Does colonialism have an impact on the current language situation in Sub-Saharan Africa?

Buzasi, Katalin

Utrecht University Faculty of Humanities Research Institute for History and Culture

22 November 2012

Does colonialism have an impact on the current language situation in Sub-Saharan Africa?

Katalin Buzasi, Utrecht University

Abstract

Recent studies show that colonization and ethnolinguistic fragmentation have relevant long-lasting impacts on ex-colonies and might explain development and underdevelopment of different parts of the world. The aim of this paper is to connect these two strands of the literature to some extent by investigating the effect of colonizer's identity on the current language situation in Sub-Saharan Africa. To establish the relationship between colonization and linguistic circumstances we conduct basic statistical analysis at the country level and more detailed econometric analyses at the level of individuals. The variable of interest is the Communication Potential Index (CPI) which reflects the probability that two randomly selected people in a society can communicate based on commonly spoken languages. Basic observations at the country level and multilevel analysis at the individual level suggest that current linguistic situation measured with the CPI cannot be explained by the identity of the colonizer.

Keywords: Sub-Saharan Africa, language, communication potential, colonialism, multi-level regression analysis

JEL Codes: N97 (Regional and Urban History, Africa and Oceania), Z19 (Cultural Economics; Economic Sociology; Economic Anthropology; Other)

Acknowledgements: I would like to thank Selin Dilli (UU), Ewout Frankema (UU and WU), Sandra de Pleijt (UU), Benjamin Guilbert (UU), and Peter Foldvari (UU) for their useful comments on the previous versions of this paper regarding the structure and the methodology, and Sarah Carmichael (UU) for language correction.

1. Introduction

Recent decades have seen an increasing number of studies aiming to seek the long term effects of colonialism on current economic, social and institutional conditions of previously occupied territories. One strand of literature focuses on the importance of the colonizer's identity in explaining development diversities. Empirical evidence suggest that depending on the nationality of the colonizer, former colonies in the modern era tend to perform differently in terms of economic development (Lange et al. 2006, Bertocchi and Canova 2002, Acemoglu et al. 2001, Grier 1999), the quality of democratic and government institutions (Angeles 2009, Olsson 2009, La Porta et al. 2008, La Porta et al. 1999), and the average level of education might also vary in ex-colonies (Cogneau 2003, Brown 2000).

The aim of this study is to contribute to a less explored field within the colonialism-related literature by investigating the possible role that European nations might have played in shaping the language situation of former colonies in the long run. Although it is well documented that ethnolinguistic fragmentation and ethnolinguistic polarization are responsible for conflict potential and the severe backwardness of Africa (Montalvo and Reynal-Querol 2005, Alesina et al. 1999, Hall and Jones 1999, Easterly and Levine 1997, Mauro 1995), the potential unifying effects of European languages as possible *lingua francas* have been less well explored. From a development perspective individual and country level linguistic features should not be ignored for at least two reasons; firstly, language is a basic tool for individual and group interactions that helps to solve collective action problems and improves cooperation (Smith 2010); moreover, common language seems to be an important factor in promoting bilateral trade (Ku and Zussmann 2010, Bosker and Garretsen 2008, Choi 2010, Feenstra et al. 2001). Secondly, since language is a crucial element of ethnic identity (Taylor et al. 1973) and language learning might play role in identity formation at the individual level (Clots-Figueras and Masella 2009), we might assume that acquisition of second languages can diminish the detrimental consequences of ethnic fragmentation at the level of the society.

An obvious reason for neglecting language effects on society this far might be that it is almost impossible to separate ethnic identity and language. This leads to the problem that it is difficult to draw conclusions about the outcomes of language use without the effect of ethnicity. Section 3 shows that ethnic and language groups do overlap, but that this is not necessarily always the case. A second reason for the underdevelopment of the field is that ethnolinguistic measures are mostly exploited for secondary goals like instrumenting institutional development or proxying cultural diversity in empirical studies and the distinction between ethnicity and language is therefore not of main concern. Data scarcity is a third obvious problem. Although sources providing data on mother tongue are available, sources on other spoken languages

which might be eligible for comparative investigations are difficult to find and hardly exploitable for economic, political or institutional research.

Based on colonial political and historical studies (Frankema 2011, Woodberry and Gallego 2010, Madeira 2006, Lange 2004, Posner 2004, Cogneau 2003, Posner 2003, Blanton et al. 2001, White 1996) we can consider at least four channels through which European colonization might have affected colonial language situation;(1) National boundaries in Africa, as set in 1885 at the Berlin Conference, are commonly accepted as a result of arbitrary decisions of colonizers without taking natural geographical circumstances, original social organization and ethnic group distribution into account; (2) The general concept behind the colonial rules involved attitudes towards the colonies, the recognition of the authority of local political leaders and social organizations; (3) The educational and (4) the elite formation strategies originating from the general colonial conception. The educational policy settled the medium of instruction, the perception of missionary education, rules for different denominations (Catholic and Protestant), investment in the education system and the curriculum itself. It is an established fact that the French promoted their own language even at lower level of education, the British encouraged teaching in local languages. While the medium of instruction had a direct impact on language, other elements of the colonial strategy might have affected language indirectly through influencing social stratification and ethnic relations. Since ethnicity and language are highly connected, colonial actions which distorted relative positions of ethnic groups or altered the costs of identity maintenance and identity shift are likely to have influenced the motivation for learning the language of the colonizer and the interest in maintaining one's ethnic identity.

However, the ultimate linguistic consequences of colonial policies implemented by European nations are far from unambiguous. In French colonies, all classes were taught in French which increases the expectation with respect to the share of the society speaking the colonizer's language. On the other hand, since ethnic groups experienced unequal chances to attain positions within the colonial administration, individual motivation to learn French might have remained low among the members of less preferred ethnic groups. The indirect ruling policy in the British colonies, on the one hand, might have contributed to maintain original ethnic constructions within the society. On the other hand, interest in learning the language of the colonizer might have been higher than in ex-French colonies since the British system provided ethnic groups with more equal chances to fulfil positions within the colonial administration.

The aim of this paper is to explore the possible persistent effects of colonization on the current language situation of twenty Sub-Saharan African countries covered by the fourth round

of Afrobarometer Survey.[1] The novelty of this study, when compared to previous literature, is that we do not restrict our analysis to mother tongues exclusively but exploit additional information on second spoken languages. We introduce a new measure, the Communication Potential Index (CPI) which shows the possibility that two randomly selected individuals in a country might be able to communicate based on commonly spoken languages. Different modifications of the country level CPI make it possible to shed light on the role that indigenous and European languages play in communication possibilities. Exploiting the CPI at the individual level we conduct multilevel regression analysis to investigate the language effects of European colonization.

The structure of this paper is as follows. The second section introduces the Communication Potential Index (CPI), which is the variable of interest in the empirical analysis. Section three provides an overview of current ethnolinguistic circumstances in Sub-Saharan Africa that is derived from data of the Afrobarometer Survey Round 4 (conducted in 2008 and 2009). In section four multilevel regression analyses are conducted to examine the possible persistent effect of colonizer's identity on current individual level communication potential. The final section summarizes.

2. The Communication Potential Index

In order to measure the effect of colonization on the current language situation we compute a new measure for twenty Sub-Saharan African countries which controls for mother and second spoken languages. The Communication Potential Index (CPI) measures the share of population which one randomly selected person is able to communicate with based exclusively on spoken languages. In other words, CPI shows the probability that a randomly selected person is able to communicate with another randomly selected person within a country. This measure accounts for only one influencing factor of communication (spoken language) whereas it ignores other possible relevant elements (the number of commonly known languages, geographical distance, willingness to speak to people of another ethnic group).

The Afrobarometer Survey is used to construct the CPI measure. The Afrobarometer is a politically independent research project that measures the social, political, and economic atmosphere in Africa. Surveys are repeated on a regular cycle.[2]Since the survey asks standard sets of questions, countries can be systematically compared. The sample is designed as a representative cross-section of all citizens of voting age in a given country. The goal is to give

[1] www.afrobarometer.org

[2] Round 1 between 1999 and 2001 with 12 countries included, Round 2 between 2002 and 2004 with 16 countries included, Round 3 in 2005 and 2006 with 18 countries included and Round 4 in 2008 and 2009 with 20 countries included. (Round 5 is being conducted at the moment with one additional country.)

every adult citizen an equal and known chance of selection for the interview.[3] Unfortunately, languages other than mother-tongue are only covered in Round 4 (2008/2009), which makes it impossible to consider changes over time.

Our indicator differs from previously used language related measures in several aspects. As mentioned above, instead of utilizing aggregate level data, we are given the chance to conduct analysis at the individual level. The widely used ethnolinguistic fragmentation index (Hall and Jones 1999, Easterly and Levine 1997, Mauro 1995, Taylor and Hudson 1972) and the language fractionalization measure of Alesina et al. (2003) focus solely on mother tongue and disregard the role of other spoken languages. Taking second languages into account enables us to analyze whether the economic consequence of natural separation by home language might be offset by acquiring second languages. Finally, data on the individual level might be used to compile a proxy for the aggregate level.

This basic form of the Communication Potential Index fails to control for the 'natural' part of the communication potential rooted in common membership of a certain ethnolinguistic group. Such a deficiency might cause problems when comparing countries with a high level of ethnolinguistic fragmentation and different relative group sizes. Countries with large, dominant ethnolinguistic groups naturally produce high communication potentials, since individuals of the same group can communicate based on their shared language.[4] To solve this problem, another type of the communication potential is calculated that measures the possibility that a randomly selected individual is able to communicate with another randomly selected person outside his or her own ethnic group. Analysing the gap between these forms of the communication potential paints a picture of the importance of the role that the relative size of ethnolinguistic groups plays in communication possibilities.

3. Current ethnic and linguistic situation in Sub-Saharan Africa

The first step to get a general idea of the legal status of languages in Sub-Saharan Africa is to provide a summary based on constitutions. There are countries in Sub-Saharan Africa that implemented the colonizer's language as an exclusive official language. Most former French colonies, such as Benin, Burkina Faso, Mali, Senegal etc, and some former British (Kenya, Namibia, Uganda, Zambia) introduced French or English as exclusive official languages in their constitutions. Most ex-British colonies are different, since most of them lack articles regulating the official language explicitly in their constitutions, although some of them require that

[3] This is achieved via (1) using random selection methods at every stage of sampling, and (2) sampling at all stages with probability proportionate to population size wherever possible to ensure that larger (i. e. more populated) geographic units have a proportionall greater probability of being chosen into the sample.
[4] This is true in countries where ethnic and linguistic groups are overlapping which is mostly the case. Section 3 shows some exceptions.

government officials and members of the parliament speak English and that state affairs are conducted in English (Botswana, the Gambia, Malawi). Kenya and Tanzania recognize Kiswahili, Lesotho recognizes Sesotho, and Burundi accepts Kirundi as official language beyond the inherited tongue of the colonizer. South Africa is unique with its eleven recognized official languages that are legislated in the constitution (Sepedi, Sesotho, Setswana, Siswati, Tshivenda, Xitsonga, Afrikaans, English, Isindebele, Isixhosa, and Isizulu). Cameroon and the Seychelles use both French and British as official languages.

Table 1 reports the most important dimensions of the ethnic and linguistic situations of the twenty Sub-Saharan African countries covered in Round 4 of the Afrobarometer Survey dataset. The sample contains eleven ex-British, five ex-French, two ex-Portuguese colonies and two 'other' countries (Liberia and Namibia).[5] Liberia received long-term support from the United States, and Namibia was governed by South Africa after the German occupation. Ethnic and linguistic fragmentation measures the probability that two randomly selected individuals within a country belong to different ethnic or linguistic groups. CPI (total) expresses the communication potential in our countries without differentiating by ethnic groups. CPI (extra) shows the probability that two randomly selected people originating in different ethnic groups are able to communicate based on commonly spoken languages. It is expected that countries with one or two dominant ethnic groups, in terms of size, will produce significant differences in these two indices. CPI (without) is the communication potential with the language of the colonizer removed from the data. Comparing this measure to the original helps to reveal the possible importance of the colonizer's language in current-day communication capacities. The drop in communication potential after excluding the colonizer's language and the share of inhabitants speaking European languages as first or second languages, functions as an appropriate tool to examine the relationship between local and ex-colonizer's languages and their contribution to the communication possibilities.

[Table 1 here]

As Table 1 reveals, there are some interesting outcomes: first, although macroeconomic research use ethnolinguistic fragmentation combined, Table 1 shows that ethnic and linguistic fragmentation might differ within countries. Ethnic fragmentation in Mali is more than ten percent higher than its linguistic fragmentation. Islands (Cape Verde and Madagascar) and Lesotho, which can be called an embedded country, experience huge gaps between these two measures. Investigating the particularities of such countries in linguistic terms is a challenge,

[5] The grouping of former colonies is taken over from Bertocchi and Canova (2002).

and provides an interesting avenue for future research. Zimbabwe has similar results; the difference between ethnic and linguistic fragmentation is about fifty percent.

In order to get some indication of the importance of European languages in Africa the communication potential and the communication potential without the colonizer's language are compared.[6] One should bear in mind that it does not perfectly reflect how the language situation would be if there had been no colonisation, since the effects of colonizers on indigenous languages cannot be controlled for.[7] By analyzing the difference between the two communication potential indices and the share of people speaking the colonizer's language as a second language, we might gain some indication of the relationship between local and European languages, as they can be substitutes or complementary. In certain countries communication potential is independent of the languages inherited from the colonizers. Botswana, Cape Verde, Kenya, Lesotho, Madagascar, Mali, Malawi, Senegal, Tanzania, and Zimbabwe experience a maximum of two percent drop in communication potential when we ignore the colonizer's language. The decrease in Burkina Faso and Zambia is about five percent. Benin and Ghana lose about ten percent and Uganda about twenty percent of their communication potential without the language of the colonizer. The cases that stand out the most are Liberia, Mozambique, Namibia, Nigeria, and South Africa with their drop of between 35 and 55 percent. The case of Cape Verde might be misleading: their first language (spoken by almost every respondent in the sample), the Cape Verdean Creole was developed from Portuguese, thus simply ignoring the Portuguese language and its transformation into another one does not reveal the true picture of the importance of the language of the colonizer. According to these basic findings, we can conclude that some African countries could adopt foreign languages, others not. In other words, colonizers were able to introduce and establish their languages successfully in some colonies, but not in the others. A clear relationship between the nationality of colonizers and the level of success is not observed. It would be a mistake to believe that in countries where the colonizer's language does not matter much for communication potential, European languages are less known. The last column of Table 1 reveals that European languages are common even in countries with high CPI(without) but they cannot be considered as exclusive lingua franca.

The Afrobarometer makes it possible to investigate the relationship between second language acquisition and education in the twenty available countries. In Section 2 the main features of the education strategies implemented by the European nations are discussed. Different underlying concepts and practices resulted in persistent variations in terms of average

[6] We computed only the communication potential without colonizer's language which takes ethnic group identity into consideration. (We computed only the CPI(extra) without the language of colonizer).

[7] Still, if the colonizer's language is expected to become dominant over time, its effect on the use of other indigenous languages should rather be negative (if we think of languages as substitutes). This means that the CPI(without) is more likely to underestimate than to overestimate the magnitude of communication potential in the absence of colonization.

years of education and the literacy rates. Table 2 compares the distribution of the individual level communication potential between education categories in our sample. Since we have only two countries in the 'Portuguese' and 'Other' categories and one of the two Portuguese colonies is an island (that makes it a special case in linguistic terms) with a language developed from Portuguese, the focus is on the British and French comparison. The average communication potential and standard deviation does not differ much and the standard deviation of the communication potential decreases with increases in education. Lower level schooling seems to lead to a higher communication potential in French colonies, which might be the result of the primacy of the French language at lower levels of education that is inherited from the colonial tradition.

[Table 2 here]

The other possible difference between colonies is the share of people who speak the language of the colonizer (Table 3). Of the population colonized by the British 46 percent speak English as a second language, whereas French is spoken by 29 percent of the people in ex-French territories. This pattern might be an overall consequence of the elements of the colonial rule (see section 2). To summarize, in former British countries there is a higher probability of speaking the language of the colonizer when citizens have completed less than primary education. French territories have the advantage in primary and secondary schooling categories, but in higher education the colonizer's language is spoken with higher probability in ex-British countries.

[Table 3 here]

4. Results of the multilevel regression analysis

The previous section attempted to present general patterns of language use in Sub-Saharan Africa at the country level. In this section we conduct multilevel regression analysis to reveal the possible effect of colonization on the individual level communication potential. Multilevel model techniques are designed for data which has a hierarchical or clustered structure. It assumes the dependent variable to be measured at the lowest level, whereas the explanatory variables can be measured at different levels. In our analysis a two-level multilevel modelling method with an individual and a country level is applied, where the country specific effects are allowed to depend on country specific factors (including the colonizer). It can be conceptualized as a two-stage system of equations in which the individual variation within each group is explained by an individual-level equation, and the variation across groups in the group-specific regression coefficients is explained by a group-level equation (Diez-Roux 2000). The dependent and explanatory variables are described in Table 4. The results of our different model specifications

are summarized in Table 5.[8] Technical details of the estimated models are provided in the Appendix.

[Table 4 here]

When using hierarchical or structured data, social scientists prefer multilevel modelling techniques over the OLS regression method for several reasons. Applying OLS regression instead of multilevel modelling when it is appropriate (i.e. analyzing variables from different levels at one single common level) creates two sets of problems (Hox 1995 p. 4-6.). The first of these is statistical. Standard statistical tests are based on the assumption of independence of the observations. In hierarchical data this assumption is violated since observations within a group tend to be more similar than observations randomly sampled from the larger population. Thus the average correlation between variables representing individuals from the same group will be higher than the average correlation between variables representing individuals from different groups. If the independency assumption is violated the estimates of the standard errors of conventional statistical tests are much too small, and this results in many spuriously 'significant' results. The other set of problems is conceptual. With a hierarchical sample it can be the case that one investigates data at a certain level and concludes at a different level.[9] Multilevel modelling techniques are designed to diminish the statistical and conceptual difficulties rooted in the hierarchical structure of the data.

We estimated six multilevel models with the maximum likelihood estimation method and compared them with the widely used likelihood ratio (LR) test. Model 1 is a two-level random intercept model assuming that regression intercepts might differ between countries. Contrary to the OLS method, instead of estimating 19 intercepts (we have 20 countries, thus 19 dummy variables) multilevel analysis estimates only the variance or the standard deviation of these intercepts. The estimated standard deviation of the intercepts in Model 1 is 0.149 and significant (see the standard error of the estimation in parentheses). Individual level variables turn out to be significant in explaining individual level communication potential. While the age and the female dummy are negatively associated with the dependent variable, education increases it *ceteris paribus*. Among the country level variables linguistic fragmentation and population density have significant and negative coefficients, however, the coefficient of the latter is very close to zero. The remaining two country level variables (GDP per capita and being landlocked)

[8] Since regression results using the CPI(total) and CPI(extra) as dependent variables do not differ essentially we report only coefficients of the models with the CPI_extra.

[9] Ecological fallacy means the mistake of interpreting aggregated data based results at the individual level (Robinson 1950). Atomistic or individualistic fallacy refers to drawing inferences at a higher level from analyses performed at a lower level (Diez-Roux 1998).

are not significant. Model 2 takes colonial history into account, Model 3 introduces interaction variables assuming that the effect of education on communication potential might depend on the identity of the former colonizer. Colonizer dummies in Model 2 are not significant. In Model 3 the British dummy and the interaction variables between education and the identity of former colonizer are significant. Thus, Model 3 suggests that individuals in former British colonies have higher communication potential than colonies of the 'other' group (reference case) at the 10% significance level and the effect of education on communication potential differs between colonizers. In colonies of the 'other' category a person who has an advantage of one year in education experiences 0.026 higher communication potential *ceteris paribus*. The significant negative coefficients of the interaction variables suggest that the increasing effect of education on communication potential is less in the remaining colonies. One extra year spent in education implies 0.019 (0.026-0.07) higher communication potential in former British colonies. The corresponding effects for former French and Portuguese colonies are 0.011 (0.026-0.015) and 0.020 (0.026-0.006), respectively (see Table 5 Model 3).

Model 4 is identical to Model 1 with the exception that we assume that the coefficient of education varies randomly between countries. In this random intercept-random slope model the estimated intercepts have higher estimated standard deviation (0.268) compared to that in the random intercept models (0.149). The standard deviation of the random coefficient is 0.014 and significant. Table 5 reports correlation between the random intercept and the random slope. The significant strong negative correlation (-0.911) between the intercept and the coefficient of education indicates that for countries with a relatively high intercept, a relatively low slope is observed. Model 5 introduces colonizer dummies, Model 6 allows for interaction between the identity of colonizer and education. Similarly to Model 2, Model 5 produces insignificant coefficients for the colonizer dummies. Comparing Model 6 to Model 3 we find relevant differences. The British colony dummy is not significant anymore and the interaction variable is significant only in the French case. It means that compared to countries in the 'other' colony group education has a significantly lower effect on the communication potential only in the former French colonies.

Random intercept and random intercept-random slope models have relevant differences with respect to the non colonialism related variables. In random intercept-random slope models linguistic fragmentation becomes insignificant and population density is not significant in Model 4.

[Table 5 here]

When applying maximum likelihood estimation method, log-likelihood values might be exploited for model selection. Table 6 summarizes the steps of the model selection procedure. Likelihood ratio tests suggest that introducing colonizer dummies in Model 2 is not necessary. However, Model 3 seems to provide relevant additional information with the interaction variables. Likelihood ratio tests also suggest that assuming random slope in Model 4 is relevant. Our model selection test does not support the relevance of taking colonizer dummies and interaction variables into account in Model 5 and Model 6 after considering random slope in Model 4. The final selection might be complicated, since the remaining two strongest model specifications (Model 3 and Model 4) have essentially different conclusion on the explanatory variables of interest (colonization dummy and the interaction variables). Although the p-value of the likelihood ratio test is very close to zero in both cases, it is slightly closer when we contrast Model 4 with Model 1. Moreover, AIC and BIC information criteria prefer Model 4 over Model 3. Thus, Model 4 turns out to be the best fitted model with a random intercept and a random slope without colonizer dummies and interaction variables.

In Table 7 we present the results of the OLS estimation with country and colonizer dummies and interaction variables between countries and education and colonizers and education in order to show the misleading consequences of using the OLS method instead of multilevel modelling when we have hierarchical data. All the dummies and interaction variables are significant. The Akaike Information Criterion would prefer the OLS estimation.

5. Conclusion

Development research recognizes the role that colonization and ethnolingustic fragmentation play in explaining underdevelopment of certain area of the world. Ethnicity and mother tongue are natural factors which lead to separation of people. Elimination of these types of barriers means incurring transaction costs. Second language learning might be an effective tool to abolish elements hindering communication and cooperation between people. Although second language accumulation might have a significant impact on the economic, social, and institutional development of a country, economic literature fails to reveal them partly due to a lack of data. This paper focuses on potential long lasting effects of colonization on second language accumulation.

Based on individual level data provided by the Afrobarometer Survey Project measures describing ethnic, linguistic and education differences in twenty Sub-Saharan African countries previously colonized by different European nations (British, French, Portuguese, and other) were created. Although the literature does not differentiate between ethnic and linguistic

diversity we see that ethnic and linguistic groups in Sub-Saharan Africa do not necessarily overlap. Based on second spoken languages we introduced the Communication Potential Index which shows the chance that two randomly selected persons within a country are able to communicate with each other based on commonly spoken languages. We computed the same measure without the language of colonizers to be able to get an indication of the importance of colonizer's language in each country. Afrobarometer provided the opportunity to calculate the share of respondents speaking European languages as mother tongue and second language. Comparing these language related measures we investigate the relation between local and European languages in Africa.

In the final part of the paper we conducted econometric analysis at the individual level. We applied multilevel analysis technique to reveal the effects of colonization on communication potential. We experimented with random intercept and random intercept-random slope models with an individual and a country level. In some specifications we assumed that the effect of education on the communication potential might vary between colonizers (we introduced interaction variables). The best model selected based on LR tests and AIC and BIC information criteria suggest that the identity of the colonizer is not a significant explanatory variable in explaining communication potential in our twenty Sub-Saharan African countries. Neither are the interaction variables, controlling for the possibility of varying education effects between different colonies, significant.

Basic observations and statistical analyses conducted in this paper encourage the opening up of avenues for future research. One possible path might be to focus on interesting cases provided by diversity and communication potential comparisons and try to get closer to the factors causing surprising outcomes. For instance islands (Cape Verde and Madagascar) emerge as natural case studies. The other possible way is to choose general phenomena which had an impact in every country and compare the difference. For instance Christian missions took place in every countries of Africa which might have affected linguistic outcomes differently. Finally we could compare the pre-colonial situation in Africa and study how these initial conditions determined the success of colonizing strategies in linguistic terms.

References

Acemoglu, D., and S. Johnson, and J.A. Robinson. 2001. The colonial origins of comparative development: An empirical investigation. *American Economic Review* 91(5): 1369-1401.

Alesina, A., R. Baqir, and W. Easterly. 1999. Public goods and ethnic divisions. *The Quarterly Journal of Economics* 114 (4): 1243-84.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. 2003. Fractionalization. *Journal of Economic Growth* 8(2): 155-194.

Angeles, L. 2009. Colonialism, European descendants and democracy. *SIRE Discussion Papers 2009-50*, Scottish Institute for Research in Economics

Bertocchi, G., and Canova, F. 2002. Did colonization matter for growth? An empirical exploration into the historical causes of Africa's underdevelopment. *European Economic Review* 46 (10): 1851-1871.

Blanton, R., Mason, D. T., and Athow, B. 2001. Colonial style and post-colonial ethnic conflict in Africa. *Journal of Peace Research* 38 (4): 473-491.

Bosker, M., and H. Garretsen. 2012. Economic geography and economic development in Sub-Saharan Africa. *The World Bank Economic Review*. Advanced access 26.06.2012. (First published online 14.02.2012.)

Brown, D. S. 2000 . Democracy, colonization, and human capital in Sub-Saharan Africa. *Studies in Comparative International Development* 35(1): 20-40.

Choi, E. K. 2002. Trade and the adoption of a universal language. *International Review of Economics & Finance* 11 (3): 265-75.

Clots-Figueras, I. and P. Masella. 2009. Education, Language and Identity. Unpublished.

Cogneau, D. 2003. Colonisation, school and development in Africa. An emprical analysis. DIAL Document de Travail 2003/01

Diez-Roux, A. V. 1998. Bringing context back into epidemology: Variables and fallacies in multilevel analysis. *American Journal of Public Health* 88(2): 216-222.

Diez-Roux, A. V. 2000. Multilevel analysis in public health research. *Annual Review of Public Health* 21 (1): 171-92.

Easterly, W., and R. Levine. 1997. Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics* 112 (4): 1203-50.

Feenstra, R. C., J. R. Markusen, and A. K. Rose. 2001. Using the gravity equation to differentiate among alternative theories of trade. *Canadian Journal of Economics/Revue Canadienne d'Économique* 34 (2): 430-47.

Frankema, E. 2011. The origins of formal education in Sub-Saharan Africa. Was the British rule more benign? CGEH Working Paper Series No. 5, Utrecht University

Gallego, F. A., and Woodberry, R. 2010. Christian missionaries and education in former African colonies: How competition mattered. *Journal of African Economies* 19(3): 294-329.

Grier, R. M. 1999. Colonial legacies and economic growth. *Public Choice* 98(3-4): 317-335.

Hall, R. E., and C. I. Jones. 1999. Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics* 114 (1): 83-116.

Hox, J. J. 1995. Applied Multilevel Analysis. TT-Publikaties, Amsterdam. Online: http://igitur-archive.library.uu.nl/fss/2007-1114-201211/hox_95_applied%20multilevel%20analysis.pdf, accesses: 16.10.2012.

Ku, H., and A. Zussman. 2010. Lingua franca: The role of English in international trade. *Journal of Economic Behavior & Organization* 75 (2): 250-60.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny. 1999. The quality of government. *Journal of Law, Economics, and Organization* 15 (1): 222-79.

La Porta, R., F. Lopez-de-Silanes, and A. Shleifer. 2008. The economic consequences of legal origins. *Journal of Economic Literature* 46(2): 285-332.

Lange, M. K. 2004. British colonial legacies and political development. *World Development* 32 (6): 905-922.

Lange, M., J. Mahoney, and M. Vom Hau. 2006. Colonialism and development: A comparative analysis of Spanish and British colonies. *American Journal of Sociology* 111 (5): 1412-62.

Madeira, A. I. 2005. Portuguese, French and British discourses on colonial education: Church-state relations, school expansion and missionary competition in Africa, 1890-1930. *Paedagogica Historica: International Journal of the History of Education* 41 (1-2): 31-60.
Mauro, P. 1995. Corruption and growth. *The Quarterly Journal of Economics* 110 (3): 681-712.
Montalvo, J. G., and M. Reynal-Querol. 2005. Ethnic polarization, potential conflict, and civil wars. *The American Economic Review* 95 (3): 796-816.
Olsson, O. 2009. On the democratic legacy of colonialism. *Journal of Comparative Economics* 37 (4): 534-51.
Posner, D. N. 2003. The colonial origins of ethnic cleavages: The case of linguistic divisions in Zambia. *Comparative Politics* 35(2): 127-46.
Posner, D. N. 2004. The political salience of cultural difference: Whaóy Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. *American Political Science Review* 98 (4): 529-545.
Robinson, W.S. 1950. Ecological Correlations and the Behavior of Individuals. *American Sociological Review* 15(3): 351–357.
Smith, E. A. 2010. Communication and collective action: Language and the evolution of human cooperation. *Evolution and Human Behavior* 31(4): 231-245.
Taylor, C. L., & Hudson, M. C. (1972). *World Handbook of Political and Social Indicators, 2nd edition.* New Haven: Yale University Press, 271-274.
Taylor, D. M., J. N. Bassili, and F. E. Aboud. (1973). Dimensions of ethnic identity: An example from Quebec. *The Journal of Social Psychology* 89(2): 185-192.
White, B. (1996). Talk about school: Education and the colonial project in French and British Africa, (1860-1960). *Comparative Education* 32(1): 9-25.

Table 1

Ethnic and linguistic situation in twenty Sub-Saharan African countries

Country	Sample size	Colonizer	Ethnic fragm	Linguistic fragm	CPI (total)	CPI (extra)	Dif_1 CPI (total)-CPI (extra)	CPI (extra without)	Dif_2 CPI (extra)-CPI (without)	Colonizer language share
Benin	1200	France	0.823	0.816	0.580	0.498	0.082	0.376	0.122	0.466
Botswana	1200	Britain	0.920	0.407	0.986	0.985	0.001	0.983	0.002	0.417
Burkina Faso	1200	France	0.687	0.702	0.603	0.436	0.167	0.386	0.050	0.305
Cape Verde	1264	Portugal	0.444	0.006	0.991	0.989	0.002	0.989	0.000	0.003 (1.000)+
Ghana	1200	Britain	0.727	0.717	0.625	0.498	0.127	0.378	0.120	0.475
Kenya	1104	Britain	0.890	0.892	0.916	0.906	0.010	0.892	0.014	0.594
Lesotho	1200	Britain	0.883	0.040	1.000	1.000	0.000	1.000	0.000	0.275
Liberia	1200	Other	0.887	0.885	0.597	0.553	0.044	0.068	0.485	0.724 (0.220)++
Madagascar	1350	France	0.822	0.483	1.000	1.000	0.000	1.000	0.000	0.292
Malawi	1200	Britain	0.783	0.728	0.884	0.859	0.025	0.858	0.002	0.235
Mali	1232	France	0.839	0.719	0.803	0.776	0.027	0.764	0.012	0.223
Mozambique	1200	Portugal	0.870	0.872	0.697	0.651	0.046	0.097	0.554	0.787
Namibia	1200	Other	0.705	0.701	0.816	0.730	0.086	0.328	0.402	0.762
Nigeria	2324	Britain	0.856	0.866	0.622	0.552	0.071	0.202	0.349	0.645 (0.784)+++
Senegal	1200	France	0.697	0.604	0.891	0.853	0.038	0.851	0.002	0.216
South Africa	2400	Britain	0.866	0.855	0.582	0.518	0.064	0.249	0.269	0.656
Tanzania	1208	Britain	0.953	0.950	0.991	0.990	0.000	0.990	0.000	0.000
Uganda	2431	Britain	0.896	0.896	0.498	0.443	0.055	0.252	0.192	0.512
Zambia	1200	Britain	0.884	0.872	0.643	0.598	0.045	0.552	0.047	0.405
Zimbabwe	1200	Britain	0.827	0.331	0.876	0.855	0.021	0.837	0.019	0.443

+ The share of respondents speaking the language of colonizer (Portuguese). In parentheses we indicate that if we consider that Creole language was developed on Portuguese the colonizer's language share and thus communication potential are 100%.

++ In parentheses we show the share of people speaking English as mother tongue.

+++ 0.645 is the share of people speaking English. In parentheses we added the share of people speaking English or Pidgin English.

Table 2
Mean and standard deviation of CP (extra) in different education categories and colonies

	British		French		Portuguese		Other	
education level	CPI (extra)	std. dev.	CPI (extra)	std. dev.	CPI (extra)	std. dev.	CPI (extra)	std. dev.
No formal schooling	0.589	0.386	0.578	0.328	0.497	0.435	0.436	0.330
Informal schooling only	0.471	0.344	0.733	0.297	0.664	0.341	0.362	0.325
Some primary schooling	0.654	0.359	0.785	0.272	0.704	0.342	0.466	0.329
Primary completed	0.730	0.320	0.853	0.229	0.807	0.228	0.596	0.292
Some secondary	0.700	0.275	0.789	0.231	0.849	0.163	0.716	0.233
Secondary completed	0.749	0.226	0.830	0.215	0.854	0.151	0.782	0.178
Post-secondary, not univ.	0.760	0.185	0.900	0.120	0.867	0.093	0.825	0.151
Some university	0.746	0.168	0.810	0.202	0.915	0.094	0.764	0.149
Univeity completed	0.756	0.175	0.822	0.200	0.950	0.081	0.791	0.190
Post-graduate	0.762	0.176	0.930	0.133	-	-	0.736	0.315
Total	0.697	0.302	0.715	0.301	0.766	0.291	0.646	0.293

Table 3

Share of people speaking the colonizer's language in different education categories and colonies

education level	British			French			Portuguese			Other		
	# in category (sum of weights)	# of speaking English within category	%	# in category (sum of weights)	# of speaking French within category	%	# in category (sum of weights)	# of speaking Portuguese within category	%	# in category (sum of weights)	# of speaking English within category	%
No formal schooling	1570.104	170.9912	0.109	2147.755	86.8023	0.040	277.9678	19.8904	0.072	334.8807	153.4754	0.458
Informal schooling only	440.8583	59.8176	0.136	734.4217	62.6549	0.085	27.6269	12.7357	0.461	60.1279	20.5696	0.342
Some primary schooling	2963.256	495.0513	0.167	1152.994	297.6675	0.258	639.2079	212.9777	0.333	302.5219	149.021	0.493
Primary completed	2894.062	749.019	0.259	474.0745	150.1439	0.317	363.8297	168.1547	0.462	251.0605	154.8069	0.617
Some secondary	3606.73	2028.138	0.562	1087.343	725.7615	0.667	661.3293	332.0123	0.502	641.7905	544.9931	0.849
Secondary completed	3134.535	2365.413	0.755	233.6965	187.9998	0.804	273.3542	133.2016	0.487	530.3762	493.4416	0.930
Post-secondary, not univ.	1349.461	1183.533	0.877	103.482	75.5692	0.730	65.646	27.811	0.424	115.8346	110.5362	0.954
Some university	299.8838	271.0594	0.904	149.9559	128.3256	0.856	93.8531	27.344	0.291	103.0609	100.0991	0.971
Univeity completed	333.381	291.5639	0.875	61.7537	51.8339	0.839	50.8831	10.4264	0.205	45.889	43.6467	0.951
Post-graduate	55.4963	50.0967	0.903	29.5114	23.8909	0.810	0	0	-	8.6544	8.204	0.948
Total	16647.7674	7664.6831	0.460	6174.9877	1790.6495	0.290	2453.698	944.5538	0.385	2394.1966	1778.7936	0.743

Table 4

Dependent and independent variables in empirical analyses

	Role in regression	Notes	Source
individual-level variables			
CPI (total)	dependent variable	the probability that two randommly selected people can communicate	computed based on the Afrobarometer data in Round 4 (2008/2009)
CPI (extra)	dependent variable	the probability that two randomly selected people from different ethnic groups can communicate	computed based on the Afrobarometer data in Round 4 (2008/2009)
urban dummy	explanatory variable	1' - respondent lives in urban area, '2'-rural area, '3' - semi-urban area	Afrobarometer Survey Round 4 (2008/2009)
age	explanatory variable	age of respondent's in years	Afrobarometer Survey Round 4 (2008/2009)
gender dummy	explanatory variable	1' - respondent is female	Afrobarometer Survey Round 4 (2008/2009)
educ	explanatory variable	average years of education	We turned the education cathegories in Afrobarometer Survey (2008/2009) into average years of education
country-level variables			
gdp	explanatory variable	GDP per capita in international Geary-Khamis dollar (1990)	Maddison dataset (2008)
lingfrag	explanatory variable	the probability that two randomly selected people come from different lingusitc groups	computed based on the Afrobarometer data in Round 4 (2008/2009)
popdens	explanatory variable	people per square kilometer	World Bank (2008)
landlocked	explanatory variable	1 - country has no coastal area	time invariant data
country dummies	explanatory variable		time invariant data
colonizer dummies	explanatory variable	1- ex-British colony, 2 - ex French colony, 3 - ex-Portuguese colony, 0 - other	Bertocchi-Canova (2002)

Table 5

Results of the multilevel regression analysis with maximum likelihood estimation method (OLS regression results are included for comparison)[+]

Dependent variable: individual level CPI(extra)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
const	1.188 (0.000)	1.126 (0.000)	1.074 (0.000)	0.863 (0.000)	0.836 (0.000)	0.585 (0.005)
age	-0.0003 (0.003)	-0.0003 (0.003)	-0.0003 (0.006)	-0.0005 (0.000)	-0.0005 (0.000)	-0.0005 (0.000)
female	-0.021 (0.000)	-0.021 (0.000)	-0.021 (0.000)	-0.019 (0.000)	-0.019 (0.000)	-0.019 (0.000)
educ	0.018 (0.000)	0.018 (0.000)	0.026 (0.000)	0.015 (0.000)	0.015 (0.000)	0.030 (0.002)
gdp	-0.00004 (0.132)	-0.00005 (0.105)	-0.00005 (0.090)	-0.000 (0.957)	-0.0000 (0.673)	-0.0000 (0.674)
lingfrag	-0.518 (0.000)	-0.528 (0.000)	-0.533 (0.000)	-0.156 (0.114)	-0.110 (0.254)	-0.111 (0.252)
popdens	-0.001 (0.082)	-0.002 (0.027)	-0.002 (0.026)	-0.0008 (0.124)	-0.001 (0.026)	-0.001 (0.026)
landlocked	-0.064 (0.377)	-0.105 (0.184)	-0.110 (0.166)	-0.065 (0.230)	-0.065 (0.249)	-0.065 (0.248)
rural[++]	141.89 (0.000)	141.94 (0.000)	156.92 (0.000)	211.63 (0.000)	211.58 (0.000)	211.62 (0.000)
British	-	0.163 (0.193)	0.217 (0.085)	-	0.060 (0.505)	0.315 (0.116)
French	-	0.087 (0.504)	0.175 (0.179)	-	-0.038 (0.675)	0.315 (0.144)
Portuguese	-	0.109 (0.482)	0.159 (0.308)	-	0.146 (0.186)	0.365 (0.157)
British *educ	-	-	-0.007 (0.000)	-	-	-0.015 (0.154)
French*educ	-	-	-0.015 (0.000)	-	-	-0.020 (0.070)
Portuguese*educ	-	-	-0.006 (0.000)	-	-	-0.012 (0.351)
Random-effect parameters[+++]						
sd_constant (τ_{00})	0.149 (0.024)	0.142 (0.023)	0.142 (0.023)	0.268 (0.024)	0.271 (0.051)	0.254 (0.046)
sd_educ (τ_{30})	-	-	-	0.014 (0.002)	0.014 (0.002)	0.013 (0.002)
corr(educ_const)	-	-	-	-0.911 (0.043)	-0.930 (0.034)	-0.920 (0.038)
sd_residual (σ)	0.205 (0.0009)	0.205 (0.0009)	0.204 (0.0009)	0.198 (0.0008)	0.198 (0.0009)	0.198 (0.0009)
Number of observations	25875	25875	25875	25875	25875	25875
AIC	-8359.64	-8355.52	-8553.09	-10245.48	-10243.69	-10240.76
BIC	-8261.71	-8233.11	-8406.20	-10131.22	-10104.95	-10077.54
log-likelihood	4191.822	4192.7616	4294.5467	5136.7397	5138.8451	5140.3805

[+]p-values are in parentheses

[++]For rural and semi-urban dummies, country dummies, and for interaction variable between countries and education only the results of joint significance tests (chi-square and p-value) are reported.

[+++]Regarding the random effect parameters standard errors are reported in parentheses (not p-values)

Table 6

Comparison of the multilevel models in Table 5

	test statistic of the LR test (-2*(difference in the log likelihood values))	degree of freedom of the Chi-square distribution	p-value
Model 2 vs. Model 1	2.9396	3	0.401
Model 3 vs. Model 1	104.7247	6	<0.001
Model 4 vs. Model 1	1889.8344	2	<0.001
Model 5 vs. Model 4	4.2118	3	0.239
Model 6 vs. Model 4	7.2826	6	0.296

Table 7

OLS regression results[+]

	dep. var: CPI (extra)
const	2.313 (0.000)
age	-0.0005 (0.000)
female	-0.019 (0.000)
educ	0.038 (0.000)
gdp	-0.0002 (0.000)
lingfrag	-1.254 (0.000)
popdens	-0.005 (0.000)
landlocked	0.025 (0.057)
rural[++]	112.82 (0.000)
country[++]	529.62 (0.000)
i.country*educ[++]	287.83 (0.000)
British	0.326 (0.000)
French	-0.178 (0.000)
Portuguese	-0.071 (0.016)
British *educ	-0.039 (0.000)
French*educ	-0.015 (0.000)
Portuguese*educ	omitted
N	25875
AIC	-10438.5
BIC	-10079.4
R^2	0.567

[+]p-values are in parentheses

[++]For rural and semi-urban dummies, country dummies, and for interaction variable between countries and education only the results of joint significance tests (chi-square and p-value) are reported.

Appendix

Multilevel regression models with ML estimation method (Table 5)

Model 1

Level 1

$$CP_{ij} = \beta_{0j} + \beta_{1j} age_{ij} + \beta_{2j} female_{ij} + \beta_{3j} educ_{ij} + \beta_{4j} rural_{ij} + \beta_{5j} semiurban_{ij} + r_{ij}$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01} gdp_j + \gamma_{02} lingfrag_j + \gamma_{03} popdens_j + \gamma_{04} landlocked_j + u_{0j}$$

Thus the complete Model 1 is

$$\begin{aligned} CP_{ij} &= \gamma_{00} + \gamma_{01} gdp_j + \gamma_{02} lingfrag_j + \gamma_{03} popdens_j + \gamma_{04} landlocked_j \\ &+ \beta_{1j} age_{ij} + \beta_{2j} female_{ij} + \beta_{3j} educ_{ij} + \beta_{4j} rural_{ij} + \beta_{5j} semiurban_{ij} + u_{0j} + r_{ij} \end{aligned}$$

Model 3

Level 1

$$\begin{aligned} CP_{ij} &= \beta_{0j} + \beta_{1j} age_{ij} + \beta_{2j} female_{ij} + \beta_{3j} educ_{ij} + \beta_{4j} rural_{ij} + \beta_{5j} semiurban_{ij} + \\ &\delta_{1j} British_j * educ_{ij} + \delta_{2j} French_j * educ_{ij} + \delta_{3j} Portuguese_j * educ_{ij} + r_{ij} \end{aligned}$$

Level 2

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} gdp_j + \gamma_{02} lingfrag_j + \gamma_{03} popdens_j + \gamma_{04} landlocked_j + \\ &\gamma_{05} British_j + \gamma_{06} French_j + \gamma_{07} Portuguese_j + u_{0j} \end{aligned}$$

Thus the complete Model 3 is

$$\begin{aligned} CP_{ij} &= \gamma_{00} + \gamma_{01} gdp_j + \gamma_{02} lingfrag_j + \gamma_{03} popdens_j + \gamma_{04} landlocked_j + \gamma_{05} British_j + \gamma_{06} French_j \\ &+ \gamma_{07} Portuguese_j + \beta_{1j} age_{ij} + \beta_{2j} female_{ij} + \beta_{3j} educ_{ij} + \beta_{4j} rural_{ij} + \beta_{5j} semiurban_{ij} + \\ &\delta_{1j} British_j * educ_{ij} + \delta_{2j} French_j * educ_{ij} + \delta_{3j} Portuguese_j * educ_{ij} + u_{0j} + r_{ij} \end{aligned}$$

Model 2 equals Model 3 without the interaction variables controlling for connection between the identity of the colonizer and education.

Model 1, Model 2 and Model 3 are random intercept models.

Model 4

Level 1

$$CP_{ij} = \beta_{0j} + \beta_{1j} age_{ij} + \beta_{2j} female_{ij} + \beta_{3j} educ_{ij} + \beta_{4j} rural_{ij} + \beta_{5j} semiurban_{ij} + r_{ij}$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01} gdp_j + \gamma_{02} lingfrag_j + \gamma_{03} popdens_j + \gamma_{04} landlocked_j + u_{0j}$$

and

$$\beta_{3j} = \gamma_{30} + u_{3j}$$

The complete Model 4 is

$$CP_{ij} = \gamma_{00} + \gamma_{01} gdp_j + \gamma_{02} lingfrag_j + \gamma_{03} popdens_j + \gamma_{04} landlocked_j + \beta_{1j} age_{ij} + \beta_{2j} female_{ij} + \beta_{4j} rural_{ij} + \beta_{5j} semiurban_{ij} + \gamma_{30} educ_{ij} + u_{0j} + u_{3j} * educ_{ij} + r_{ij}$$

Model 6

Level 1

$$CP_{ij} = \beta_{0j} + \beta_{1j} age_{ij} + \beta_{2j} female_{ij} + \beta_{3j} educ_{ij} + \beta_{4j} rural_{ij} + \beta_{5j} semiurban_{ij} + \delta_{1j} British_j * educ_{ij} + \delta_{2j} French_j * educ_{ij} + \delta_{3j} Portuguese_j * educ_{ij} + r_{ij}$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01} gdp_j + \gamma_{02} lingfrag_j + \gamma_{03} popdens_j + \gamma_{04} landlocked_j + \gamma_{05} British_j + \gamma_{06} French_j + \gamma_{07} Portuguese_j + u_{0j}$$

and

$$\beta_{3j} = \gamma_{30} + u_{3j}$$

The complete Model 6 is

$$CP_{ij} = \gamma_{00} + \gamma_{01} gdp_j + \gamma_{02} lingfrag_j + \gamma_{03} popdens_j + \gamma_{04} landlocked_j + \gamma_{05} British_j + \gamma_{06} French_j + \gamma_{07} Portuguese_j + \beta_{1j} age_{ij} + \beta_{2j} female_{ij} + \beta_{4j} rural_{ij} + \beta_{5j} semiurban_{ij} + \delta_{1j} British_j * educ_{ij} + \delta_{2j} French_j * educ_{ij} + \delta_{3j} Portuguese_j * educ_{ij} + \gamma_{30} educ_{ij} + u_{0j} + u_{3j} * educ_{ij} + r_{ij}$$

Model 5 equals Model 6 without the interaction variables of the colonizer and education.

Model 4, Model 5, and Model 6 are called random intercept random slope multilevel models.

CP_{ij} is the communication potential measure for individual *i* in country *j*, r_{ij} is the individual error term with $r_{ij} \sim N(0, \sigma^2)$ distribution in every model specification. u_{0j} is the random part of the intercept with distribution $u_{0j} \sim N(0, \tau_{00}^2)$ in Model 1 to Model 6. u_{3j} is the random part of the coefficient of the education with distribution $u_{3j} \sim N(0, \tau_{30}^2)$ in Model 4 to Model 6.

The OLS estimation (Table 7)

$$CP_{ij} = \beta_0 + \beta_1 age_{ij} + \beta_2 female_{ij} + \beta_3 educ_{ij} + \beta_4 gdp_j + \beta_5 lingfrag_j +$$
$$\beta_6 popdens_j + \beta_7 landlocked_j + \beta_8 rural_{ij} + \beta_9 semiurban_{ij} +$$
$$\sum_{j=2}^{20} \delta_{j-1} country_j + \sum_{j=2}^{20} \lambda_{j-1} country_j * educ_{ij} +$$
$$\varphi_1 British_j + \varphi_2 French_j + \varphi_3 Portuguese_j +$$
$$\rho_1 British_j * educ_{ij} + \rho_2 French_j * educ_{ij} + \rho_3 Portuguese_j * educ_{ij} + \varepsilon_{ij}$$

CP_{ij} is the communication potential of individual i in country j $(j = 1, 2, ..., 20)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$ is the error term in every model specification.