164)
Luhya	0.036	0.038	0.013	0.034	0.037	0.014	0.037	0.040	0.013
	(0.186)	(0.192)	(0.115)	(0.182)	(0.188)	(0.117)	(0.189)	(0.195)	(0.113)
Luo	0.101	0.109	0.032	0.103	0.111	0.035	0.100	0.107	0.030
	(0.302)	(0.312)	(0.177)	(0.304)	(0.314)	(0.183)	(0.300)	(0.309)	(0.170)
Maasai	0.042	0.035	0.106	0.042	0.035	0.099	0.043	0.035	0.113
	(0.201)	(0.184)	(0.308)	(0.200)	(0.184)	(0.299)	(0.202)	(0.184)	(0.317)
Meru	0.040	0.043	0.019	0.043	0.046	0.017	0.037	0.039	0.020
	(0.196)	(0.202)	(0.136)	(0.203)	(0.210)	(0.131)	(0.190)	(0.194)	(0.141)
Mijikenda	0.024	0.024	0.029	0.024	0.023	0.031	0.024	0.024	0.026
	(0.154)	(0.153)	(0.167)	(0.154)	(0.151)	(0.174)	(0.154)	(0.154)	(0.159)
Somali	0.052	0.032	0.237	0.051	0.025	0.273	0.054	0.039	0.198
	(0.223)	(0.175)	(0.425)	(0.219)	(0.155)	(0.446)	(0.226)	(0.193)	(0.399)
Swahili-coast	0.056	0.050	0.109	0.056	0.048	0.123	0.056	0.052	0.094
	(0.230)	(0.219)	(0.312)	(0.230)	(0.215)	(0.329)	(0.230)	(0.222)	(0.293)
Swahili-rest	0.229	0.229	0.221	0.224	0.226	0.203	0.233	0.232	0.241
	(0.420)	(0.421)	(0.415)	(0.417)	(0.419)	(0.403)	(0.423)	(0.422)	(0.428)
English	0.029	0.027	0.055	0.030	0.027	0.056	0.029	0.026	0.054
	(0.169)	(0.161)	(0.227)	(0.170)	(0.162)	(0.229)	(0.168)	(0.160)	(0.226)

Table A3 Three-level random intercept probit model estimates with different reference groups (standard errors in parentheses)

	Reference group						
	Embu	Kamba	Kikuyu	Kalenjin	Kisii	Luhya	Luo
Kamba	0.347		0.315	0.500 **	0.377	0.387	0.353
	(0.310)		(0.214)	(0.199)	(0.255)	(0.258)	(0.217)
Kikuyu	0.032	-0.315		0.185	0.063	0.072	0.038
	(0.299)	(0.214)		(0.180)	(0.250)	(0.242)	(0.194)
Kalenjin	-0.153	-0.500 **	-0.185		-0.123	-0.113	-0.147
	(0.291)	(0.199)	(0.180)		(0.239)	(0.222)	(0.167)
Kisii	-0.031	-0.377	-0.063	0.123		0.009	-0.024
	(0.338)	(0.255)	(0.250)	(0.239)		(0.289)	(0.254)
Luhya	-0.040	-0.387	-0.072	0.113	-0.009		-0.033
	(0.325)	(0.258)	(0.242)	(0.222)	(0.289)		(0.227)
Luo	-0.007	-0.353	-0.038	0.147	0.024	0.033	
	(0.290)	(0.217)	(0.194)	(0.167)	(0.254)	(0.227)	
Maasai	-1.042 ***	-1.389 ***	-1.074 ***	-0.889 ***	-1.011 ***	-1.002 ***	-1.035 ***
	(0.315)	(0.252)	(0.233)	(0.201)	(0.288)	(0.259)	(0.207)
Meru	-0.395	-0.742 ***	-0.427*	-0.242	-0.365	-0.356	-0.389*
	(0.316)	(0.246)	(0.234)	(0.227)	(0.281)	(0.270)	(0.227)
Mijikenda	-0.323	-0.670 **	-0.355	-0.170	-0.293	-0.283	-0.317
	(0.353)	(0.282)	(0.280)	(0.271)	(0.314)	(0.310)	(0.277)
Somali	-1.315 ***	-1.662 ***	-1.347 ***	-1.162 ***	-1.284 ***	-1.275 ***	-1.308 ***
	(0.343)	(0.261)	(0.263)	(0.245)	(0.299)	(0.296)	(0.261)
Swahili-coast	-0.623 **	-0.970 ***	-0.655 ***	-0.470 **	-0.592 **	-0.583 **	-0.616 ***
	(0.309)	(0.226)	(0.218)	(0.191)	(0.267)	(0.252)	(0.205)
Swahili-rest	-0.261	-0.608 ***	-0.293 *	-0.108	-0.230	-0.221	-0.255 *
	(0.277)	(0.187)	(0.166)	(0.123)	(0.227)	(0.204)	(0.146)
English	-0.473	-0.820 ***	-0.505 *	-0.320	-0.442	-0.433	-0.466 *
	(0.352)	(0.287)	(0.280)	(0.259)	(0.320)	(0.304)	(0.268)
Embu		-0.347	-0.032	0.153	0.031	0.040	0.007
		(0.310)	(0.299)	(0.291)	(0.338)	(0.325)	(0.290)

	Reference group						
	Maasai	Meru	Mijikenda	Somali	Swa-hili-coast	Swahili-rest	English
Kamba	1.389***	0.742***	0.670**	1.662***	0.970***	0.608***	0.820***
	(0.252)	(0.246)	(0.282)	(0.261)	(0.226)	(0.187)	(0.287)
Kikuyu	1.074***	0.427*	0.355	1.347***	0.655***	0.293*	0.505*
	(0.233)	(0.234)	(0.280)	(0.263)	(0.218)	(0.166)	(0.280)
Kalenjin	0.889***	0.242	0.170	1.162***	0.470**	0.108	0.320
	(0.201)	(0.227)	(0.271)	(0.245)	(0.191)	(0.123)	(0.259)
Kisii	1.011***	0.365	0.293	1.284***	0.592**	0.230	0.442
	(0.288)	(0.281)	(0.314)	(0.299)	(0.267)	(0.227)	(0.320)
Luhya	1.002***	0.356	0.283	1.275***	0.583**	0.221	0.433
	(0.259)	(0.270)	(0.310)	(0.296)	(0.252)	(0.204)	(0.304)
Luo	1.035***	0.389*	0.317	1.308***	0.616***	0.255*	0.466*
	(0.207)	(0.227)	(0.277)	(0.261)	(0.205)	(0.146)	(0.268)
Maasai		-0.646**	-0.719**	0.273	-0.419*	-0.781***	-0.569**
		(0.253)	(0.305)	(0.269)	(0.224)	(0.182)	(0.279)
Meru	0.646**		-0.072	0.919***	0.227	-0.134	0.077
	(0.253)		(0.296)	(0.284)	(0.247)	(0.209)	(0.302)
Mijikenda	0.719**	0.072		0.992***	0.300	-0.062	0.150
	(0.305)	(0.296)		(0.278)	(0.262)	(0.255)	(0.307)
Somali	-0.273	-0.919***	-0.992***		-0.692***	-1.054***	-0.842***
	(0.269)	(0.284)	(0.278)		(0.200)	(0.226)	(0.240)
Swahili-coast	0.419*	-0.227	-0.300	0.692***		-0.362**	-0.150
	(0.224)	(0.247)	(0.262)	(0.200)		(0.169)	(0.232)
Swahili-rest	0.781***	0.134	0.062	1.054***	0.362**		0.212
	(0.182)	(0.209)	(0.255)	(0.226)	(0.169)		(0.242)
English	0.569**	-0.077	-0.150	0.842***	0.150	-0.212	
	(0.279)	(0.302)	(0.307)	(0.240)	(0.232)	(0.242)	
Embu	1.042***	0.395	0.323	1.315***	0.623**	0.261	0.473
	(0.315)	(0.316)	(0.353)	(0.343)	(0.309)	(0.277)	(0.352)
Characteristics, by level							
Child	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Community or district	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** 1% significance level, ** 5% significance level, and * 10% significance level.

Table A4 Three-level random intercept probit model estimates with different age sample (standard errors in parentheses)

	Age 6-14 n = 11138	Age 6-14 n = 11138	Age 7-14 n = 9749	Age 8-14 n = 8472	Age 9-14 n = 7134	Age 10-14 n = 5919	Age 11-14 n = 4630
Embu	-0.033	-0.032	-0.028	0.007	0.257	0.191	0.097
	(0.299)	(0.299)	(0.351)	(0.379)	(0.413)	(0.460)	(0.525)
Kamba	0.314	0.315	0.396	0.420	0.481 *	0.429	0.446
	(0.213)	(0.214)	(0.253)	(0.266)	(0.274)	(0.303)	(0.345)
Kalenjin	-0.186	-0.185	-0.142	0.022	0.125	0.275	0.265
	(0.180)	(0.180)	(0.208)	(0.220)	(0.226)	(0.258)	(0.298)
Kisii	-0.062	-0.063	0.167	0.189	0.763 *	1.234 **	1.264 **
	(0.250)	(0.250)	(0.311)	(0.325)	(0.421)	(0.586)	(0.635)
Luhya	-0.073	-0.072	0.188	1.154 **	1.147 **	1.185 **	1.103 *
	(0.242)	(0.242)	(0.305)	(0.527)	(0.528)	(0.577)	(0.634)
Luo	-0.038	-0.038	0.017	0.203	0.320	0.333	0.363
	(0.194)	(0.194)	(0.227)	(0.244)	(0.248)	(0.279)	(0.327)
Maasai	-1.072 ***	-1.074 ***	-1.128 ***	-1.078 ***	-0.879 ***	-0.830 ***	-0.919 **
	(0.233)	(0.233)	(0.266)	(0.277)	(0.279)	(0.315)	(0.369)
Meru	-0.427 *	-0.427 *	-0.422	-0.416	-0.256	-0.287	-0.393
	(0.234)	(0.234)	(0.274)	(0.287)	(0.299)	(0.335)	(0.387)
Mijikenda	-0.354	-0.355	-0.279	0.144	0.280	0.300	0.277
	(0.280)	(0.280)	(0.319)	(0.357)	(0.363)	(0.412)	(0.471)
Somali	-1.345 ***	-1.347 ***	-1.369 ***	-1.383 ***	-1.125 ***	-1.217 ***	-1.059 **
	(0.263)	(0.263)	(0.296)	(0.315)	(0.321)	(0.368)	(0.425)
Swahili-coast	-0.654 ***	-0.655 ***	-0.766 ***	-0.651 **	-0.446 *	-0.400	-0.364
	(0.218)	(0.218)	(0.246)	(0.262)	(0.269)	(0.309)	(0.356)
Swahili- rest	-0.294 *	-0.293 *	-0.197	-0.000	0.180	0.222	0.178
	(0.166)	(0.166)	(0.193)	(0.206)	(0.213)	(0.240)	(0.276)
English	-0.504 *	-0.505 *	-0.542 *	-0.511	-0.414	-0.495	-0.385
	(0.280)	(0.280)	(0.311)	(0.323)	(0.327)	(0.369)	(0.428)
Reform	-0.055						
	(0.107)						
Characteristics, by level							
Child	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Community or district	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random effect variance							
Small community	0.326 ***	0.326 ***	0.380 ***	0.359 ***	0.309 ***	0.344 ***	0.372 ***
	(0.053)	(0.053)	(0.071	(0.077)	(0.078)	(0.098)	(0.131)
Household within community	0.359 ***	0.360 ***	0.418 ***	0.452 ***	0.401 ***	0.610 ***	0.822 ***
	(0.071)	(0.071)	(0.091)	(0.110)	(0.128)	(0.194)	(0.303
Log-likelihood		-2531	-1938	-1592	-1281	-1080	-840

Notes: *** 1% significance level, ** 5% significance level, and * 10% significance level.

Table A5 Robustness test "model" estimates (standard errors in parentheses)

	Random intercept model at household and community level		Probit				Probit, Marginal effects	
			Without clusters		Clusters at household level		Clusters at community level	
Constant	-2.754	***	-2.376	***	-2.376	***	-2.376	***
			(0.349)		(0.348)		(0.373)	
Embu	-0.032		-0.153		-0.153		-0.153	-0.021
	(0.299)		(0.186)		(0.259)		(0.258)	(0.039)
Kamba	0.315		0.270	**	0.270	*	0.270	0.031 *
	(0.214)		(0.136)		(0.154)		(0.171)	(0.017)
Kalenjin	-0.185		-0.173		-0.173		-0.173	-0.024
	(0.180)		(0.113)		(0.125)		(0.137)	(0.020)
Kisii	-0.063		-0.110		-0.110		-0.110	-0.015
	(0.250)		(0.150)		(0.170)		(0.209)	(0.030)
Luhya	-0.072		-0.010		-0.010		-0.010	-0.001
	(0.242)		(0.160)		(0.170)		(0.176)	(0.023)
Luo	-0.038		-0.048		-0.048		-0.048	-0.006
	(0.194)		(0.126)		(0.138)		(0.150)	(0.020)
Maasai	-1.074	***	-0.721	***	-0.721	***	-0.721	*** -0.126 ***
	(0.233)		(0.131)		(0.160)		(0.191)	(0.042)
Meru	-0.427	*	-0.412	***	-0.412	**	-0.412	** -0.064 *
	(0.234)		(0.149)		(0.170)		(0.186)	(0.034)
Mijikenda	-0.355		-0.253		-0.253		-0.253	-0.037
	(0.280)		(0.157)		(0.179)		(0.214)	(0.035)
Somali	-1.347	***	-0.962	***	-0.962	***	-0.962	*** -0.187 ***
	(0.263)		(0.148)		(0.174)		(0.205)	(0.053)
Swahili-coast	-0.655	***	-0.518	***	-0.518	***	-0.518	*** -0.083 **
	(0.218)		(0.132)		(0.152)		(0.175)	(0.033)
Swahili-rest	-0.293	*	-0.233	**	-0.233	**	-0.233	* -0.032 *
	(0.166)		(0.106)		(0.118)		(0.130)	(0.019)
English	-0.505	*	-0.278	*	-0.278		-0.278	-0.041
	(0.280)		(0.155)		(0.171)		(0.201)	(0.033)
Characteristics, by level								
Child	Yes		Yes		Yes		Yes	Yes
Household	Yes		Yes		Yes		Yes	Yes
Community or district	Yes		Yes		Yes		Yes	Yes
Log-likelihood	-		-2660		-2660		-2660	-2660
2531								
Pseudo R^2			0.266		0.266		0.266	0.266

Notes: *** 1% significance level, ** 5% significance level, and * 10% significance level.