

This paper is a draft submission to the

## WIDER Development Conference

# Human capital and growth 

6-7 June 2016 Helsinki, Finland

This is a draft version of a conference paper submitted for presentation at UNU-WIDER's conference, held in Helsinki on 6-7 June 2016. This is not a formal publication of UNU-WIDER and may reflect work-in-progress.

THIS DRAFT IS NOT TO BE CITED, QUOTED OR ATTRIBUTED WITHOUT PERMISSION FROM AUTHOR(S).

# Language of Instruction and Student Learning: <br> Evidence from an Experimental Program in Cameroon 

David D. Laitin*, Rajesh Ramachandran ${ }^{\dagger}$ and Stephen L. Walter ${ }^{*}$

February 2016


#### Abstract

The relative educational returns on colonial - vs. indigenous-language instruction in Sub-Saharan countries have yet to be decisively estimated. This paper, to address this unanswered question, provides an impact assessment of an experiment in Cameroon in which the first 3 years of schooling was conducted in a local language instead of in English. Test results in examinations in both English and Math reveal that treated students exhibit gains of 1.1-1.4 of a standard deviation in Grades 1 and 3 as compared to the control students. By the end of $5^{\text {th }}$ Grade, two years after reverting to the English stream, treated students still exhibit gains of 0.40-0.60 of a standard deviation, though the absolute scores for both groups are low enough to suggest limited learning is taking place.


JEL:I20; I25; I28; O15
Keywords: Cameroon; Development; Language of instruction; Student attainment

[^0]
## 1 Introduction

The objective of increasing access to schooling has been successful in most of the developing world, with gross enrollment rates exceeding 100 percent even in many Sub-Saharan African states. This sharp increase in enrollment rates though has been accompanied by a worrying trend, where student-learning outcomes have either stagnated or even worsened. For instance, results indicate that less than a fourth of grade six children reached the desirable level of reading literacy in Botswana, Kenya, South Africa and Swaziland, and this reduces to fewer than 10 percent in Lesotho, Malawi, Mozambique, Namibia, Uganda and Zambia (SACMEQ II and III). ${ }^{1}$ Similarly in some countries in Sub-Saharan Africa, up to 40 percent of young people who have attended primary schooling for five years have neither the essential skills to avoid lapsing into illiteracy, nor the minimal qualifications to secure a modern sector job (UNECOSOC, 2011).

The growing body of evidence on improving student outcomes, however, has failed to provide clear solutions on how school quality and learning outcomes can be improved. As Banerjee et al. (2007) point out, a number of rigorous randomized evaluations confirm that spending more resources on conventional inputs like textbooks, flip charts, and additional teachers has no discernible impact on children's test scores. Similarly Muralidharan (2013) notes in the context of India "there is very little evidence to support the notion that improving school inputs in a "business as usual' manner will improve learning outcomes". Against this background this paper explores the role of a particular school input, the language of instruction, in improving student learning outcomes in the context of Cameroon, a country, as will become clear when we present our data, where educational outcomes are typically (as in the rest of Sub-Saharan Africa) weak.

[^1]The paper analyses an experimental local language schooling program introduced in the Boyo division of northwestern Cameroon in 2007. The program, sponsored by the SIL International, involved the introduction of instruction in Kom, the local language of the Boyo division, for the first 3 years of primary schooling in 12 experimental schools. ${ }^{2}$ These 12 experimental schools were then matched with 12 comparison schools to form the control group for the study. At the end of the 3 years of treatment the students in the experimental schools reverted back to the standard practice of being instructed solely in English. At the end of each academic year independent evaluators conducted student assessment tests, in Math and English, to assess learning outcomes in the treated and control schools. The data for our study comes from the tests conducted at the end of Grade 1, Grade 3 and Grade 5.

The design of the experiment did not involve randomization but instead matching based on heuristics and other criteria discussed in detail in the next section. Given the potential for bias due to such an experimental design, we explore various ways to determine whether the treated and control students and schools are indeed comparable to each other, and also to the other schools in the region. We show that the treated and control schools exhibit no statistically significant differences on a host of observable characteristics, including on the results of primary school leaving examinations from the pre-intervention year of 2006. Next, using data on socio-economic characteristics such as parents' education, housing quality, and ownership of durables, we show that the treated and control student groups exhibit no statistically significant
${ }^{2}$ SIL International, formerly known as the Summer Institute of Linguistics, engages in a wide range of work involving minority or lesser-known languages. This work includes linguistic research, language documentation, language development (alphabets, dictionaries, grammars, instructional materials), adult literacy, multilingual education, advocacy of minority rights, and the development of a variety of literatures including translations of the New Testament. Because of its extensive work in language documentation, advocacy, and literacy, SIL International has been a consulting agency to UNESCO for the past 20 years.
differences on most socio-economic characteristics.
The assignment of local language instruction is done at the level of the school and hence there are common factors affecting the tests scores of students within the same school. In order to account for the potential interdependence of standard errors of pupils within the same school, we employ a clustered bootstrap methodology and estimate bias-corrected accelerated 95 percent confidence intervals. The estimates suggest that the provision of local language instruction had sizeable effects on Grade 1 and Grade 3 student achievement tests on Math and English amounting to 1.17-1.71 and 0.68-1.53 of a standard deviation, respectively. ${ }^{3}$ The raw scores show that the treated students and control students have an overall score of 52 and 16 percent, respectively in Grade 1, and 45 and 22 percent, respectively, in Grade 3. Thus, not only do the treated students score more than twice as high as the untreated students but also the absolute scores suggest that at least some basic knowledge is being garnered by the treated sample. The Grade 5 data show treatment effects on the overall achievement score of $0.007-0.84$ of a standard deviation. The raw scores however show that the treated and control groups on average score 17 and 36 percent and 18 and 29 percent, in Math and English, respectively. Our favored interpretation of the results is that the student level of knowledge in Grade 5 for both treated and control groups is very low, and any small advantage the treated students possess has little long-term consequence in determining actual human capital. ${ }^{4}$

The available evidence on absolute test scores in the context of countries employing nonindigenous language for schooling, consistent with our evidence, exhibits a bleak picture. Blimpo et al. (2011), in the context of Benin, which uses French as the medium of instruction, find that
${ }^{3}$ The range is based on the 95 percent confidence intervals of the treatment effect.
${ }^{4}$ It is important to note that the evidence suggests that treated students possess no significant advantage on Grade 5 level tests, which however does not rule out the possibility that they might still possess significant advantage in terms of basic numeracy and reading skills that are imparted earlier.
$4^{\text {th }}$ and $6^{\text {th }}$ graders read 24 and 41 words per minute, respectively, whereas international standards suggest that about 45 to 60 words per minute are required for comprehension (Abadzi, 2008). In a study based in Kenya Glewwe et al. (2009) report that the average test score for Grade 6 students in their sample is as low as $37 \%$, again suggesting that most students are unable to master the lion's share of the curriculum. Similarly the Demographic and Health Survey (DHS) data of Cameroon from the year 2011 show that almost $30 \%$ of the students who are coded as having between 4 to 7 years of primary schooling are unable to read a complete sentence. Our finding of large treatment effects and the fading of these treatment effects after reverting to the English stream suggest that local language instruction might be essential for a much longer duration if sustainable gains in student achievement are the goal.

The role of language in affecting student outcomes, in the context of Sub-Saharan Africa, has been highlighted by a small group of educationalists and pedagogues (Alidou et al., 2006) though little systematic quantitative evidence has been brought to bear on this. Some recent papers however explore the importance of the question of language use in education. Eriksson (2014) exploits a language policy change in South Africa, and in line with our findings shows that the provision of two extra years of local language instruction, instead of in English or Afrikaans, had a positive effect on wages, the ability to read and write, on educational attainment, and on the ability to speak English. Ramachandran (2015) using a triple difference-in-differences strategy finds that introduction of mother tongue schooling for the largest ethnic group in Ethiopia in 1994 resulted in increasing the ability to read a complete sentence by 26 percent. Taylor et al. (2013) employing a school fixed-effects model in South Africa find that provision of mother tongue instruction in the early grades significantly improves English acquisition, as measured in grades 4, 5 and 6. Angrist et al. (2008) find that introduction of Spanish as the language of instruction, instead of English after Grade 5, did not affect English language skills in Puerto Rico. The above studies are in line with the findings of a positive (or at least no negative) impact of instruction in a proximate language on second language acquisition. The
above findings, however, are in contrast with those of Angrist and Lavy (1997) who find that the program of 'Arabization' had negative effects on wages and French writing skills in Morocco. However their setting differs from ours in two crucial ways - (i) they analyze extension of secondary schooling in Arabic whereas a large majority of the treated speak Berber as their mother tongue. ${ }^{5}$; (ii) they look at a change that happens at the secondary schooling level whereas the pedagogical literature stresses the importance of instruction in a proximate language in the formative years.

In the context of OECD countries and the United Kingdom, Dustmann et al. (2010) and Dustmann et al. (2012) highlight language as the single most important factor in explaining differences between immigrant and native children's schooling outcomes. Thomas and Collier (2002), Chin et al. (2013) and Slavin et al. (2011) evaluate bilingual education programs in the United States for Hispanic students and find either positive or no impacts of extension of mother tongue instruction.

Laitin and Ramachandran (2015) show that the average distance from the official language has a significant negative impact on the cognitive test scores of students, as well as wage and occupational outcomes. ${ }^{6}$ The choice of language is assumed to affect the process of human capital formation through two distinct channels - the individual's distance to the official language and exposure of the individual to the official language. ${ }^{7}$ Due to their distance from popular speech,

[^2]continued reliance on colonial languages might be responsible for imposing insurmountable learning costs on the school-going population. For instance, despite more than 50 years of using a colonial language as the language for schooling and administration, in more than 25 Sub-Saharan African countries, less than $20 \%$, and often less than $10 \%$, of the population is estimated to be able to speak the colonial language, let alone read and write it (Albaugh, 2014). The results seem to provide further evidence that the language of instruction in schooling might be an important input that has been overlooked in the economic literature analyzing human capital formation and socio-economic development in Sub-Saharan Africa.

## 2 The Kom experimental mother tongue project - Overview

As of 2014, an experimental program in multilingual education has been functioning for seven years in the Boyo Division of Cameroon. The program involves introducing the local language, Kom, as a medium of instruction in classes 1-3 for 12 experimental schools in the region.

### 2.1 Nature of intervention and design

The design of the intervention involved selecting 12 schools in the region which would be allocated local language instruction for three years. The selection of the treatment schools involved the criterion of choosing schools which were perceived by the local education inspectors to be the 'low' performing schools in the region. These 12 selected schools were each matched with a comparison school to provide the relevant counterfactual to be able to assess the effects of the local language program.

The choice of the comparison school was based on what were perceived by the local edGhana, imply that the role of language might be much more important in the developing country context, and might be responsible for the inconclusive findings on the role of language of instruction on student achievement in the United States.
ucation inspectors to be most similar to the treatment schools, and were explicitly guided by three factors - (a) geographic proximity to the matched experimental school; (b) similar size; and (c) similar 'type' of school. By similar size we refer to the total number of students, as well as to the number per classroom. The type refers to whether the school was public or private. And if private, whether it was sponsored by the church or any other organization. The design of the intervention meant that treatment and control status were not randomly allocated but were in fact based on the criteria described above. The results or findings hence need to be interpreted given the imperfect nature of the design. This said it should be mentioned that the treated and control schools exhibit no statistically significant differences on a host of available observable characteristics, including the pre-intervention test scores for the school leaving examinations conducted in Grade 6. ${ }^{8}$ Thus, the heuristic decision making process of the local education inspectors seems to have done a reasonable job. Moreover, tests conducted later also reveal that the other schools in the region, i.e. those that were neither chosen as treatment or control schools, also exhibit no statistically different performance on the $6^{\text {th }}$ Grade school leaving exam results. This in turn suggests that the treated and control schools could be considered representative of the schools in the region.

The intervention lasted for the first three years of primary schooling during which children were instructed in the local language Kom in all subjects and learnt English as a subject. The comparison schools, on the other hand, continued with the normal practice of English-medium instruction. After 3 years of intervention the students in the experimental schools reverted to the standard stream of English-only instruction.

The teachers in the treatment schools were those who were already working there prior to the intervention and no special teachers were hired for the purpose of the project. The teachers

[^3]were, however, provided a two-week training course to be able to teach in the local language, as they had no prior experience in Kom-medium teaching. This corresponds to the normal teacher training provided in the schools following the standard English-only stream.

Students in the 'experimental' or 'treated' schools were to be provided instruction in the local language Kom. As Kom language textbooks were not available on the market, they were provided free of cost to all students at the treatment schools. The control school students, using English language textbooks freely available on the market, were expected to purchase the required textbooks by themselves. As students (or parents) do not often buy textbooks, the nonprovision of textbooks to the control school could potentially introduce bias in the interpretation of the treatment effects. However (Glewwe et al. 2009) in their data from Kenya report that provision of textbooks show small or insignificant effects. One of their proposed explanations for their finding of the failure of the textbooks to have an effect on student achievement is the language of the textbook. If their proposed mechanism turns out to be correct, this would mean that the language of the textbook (or, in our case, the language of instruction) is a near-perfect complement to the provision of textbooks. If English-medium textbooks and English-language medium of instruction are measuring the same thing - i.e. the inability of students to garner much information if provided in their ill-learned supplementary language - the worry about this omitted variable is not a major cause for concern.

The experiment involved testing the students enrolled in the experimental and control schools at the end of each academic year till the end of primary schooling. ${ }^{9}$ At the end of each academic year detailed student assessments, in English and Math, were carried out by independent evaluators for both the treated and control schools. ${ }^{10}$ The tests conducted every year were designed to be compatible with the level of knowledge that should be attained by a student enrolled at the

[^4]current Grade level, as prescribed by the national curriculum. ${ }^{11}$

## 3 Data and identification strategy

### 3.1 Basic framework

The data for the evaluation of the effects of local language instruction is from Grade 1, Grade 3 and Grade 5 student assessment tests in Math and English from the years 2008, 2010 and 2012. ${ }^{12}$ As mentioned before, the treatment consisted of 3 years of local language instruction for the treated students (schools) and then at the end of 3 years the treated students revert back to the standard stream of English-only instruction.

The process of human capital formation is conceptualized as depending upon the cost of learning and represented by:

$$
\begin{equation*}
h_{i}=\alpha_{0}+\beta_{0} C_{i}+\varepsilon_{i} \tag{1}
\end{equation*}
$$

where $h_{i}$ and $C_{i}$ represent the level of human capital and the cost of acquiring human capital, respectively, for student $i$, and $\varepsilon_{i}$, the error term, satisfies the assumption of normality. The cost of learning for an individual $i$ in turn is dependent on the language of instruction and other individual, school and family level factors. We represent the cost function as:

$$
\begin{equation*}
C_{i}=\alpha_{1}+\beta_{1} L\left(d_{i m}, e_{i m}\right)+\mu_{i} \tag{2}
\end{equation*}
$$

[^5]where $L$ captures the cost affected by the language of instruction and $\mu_{i}$ includes all other factors, including school, family and individual level variables, that affect the process/cost of human capital formation. As can be seen in Equation 2, the cost imposed by $L$ is in turn assumed to depend on two factors. First, $d_{i m}$, which captures how distant is the language of schooling $m$ from the first language of the student $i$. Second, $e_{i m}$, which represents what is the exposure of the student $i$ to the medium of instruction $m$ on a day-to-day basis. ${ }^{13}$ The crucial assumption is human capital is increasing the lower the distance and higher the exposure to the medium of instruction in schooling. Now replacing for $C_{i}$ in Equation 1 gives us:
\[

$$
\begin{equation*}
h_{i}=\psi+\varphi L\left(d_{i m}, e_{i m}\right)+v_{i} \tag{3}
\end{equation*}
$$

\]

where $\psi=\alpha_{0}+\beta_{0} \alpha_{1}, \varphi=\beta_{0} \beta_{1}$ and $v_{i}=\varepsilon_{i}+\beta_{0} \mu_{i}$. In our empirical setting we compare students who receive instruction in either the local language Kom or the foreign language English. The treated students who receive the local language instruction are primarily from the linguistic group Kom. There are only 4 treated individuals from an alternative language group. This implies that there is little or no variation in the distance from the medium of instruction within the treated group, and the precise effect of changing language distance on student attainment cannot be retrieved. Similarly, the exposure of the students to the local language does not vary across the treated groups as they are all from very similar socio-economic backgrounds, implying we cannot disentangle the $\varphi$ parameter into its two theoretical components: (i) the effect of distance of students' language from the medium of instruction on student attainment (ii) the effect of increased exposure to the medium of instruction on student attainment. We thus adopt the strategy of estimating the treatment effect through the use of a dummy indicating whether

[^6]the students receive local language instruction or not. ${ }^{14}$
If the allocation of local language instruction is orthogonal to $v_{i}$ (i.e. the host of other factors that affect human capital formation plus the error term), unbiased estimates of local language instruction can be retrieved by estimating Equation 3. The assessment data allows us to evaluate effects of mother tongue instruction, one year after the intervention started, immediately after the intervention ended, and also 2 years after the students reverted back to the Standard English-only stream. The allocation of treatment and control was, as noted before, carried out at the school level and we show below the treated and control schools demonstrate no significant differences on a host of school level characteristics. The data from Grade 5, besides the scores on the achievement tests, also provides detailed information on the socio-economic characteristics of the students enrolled in the control and treated schools. ${ }^{15}$ The additional student level data from Grade 5 allows us to ensure that there are also no systematic differences in the socioeconomic attributes of the students classified as treated and control. This provides evidence that the assumption, $\operatorname{cov}\left[L, v_{i}\right]=0$, is supported by the available data. ${ }^{16}$ The socio-economic variables are collected at the compound level. The compound in the Boyo district usually refers

[^7]to a household but may include two or three houses close together that form part of an extended family. Collection of socio-economic variables at the compound level hence also accounts for positive externalities, due to available resources in extended families, which have been shown to be important in settings such as Cameroon (Angelucci et al. 2010, Loury 2006).

As the data allows us to identify whether the student attended the first 3 years of schooling in the local language or in English, the basic empirical strategy consists of estimating the effect of treatment by comparing the outcomes of students in the two different streams

### 3.2 Accounting for interdependence of errors within clusters

The assignment of treatment was done at the school level and there exists a strong possibility of the standard errors of the student within the same school do not not satisfy the assumption of being independent. The violation of the assumption of independence of errors within clusters will typically lead ordinary least squares (OLS) regression to underestimate standard errors leading to over rejection of the null hypothesis and too many false positives (Moulton, 1990). The typical way to deal with this problem is to compute cluster-robust variance estimates (Arellano 1987, White 2001). The cluster-robust variance estimates (CRVE) control for both error heteroskedascticity across clusters and quite general correlation and heteroskedascticity within clusters, under the condition that the number of clusters is sufficiently large. Hence, when the number of clusters is not large using CRVE might lead to biased estimates of the standard error, and these errors will be typically biased downward.

In order to deal with this we first estimate treatment effects using OLS and cluster errors at the school level, and calculate the 95 percent confidence interval arising from this exercise to serve as a reference point. We then perform a clustered bootstrap with 1000 repetitions that samples the cluster with replacement and present 95 percent confidence intervals pertaining to the "normal-based" (NB) and "bias-corrected accelerated" (BCa) confidence intervals. The
normal-based confidence interval for $\theta$ is calculated as:

$$
\left[\hat{\theta}-1.96 \times \operatorname{se}_{\text {Boot }}(\hat{\theta}), \hat{\theta}+1.96 \times \operatorname{se}_{\text {Boot }}(\hat{\theta})\right]
$$

where the above is a standard Wald asymptotic confidence interval, except that the bootstrap is used to compute the standard errors. The BCa confidence interval in turn makes three additional adjustments to the normal-based confidence interval. First, it uses the relevant percentiles of the empirical distribution of the B (here 1000) bootstrap estimates, and has the advantage of the confidence interval around $\hat{\theta}$ being asymmetric and invariant to monotonic transformations of $\theta$. Second, it incorporates a bootstrap estimate of the finite-sample bias in $\hat{\theta}$. Finally, it adds an acceleration component that permits the asymptotic variance of $\hat{\theta}$ to vary with $\theta$. Thus, the BCa confidence interval has the theoretical advantage of offering an asymptotic refinement, which for instance the methodology used in Bertrand et al. (2004) does not. ${ }^{17}$

### 3.3 Dealing with selection issues

One of the confounding factors in the estimation of treatment effects for the later grades (Grade 3 and 5) is attrition in the data. Table I shows the sizes of the original cohort by treatment status, and their attrition level by Grade 3 and 5. As can be seen there were a total of 323 treated students in Grade 1. Out of these 166 proceed to Grade 3, and finally 85 are left in Grade 5. In case of the control students, 335 students were present in Grade 1. Out of these 100 are still present in Grade 3, and finally 39 still in Grade 5. The level of attrition for the treated by Grade
${ }^{17}$ The BCa confidence interval and bootstrap- $t$ (also known as the percentile- $t$ method) confidence interval provide the same asymptotic refinement but the BCa is invariant to monotonic transformations of $\theta$, whereas the bootstrap-t confidence interval is not. Otherwise there is no strong theoretical reason to prefer one method to the other. For further details regarding the bootstrap confidence intervals refer to Cameron et al. (2008) and Cameron and Trivedi (2009).

Table I: Level of attrition by treatment status

|  | No. of <br> Treated | Percentage of <br> Attrition for the Treated | No. of <br> Untreated | Percentage of <br> Attrition for the Untreated |
| :--- | :---: | :---: | :---: | :---: |
| Present in Grade 1 | 323 | . | 335 | .. |
| Present in Grade 3 | 166 | $49 \%$ | 100 | $70 \%$ |
| Present in Grade 5 | 85 | $74 \%$ | 39 | $88 \%$ |

3 is 49 percent whereas for the control group attrition by Grade 3 is equal to 70 percent. By the end of Grade 5, 74 and 88 percent of the original cohort of the treated and control group are no longer present in the data. Given the levels of observed attrition it is important to understand how it might bias our estimation of treatment effects for Grades 3 and 5. First, it is important to note that though the levels of attrition in Table I might seem inordinately high, they need to be compared with the survival rate in primary schooling in Cameroon to be able to put them in context. The data shows that the survival rate to Grade 5 in Cameroon in 2007, the year the program started, was 62 percent. ${ }^{18}$ Similarly the percentage of repeaters in Grade 5 is around 20 percent. These would imply that if we would start off with an original cohort size of 100, based on the national repetition and survival rates, only 49 of the original students should be still present in Grade 5, or in other words attrition levels of around 50 percent. Moreover, the above figures on repetition and survival are based on country level averages, whereas our setting is in a rural area. Given the well-known gaps in outcomes between rural and urban areas in educational attainment in Sub-Saharan Africa (Zhang, 2006), it is reasonable to assume that combining the repetition and survival rates for rural areas would generate attrition levels of around 65 percent by Grade $5 .{ }^{19}$ As we track students only if they are present in these 24 schools, the 15 percentage point discrepancy between the national attrition rate and that in our

[^8]data could be due to the students moving to other schools who we are consequently unable to track.

Comparing attrition rates across the treated and untreated group in Table I allows us to establish that the levels of attrition are higher for the control rather than the treated group. Concerning the direction in which selection is working, a plausible explanation is that the students who are disappearing from the data are drop-outs, and the sequence is that the worst performing students are the first ones to drop out (i.e. the lower end of the ability distribution). As we have longitudinal data on the students, we can calculate the test scores of students by attrition status. In Table II are shown the test scores of the students in Grade 1 tabulated by when they disappear from the data. ${ }^{20}$ The treated students, who are observed in Grade 1, but not in Grade

Table II: Test scores and attrition by treatment status

|  | No. of <br> Treated | Overall Score of <br> Treated <br> in Grade 1 | No. of <br> Untretaed | Overall Score <br> of Untreated <br> in Grade 1 |
| :--- | :---: | :---: | :---: | :---: |
| Present in Grade 1 but not in Grade 3 or 5 | 153 | 42.78 | 230 | 13.60 |
| Present in Grade 1 and 3 but not in Grade 5 | 85 | 58.52 | 64 | 19.40 |
| Present in Grade 1, 3 and 5 | 85 | 63.15 | 39 | 26.19 |
|  |  |  |  |  |
| TOTAL | 323 | 52.31 | 335 | 16.12 |

The scores are out of a total possible maximum of 100 points.

3 or 5, have an average overall tests score of 42.78 percent in Grade 1. Those who are present in Grade 1 and 3, but not in Grade 5, have an average overall test score of 58.52 percent, and those present in Grade 1, 3 and 5 have an average test score of 63.15 , in Grade 1. Similarly for the control group the test scores in Grade 1 corresponding to students observed in Grade 1 but not in Grade 3 or 5, those observed in Grade 1 and 3 but not in Grade 5, and those present in Grade 1, 3 and 5, are 13.60, 19.40 and 26.19 percent, respectively.

[^9]Table II hence shows, first, that the trend presumed by the assumption of the worst-off students tending to disappear (and interpreted as dropping out) is indeed borne out in the data. Second, even the worst-off treated students perform substantially better than the best-off control students. Finally, from Table I we know the rate of attrition is higher for the control than the treated students. Assuming that we start off with identical ability distributions, the extra attrition happening in the control schools implies that the relative proportion of lower ability students from the treated group are overrepresented. The analysis in this subsection suggests that one, the levels of attrition are in line with the national average if we assume that some students change schools who we are unable to track, and two, if our estimates are biased, they are downwardly so. Thus if anything, we underestimate the effect of local language instruction on test scores.

## 4 Results

### 4.1 Comparison of treated and control schools

Table III shows the comparison of the treated and control schools on a set of school level characteristics. There are no systematic differences in terms of total number of students, toilet student ratios, availability of library or playground, years of operation, state of school and school building material. The available characteristics of teacher skills - years of education, years of experience and a subjective rating of teachers by the local education inspectors - also show no statistically significant differences. The only two variables which show statistically significant differences are the availability of a toilet and primary school building material, and in these the control schools are slightly better off.

Table III also reports the 2-group Hotelling's T-squared statistics, as well as the joint Fstatistic, with the null hypothesis that the group means for the various covariates considered are equal. As can be seen, the null hypothesis cannot be rejected at reasonable significance levels

Table III: Baseline Group Comparison on School Characteristics - Test of Means

| Variable | Control <br> School <br> Mean | Treated <br> School <br> Mean | Diff | p-value |
| :--- | :---: | :---: | :---: | :---: |
| Total Students | 163.08 | 153.33 | 9.75 | 0.80 |
| Toilet Dummy | 1 | 0.75 | 0.25 | 0.07 |
| Separate Toilet For Girls Dummy | 0.58 | 0.67 | -0.08 | 0.69 |
| Toilet Student Ratio | 99.46 | 102.03 | -2.57 | 0.93 |
| Library Dummy | 0 | 0 | 0 | . |
| Playground Dummy | 0.75 | 0.83 | -0.08 | 0.63 |
| Roof | 2 | 2 | 0 | . |
| No. of Rooms | 6.83 | 6.58 | 0.25 | 0.80 |
| No. of years of operation | 24.5 | 28.5 | -4 | 0.62 |
| Primary building material | 4.33 | 3.5 | 0.83 | 0.04 |
| State of school | 1.5 | 1.75 | -0.25 | 0.36 |
| Avg. years of educ. of Teachers | 13.35 | 14.03 | -0.68 | 0.22 |
| Avg. Years of exper. of Teachers | 12.38 | 14.69 | -2.32 | 0.42 |
| Subjective Rating of Teachers | 4.56 | 4.54 | 0.01 | 0.92 |
| Mean primary school leaving score | 159.02 | 158.59 | 0.43 | 0.97 |
| 2-group Hotelling's T-squared $=30.986956$ |  |  |  |  |
| F test statistic: $((22-13-1) /(22-2)(13)) \times 30.986956=.95344481$ |  |  |  |  |
| H0: Vectors of means are equal for the two groups |  |  |  |  |
| $F(13,8)=0.9534$ |  |  |  |  |
| Prob $>F(13,8)=0.5495$ |  |  |  |  |

12 control and treated schools are used to calculate the averages shown above in Table III; except for the variable mean primary school leaving score which is available for 10 and not 12 control schools.
providing us confidence that the two sets of schools are indeed comparable.
The data also provides the average tests scores at the school level for the $6^{\text {th }}$ Grade nationally organized primary school leaving examinations for a set of 12 treated, 10 control and 80 'other' schools. 'Other' schools refer to schools that were neither treatment nor comparison schools. These are pre-intervention data from the year 2006, and correspond to the year before the local language program initiative was implemented. Primary school leaving scores might be considered as the output arising from the combination of various inputs that schools possess and might be considered as a proxy for other unobservable school inputs that are not captured by the available school-level characteristics. Reassuringly, as can be seen in the last row of Table III, there is no difference between the mean school leaving scores of the control and treated
groups.
Table IV regresses the mean school leaving score on a set of categorical dummies which

Table IV: Comparing schools allocated to treatment, control and not participating in the preintervention primary school leaving exam results

Dependent variable - Average School Performance
on Primary School Leaving Exams on Primary School Leaving Exams
(1)
(2)

| Experimental Schools | -2.763 | -0.506 |
| :--- | :---: | :---: |
|  | $(8.519)$ | $(9.057)$ |
| 'Other' Schools | -0.167 | -0.208 |
| School Type Dummies | $(6.074)$ | $(6.314)$ |
| Observations | No | Yes |
| R-squared | 102 | 102 |
| Average of dependent variable | 0.001 | 0.041 |
|  |  |  |
|  | 159.70 | 159.70 |

The baseline omitted category are the control schools. The omitted school type is the Cameroon Baptist Convention School. $* p<.10 ; * * p<.05 ; * * * p<.01$. Robust SE's in parenthesis. The dependent variable is the mean score for the sixth Grade school leaving examinations at the school level.
indicate whether the school was experimental, comparison or 'Other' and a set of dummies indicating the type of school, and estimates:

$$
\begin{equation*}
\text { Mean FSLE } \text { Score }_{i}=\text { Program Status }_{i}+\text { School Type }_{i}+\varepsilon_{i}, \tag{4}
\end{equation*}
$$

where Mean FSLE Score $_{i}$ is the average school leaving score for the $6^{\text {th }}$ Grade examination in school i, Program Status indicates whether the school was an experimental, control or an 'other' school. The School Type $i_{i}$ indicates whether the school $i$ is a government, Catholic, Presbyterian, Baptist or Islamic school.

The results in Table IV show the coefficient on both categories i.e. experimental and other schools are very close to zero with p-values in the range of 0.95 . The results show that not only
are the treated schools very similar to the chosen control schools in terms of primary school leaving results but also are very similar to the other schools in the region.

### 4.2 Comparison of treated and control students

Table V: Baseline Group Comparison on Student Characteristics - Test of Means

| Variable | Control <br> Students Observations | Control <br> Students <br> Mean | Treated <br> Students Observations | Treated <br> Students <br> Mean | Diff | p-value |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Years of educ of Mother | 26 | 4.65 | 49 | 4.04 | 0.61 | 0.33 |
| Years of educ of Father | 27 | 5.30 | 63 | 5.81 | -0.51 | 0.56 |
| Age | 35 | 10.83 | 76 | 10.97 | -0.15 | 0.64 |
| Student has a cellphone | 36 | 0.06 | 85 | 0 | 0.06 | 0.03 |
| Compound has a cellphone | 35 | 0.60 | 84 | 0.83 | -0.23 | 0.01 |
| Compound has a radio | 36 | 0.72 | 84 | 0.71 | 0.01 | 0.93 |
| Compound has a television | 36 | 0.56 | 84 | 0.33 | 0.22 | 0.02 |
| Compound has a motorcycle | 36 | 0.33 | 84 | 0.38 | -0.05 | 0.62 |
| Compound has a car/truck | 36 | 0.44 | 83 | 0.46 | -0.01 | 0.89 |
| Compound has a refrigerator | 36 | 0.33 | 84 | 0.13 | 0.20 | 0.01 |
| Compound has a gas stove | 36 | 0.31 | 84 | 0.29 | 0.02 | 0.83 |
| Someone in the compound have a business | 36 | 0.44 | 84 | 0.52 | -0.08 | 0.43 |
| Someone in the compound have a govt. job | 36 | 0.39 | 83 | 0.30 | 0.09 | 0.35 |
| Compound has a cement floor | 36 | 0.53 | 84 | 0.55 | -0.02 | 0.84 |
| Compound has a metal roof | 36 | 0.61 | 84 | 0.48 | 0.13 | 0.18 |
| Compound has a toilet | 36 | 0.39 | 84 | 0.23 | 0.16 | 0.07 |
| Compound has electricity | 36 | 0.36 | 84 | 0.24 | 0.12 | 0.17 |

2-group Hotelling's T-squared $=33.056102$
F test statistic: $((61-16-1) /(61-2)(16)) \times 33.05=1.54$
H 0 : Vectors of means are equal for the two groups
$\mathrm{F}(16,44)=1.5408$
Prob $>F(16,44)=0.1283$
The above is based on the comparison of treated and control students from the original cohort enrolled in the 24 schools in 2011/12.

Information on socioeconomic characteristics of the students was collected in the year 201112 and is available only for the students who were still present in these 24 schools in Grade 5, and not the entire original cohort of 658 students. As noted in Table I there is very high attrition after Grade 1 and only 26 and 12 percent, respectively, of the treated and control students are still present after four years of schooling in Grade 5. At the student level the socioeconomic characteristics collected include education of father and mother; age; does someone in the compound own a cellphone, television, radio, motorcycle, car or truck, refrigerator, gas
stove; whether someone in the compound owns a business or holds a government job; and does the compound have a cement floor, metal roof, toilet (as opposed to an open pit/outhouse) and electricity.

The results of the comparison of means in Table V show there are no statistically significant differences on most available socioeconomic characteristics. The 2-group Hotelling's T-squared statistics, as well as the joint F-statistic, show that the null hypothesis of the equality of the group means cannot be rejected. The only variables for which there are statistically significant differences between the treated and comparison students are - whether (i) student has a cellphone (ii) compound has a cellphone (iii) compound has a television (iv) compound has a refrigerator (v) compound has a toilet. Except for the variable "whether compound has a cellphone", in all the categories the control group does better. This is consistent with the evidence presented in Section 3.3 which argues that the proportion of lower ability students from the treated group are overrepresented. If ability is correlated with socioeconomic characteristics, we should expect that the higher attrition in the control group should bias the characteristics in favor of the control group students; the comparison of covariates in Table V shows that this is indeed the case.

The extent of observed differences in the available covariates is consistent with what might be expected to arise even from a design which relies on random assignment of treatment and control status. Thus overall the evidence seems to support the assumption that assignment of treatment status is orthogonal to family, school and individual level factors.

### 4.3 Estimates of the effect of local language instruction

### 4.3.1 Kernel density plots

Figure I plots the kernel density of the standardized overall test scores by treatment status for Grades 1, 3 and 5. The overall raw scores and the standardized test score corresponding to the plotted curves, as well as the scores on English and Math for the treated and control students in

Grades 1, 3 and 5, are shown in Table A1 in the appendix.
The treated and control students in Grade 1 have an overall raw score of 52 and 16 percent, respectively. In Panel A, the kernel density plot of test scores of the treated students in Grade 1 is sharply shifted to the right and is centered at much higher mean ( 0.73 vs -0.71 ). Panel B which


Figure I: Kernel density of standardized overall test scores in Grade 3 by treatment status plots the kernel density of the standardized overall test scores in Grade 3, for the treated and control students, exhibits a very similar picture to the one shown in Panel A. The kernel density of the treated students is sharply shifted to right with the mean centered at 0.43 compared to a mean of -0.68 for the control students. Examining the raw scores shows that the treated students overall score ( 45.27 out of 100 ) is nearly double that of the control students. In Section 4.3.2 we show that these differences shown in Panel A and B are statistically significant, and sizeable in magnitude.

The kernel density plots corresponding to the standardized overall tests scores of students
in Grade 5 by treatment status is shown in Panel C. The kernel density plots seem to almost overlap, however exhibiting a small advantage in favor of the treated students. The raw scores however show that the treated and control students score on an average 28 and 24 percent, respectively. Sadly, not even the best student in our data scores 50 percent. This suggests that the level of learning is very low and these students might lapse back into illiteracy in a few years. The raw gaps are around 20 percent compared to the 100 percent noted in Grade 3, though still significant.

### 4.3.2 Regression estimates of the effect of local language instruction

The regressions estimated are motivated by the schematic framework of student learning presented in Section 3.1, and the reduced form equation estimated is given by:

$$
\begin{equation*}
\text { Score }_{i j k}=\alpha+\varphi \text { Treated }_{i j}+v_{i j} . \tag{5}
\end{equation*}
$$

Score $_{i j k}$ refers to the test score on the overall standardized achievement test in Math and English of student $i$ from school $j$, in Grade $k$. Treated ${ }_{i j}$ is a dummy indicating whether student $i$ in school $j$ was part of the experimental program or part of the normal program in a control school during the first 3 years of schooling, and $v_{i j}$ is the error term. $\varphi$, the coefficient of interest, captures by how many standard deviations does the test score increase due to the provision of local language instruction.

The results are shown in Table VI; in column (1) the dependent variable is the standardized overall test score in Grade 1 and presents the results from an OLS, where the standard errors are clustered at the level of the school. The 95 percent confidence interval on the two treatment dummies shows that local language instruction after the first year increased the overall test score by 1.17-1.71 of a standard deviation, with a mean effect of 1.44 of a standard deviation. Column (2) again considers the standardized overall test score in Grade 1 as the dependent

Table VI: Effect of local language instruction on standardized overall test score in Grade 1, 3 and 5

|  | $(1)$ | $(2)$ | $(3)$ | $(4)$ | $(5)$ | $(6)$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Std. Overall | Std. Overall | Std. Overall | Std. Overall | Std. Overall | Std. Overall |
|  | Score - Grade 1 | Score - Grade 1 | Score - Grade 3 | Score - Grade 3 | Score - Grade 5 | Score - Grade 5 |
| Treated | $1.44^{* * *}$ | $1.44^{* * *}$ | $1.11^{* * *}$ | $1.11^{* * *}$ | $0.42^{*}$ | $0.42^{*}$ |
|  | $(0.13)$ | $(0.13)$ | $(0.20)$ | $(0.20)$ | $(0.24)$ | $(0.24)$ |
|  | $\{1.17-1.71\}$ | $\{1.18-1.70\}$ | $\{0.68-1.53\}$ | $\{0.72-1.49\}$ | $\{-0.079-0.93\}$ | $\{-0.056-0.90\}$ |
|  |  | $[1.18-1.71]$ |  | $[0.72-1.50]$ |  | $[-.03-0.93]$ |
| Observations | 658 | 658 | 266 | 266 | 124 | 124 |
| R-squared | 0.518 | 0.518 | 0.290 | 0.290 | 0.041 | 0.041 |
| Colun |  |  |  |  |  |  |

Columns (1) and (3) and (5) presents results from estimating an OLS regression with errors clustered at the level of the school. Columns (2), (4) and (6) present results from running a clustered bootstrap with 1000 repetitions. The normal-based and bias controlled accelerated 95 percent confidence interval are presented in braces and square brackets, respectively. The standard errors are shown in the parenthesis. $* p<.10 ; * * p<.05 ; * * * p<.01$.
variable, but now however we estimate a clustered bootstrap with 1000 repetitions. It can be seen that the standard errors on the treatment dummy in columns (1) and (2) are very similar. The bias corrected accelerated ( BCa ) confidence interval is also shown in column (2), which as discussed in section 3.2 has the theoretical advantage of providing an asymptotic refinement that other bootstrapped confidence intervals do not. The 95 percent BCa confidence interval shows that local language instruction for one year increased the overall test score in Grade 1 by 1.18-1.71 of a standard deviation. In fact the bias correction procedure suggests that the OLS might be underestimating the effect of local language instruction.

Columns (3) and (4) now consider the standardized overall test score from Grade 3, as the dependent variable. Column (3) presents results of an OLS regression where errors are clustered at the level of the school, and column (4) provides the results arising from the clustered bootstrap procedure. The results from the clustered OLS in column (3) show a treatment effect of 1.11 of a standard deviation, with the associated 95 percent confidence interval being 0.68 1.53. Accounting for potential interdependence of error terms of the students within the same school has no appreciable effect on the results. The 95 percent bias corrected accelerated (BCa)
confidence interval estimate is $0.72-1.50$.
In columns (5) and (6) the dependent variable is the standardized overall test score in Grade 5. The OLS estimate of the treatment effect is 0.42 of a standard deviation, with the lower bound of the 90 percent confidence interval being 0.007 , whereas the 95 percent confidence interval does not exclude zero. ${ }^{21}$ In column (6), we estimate a clustered bootstrap with 1000 repetitions, the 95 percent BCa confidence intervals are again seen to include a zero, whereas the 90 percent BCa confidence intervals (not shown) are $0.02-0.83$.

### 4.3.3 Robustness tests

The results in Table VI show that provision of schooling in a proximate language increased the average tests scores by more than a standard deviation in Grades 1 and 3, and by around 0.40 of a standard deviation in Grade 5. As discussed in Section 2.1 the allocation of control and treatment status was not randomly assigned, but implemented through a heuristic matching procedure. Tables III, IV and V show that the treated and control schools and students are balanced along the available set of characteristics. In Table VII we examine the sensitivity of our results to including the available school and student level characteristics as controls, where the estimation is based on a clustered bootstrap with 1000 repetitions. Columns (1), (2) and (3) consider the standardized overall test score in Grades 1, 3 and 5, respectively, but now additionally control for the standardized primary school leaving exam scores at the school level. ${ }^{22}$ Examining the 95 percent BCa confidence intervals or the point estimate shows that controlling for primary school leaving average exam score does not affect neither the significance nor the magnitude of the treatment effect (Grades 1 and 3), whereas in the case of Grade 5 it increases

[^10]|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Std. Overall Score Grade 1 | Std. Overall Score Grade 3 | Std. Overall Score Grade 5 | Std. <br> Overall <br> Score <br> Grade 1 | Std. Overall Score Grade 3 | Std. <br> Overall <br> Score <br> Grade 5 | Std. Overall Score Grade 1 | Std. Overall Score Grade 3 | Std. Overall Score Grade 5 |
| Treated | $\begin{gathered} 1.44^{* * *} \\ (0.13) \\ \{1.19-1.69\} \\ {[1.15-1.66]} \end{gathered}$ | $1.14 * * *$ $(0.22)$ $\{0.71-1.56\}$ $[0.71-1.59]$ | $0.65 * * *$ $(0.22)$ $\{0.22-1.09\}$ $[0.24-1.11]$ | $\begin{gathered} 1.56 * * * \\ (0.16) \\ \{1.23-1.88\} \\ {[1.12-1.81]} \end{gathered}$ | $\begin{gathered} 1.18 * * * \\ (0.41) \\ \{0.37-1.99\} \\ {[.31-1.89]} \end{gathered}$ | $\begin{gathered} 0.44^{*} \\ (0.26) \\ \{-0.067-0.95\} \\ {[-0.16-0.89]} \end{gathered}$ | $\begin{gathered} 1.48 * * * \\ (0.16) \\ \{1.16-1.79\} \\ {[1.16-1.79]} \end{gathered}$ | $1.20^{* * *}$ $(0.38)$ $\{0.45-1.94\}$ $[0.27-1.82]$ | $\begin{gathered} 0.52^{*} \\ (0.29) \\ \{-0.050-1.10\} \\ {[-0.10-1.07]} \end{gathered}$ |
| Standardized school leaving exam scores | $\begin{gathered} 0.17 * * \\ (0.067) \\ {[0.037-0.30]} \end{gathered}$ | $\begin{gathered} -0.029 \\ (0.12) \\ {[-0.26-0.21]} \end{gathered}$ | $\begin{gathered} -0.19^{*} \\ (0.10) \\ {[-0.39-0.010]} \end{gathered}$ |  |  |  |  |  |  |
| Years of educ of Father |  |  |  | $\begin{gathered} 0.022 \\ (0.019) \\ {[-0.016-0.060]} \end{gathered}$ | $\begin{gathered} -0.011 \\ (0.037) \\ {[-0.083-0.061]} \end{gathered}$ | $\begin{gathered} 0.0038 \\ (0.026) \\ {[-0.047-0.054]} \end{gathered}$ |  |  |  |
| Assets | No | No | No | No | No | No | Yes | Yes | Yes |
| Observations | 586 | 253 | 119 | 90 | 85 | 90 | 117 | 111 | 117 |
| R-squared | 0.530 | 0.287 | 0.116 | 0.613 | 0.243 | 0.046 | 0.617 | 0.342 | 0.193 |

R-squared 10.50 variable is the standardized overall score in Grade 3; and in Columns (3), (6) (9) the dependent variable is the standardized overall score in Grade 5.The normal-based and bias controlled accelerated 95 percent confidence interval are presented in braces and square brackets, respectively. The standard errors are shown in the parenthesis. The list of assets include; ownership of a personal cellphone; whether the compound has a cellphone, radio, television, motorcycle, car or truck, refrigerator, gas stove; whether someone in the compound owns a business or holds a government job; and whether the compound has a cement floor, metal roof and access to electricity. $* p<.10 ; * * p<.05 ; * * * p<.01$.
the magnitude of the treatment effect with the BCa 95 percent confidence interval now excluding zero. Columns (4), (5) and (6) control for the years of education of the father ${ }^{23}$; consistent with the reasoning that the characteristics are balanced across the treated and control group, controlling for the years of education of father again has a negligible effect on the magnitude and significance of the treatment effect found in Grades 1, 3 and 5. Finally, columns (7), (8) and (9) control for a list of items: ownership of a personal cellphone; whether the compound has a cellphone, radio, television, motorcycle, car or truck, refrigerator, gas stove; whether someone in the compound owns a business or holds a government job; and whether the compound has a cement floor, metal roof and access to electricity. The coefficient on the treatment dummy, as well as the 95 percent BCa confidence intervals show that accounting for these assets and socioeconomic characteristics does not change our results.

### 4.3.4 Examining the impact of attrition on treatment effects

The high level of observed attrition in the data raises concerns that the estimation of treatment effects in Grade 3 and 5 might be biased due to the presence of selection effects. The evidence presented in Section 3.3 however suggests that attrition might be downwardly biasing our estimates. We now provide further suggestive evidence that the effects of local language instruction estimated for Grades 3 and 5 cannot be attributed to selective attrition. Selective attrition being the underlying reason for the observed impact of local language instruction would imply that the better students (or academically higher achievers) from the treated group survive at a higher rate than from the control group. One way to test if this is indeed the case is to restrict the estimation of treatment effects in Grades 1 and 3 to the sample of students who survive up to Grade 5. Positive selection operating in the treated group should imply that the magnitude of

[^11]the treatment effect on this subsample of positively selected students should be larger than those found in Table VI. The results of estimating the treatment effects on the sample of treated and control students who survive up to Grade 5 is shown in Table VIII.

Table VIII: Effect of local language instruction on standardized overall test score in Grade 1and 3 - Sample of students present in Grade 5

|  | $(1)$ | $(2)$ | $(3)$ | $(4)$ |
| :--- | :---: | :---: | :---: | :---: |
|  | Std. Overall | Std. Overall | Std. Overall | Std. Overall |
|  | Score - Grade 1 | Score - Grade 1 | Score - Grade 3 | Score - Grade 3 |
| Treated | $1.51^{* * *}$ | $1.51^{* * *}$ | $1.20^{* * *}$ | $1.20^{* * *}$ |
|  | $(0.15)$ | $(0.15)$ | $(0.31)$ | $(0.30)$ |
|  | $\{1.21-1.81\}$ | $\{1.22-1.80\}$ | $\{0.56-1.84\}$ | $\{0.61-1.79\}$ |
|  |  | $[1.14-1.76]$ |  | $[0.60-1.79]$ |
| Observations | 124 | 124 | 118 | 118 |
| R-squared | 0.563 | 0.563 | 0.283 | 0.283 |

Columns (1) and (3) presents results from estimating an OLS regression with errors clustered at the level of the school. Columns (2) and (4) present results from running a clustered bootstrap with 1000 repetitions. The normal-based and bias controlled accelerated 95 percent confidence interval are presented in braces and square brackets, respectively. The standard errors are shown in the parenthesis. $* p<.10 ; * * p<.05 ; * * * p<.01$.

In columns (1) and (2) the dependent variable is the standardized overall test score from Grade 1, where column (1) employs an OLS estimator and column (2) a clustered bootstrap with 1000 repetitions. Comparing the coefficient on the treatment dummy in columns (1) and (2) of Table VI and VIII and conducting a formal test of equality of coefficients shows that the null hypothesis of equality is not rejected at conventional significance level $(z=0.35)$. Similarly, comparing the magnitude of the treatment effect on Grade 3 achievement tests across columns (3) and (4) of Table VI and VIII shows that the null hypothesis of equality of coefficients cannot be rejected $(z=0.24)$. The above discussion shows that the impact of local language instruction cannot be attributed to selective attrition resulting in positive selection among the treated group.

## 5 Discussion and methodological concerns

The results presented in the previous section show large treatment effects of around 1.2-1.3 of a standard deviation on the overall test score in Grade 3, and around $0.40-0.70$ of a standard deviation by the end of $5^{\text {th }}$ Grade. To be able to interpret the importance or magnitude of these treatment effects we need to look not just at the relative performance but also the absolute performance of the students. ${ }^{24}$

In Grade 1 the treated students obtain an average of 52, 45 and 62 percent in oral English, Math and reading (in Kom), respectively, and in Grade 3 obtain an average score of 50, 41.5 and 65 percent in Math, English and Kom, respectively. On the other hand, the control students obtain scores of 35, 10 and 8.5 percent in Oral English, Math and reading (in English), respectively, and in Grade 3 obtain scores of 22 and 20 percent in Math and English, respectively. The scores thus indicate that the control students are learning little or nothing as with multiple choice questions even random guessing should result in an average score of 25 . Given this context the high level of attrition observed for the control students is not surprising. The treated students, if not performing satisfactorily, at least demonstrate a minimum level of comprehension scoring more than 50 percent in the two grades on an average.

In Grade 5 the raw scores suggest a bleaker scenario. The treated students score 17.37 and 35.6 percent in Math and English, respectively. On the other hand control students score 17.61 and 29.36 percent in Math and English, respectively. Though we find statistically significant differences between the two groups, the raw scores show that both treated and control students are learning very little. Assuming a minimum body of knowledge is required for it to be useful, given the absolute raw scores, the small relative advantage of the treated students might be of

[^12]no consequence in terms of actual human capital formation.
The data on the whole shows that students in the English stream are passing through their grades effectively accumulating no knowledge or skills in the subject matter being taught. On the other hand the treated students seem to be performing significantly better and accumulating useful knowledge at least while they are being instructed in the local language, i.e. until the end of Grade 3. The results from Grade 5 suggest that the switch in the medium of instruction implies that the treated students also effectively are learning little in the classroom, though their scores exhibit a small statistically significant advantage relative to the control students. The test scores on the achievement tests seem to indicate that the educational system is producing a generation of students, many of whom are in danger of lapsing back into illiteracy.

The results also shed light on the debate regarding early-vs-late exit local language instruction programs. Early-exit programs refer to programs where instruction in the first language is only provided for a period of 1-3 years. Late-exit programs, however, involve teaching through the first language for a period of at least 6-8 years before switching to the second/foreign language. Existing evidence from developed countries however suggest that late-exit programs have higher and longer lasting effects on minority student achievement, whereas early-exit programs like the one in Cameroon are seen to be too short for individuals to shift skills from the first to the second language (Benson 2004, Cummins 1979, Thomas and Collier 2002). The finding that once the treated students revert back to English instruction there is a the steep fall and convergence in test scores with the control group, points to the fact that local language instruction for only 3 years may be too little for individuals if sustainable gains in student learning is the objective. Thus our data are consistent with those advocating for late-exit programs.

Language of instruction assumes an even more important role in a setting such as Cameroon due to the low levels of exposure to the colonial language on a day-to-day basis, and the teach-
ers themselves having less than perfect command of the dominant colonial language. ${ }^{25}$
The interpretation of the above results as being the effect of the change of language of instruction can only be made subject to the caveats noted in the design of the experiment in Section 2.1. The allocation of treatment and control as noted before was made on the basis of heuristic matching rather than randomization. The tests shown in Table III, IV and V however suggest that both schools and students which were assigned to treatment and control look very similar to each other.

As regards the potential upward bias introduced due to provision of textbooks only to the treatment group, it is important to note that we are not claiming textbooks are unimportant for student performance, but that we need to distinguish between availability and use (Moulton, 1994). If positive effects of provision depend upon utilization, and utilization, as suggested by Glewwe et al. (2009) and All Children Reading (2013), in turn depends on language appropriate instruction material, then language of instruction is a perfect complement to the provision of textbooks.

A final question: is anyone getting educational returns in this English-medium environ-

[^13]ment? Banerjee and Duflo (2011) argue that school systems remain elitist in many developing countries today. The fact that the organization and curriculum demonstrate a large amount of continuity from the colonial past, where the system was intended for a small local elite, often make them unsuitable and too demanding for first generation learners. The results presented here suggest that not only the curriculum but also the language of instruction, a relic from the colonial past, might be favoring a tiny elite at the expense of the large majority. As Albaugh $(2014,35)$ argues "in an environment of stable borders and low direct taxation, African leaders were motivated neither to spread a common language throughout their territories nor to elevate local languages out of a sense of nationalist pride. This resulted in little attention to education in general and the maintenance of colonial precedents in the specifics of language medium, except in unusually ideologically driven instances." The implications of this continuity of the colonial policy can be understood by examining the objectives of the education policy of the colonists. Despite differing influence and ideologies of the catholic and protestant missionaries, the overarching educational objective of the colonizers was identical - training a few elites to help administer the country with little concern for a productive and knowledgeable workforce (Fabunmi 2009, Whitehead 2005, Bokamba 1984). Indeed, secondary education barely reached 3 percent of Africas school-aged population at independence (Mitchell, 2003). The convergence of the education policy of the different colonists and the uninterrupted continuation of this policy can be seen by the fact that whatever the language policy in primary education, secondary education exclusively was, and still is, provided only in the colonial language. In sum, educational policies in independent Africa, as was the case in the colonial era, serve the interests of a small elite and fail to provide intellectual capital broadly.

## 6 Conclusion

Investment in human capital has long been recognized as a sine qua non for economic development. Research in development economics has, however, inadequately provided guidance on how best to invest in human capital, especially for young citizens in poor societies. Taking advantage of a promising (though imperfect) intervention in primary school education, data from a project administered by the SIL International reveal a substantial positive effect of mothertongue medium of instruction in Cameroon. The findings suggest that in Africa, where primary school teachers have quite limited capacity in the former imperial languages, educational achievement in schools relying on these languages as media of instruction is shockingly low. In contrast, education at the primary level through local media of instruction reveals gains not only in math, but also (when taught as a subject) in imperial language proficiency. The results from this intervention, if upheld in better identified treatments, suggest a radical redirection of educational funding in Africa.

## References

Abadzi, H. (2008). Efficient learning for the poor: New insights into literacy acquisition for children. International Review of Education 54(5-6), 581-604.

Albaugh, E. A. (2014). State-Building and Multilingual Education in Africa. Cambridge University Press.

Alidou, H., A. Boly, B. Brock-Utne, Y. S. Diallo, K. Heugh, and H. E. Wolff (2006). Optimizing learning and education in Africa-the language factor. Paris: ADEA.

All Children Reading (2013). Teaching and learning materials. Technical report, Retrieved at http://www.allchildrenreading.org/teaching-and-learning-materials.

Angelucci, M., G. De Giorgi, M. A. Rangel, and I. Rasul (2010). Family networks and school enrolment: Evidence from a randomized social experiment. Journal of Public Economics 94(3), 197-221.

Angrist, J., A. Chin, and R. Godoy (2008). Is Spanish-only schooling responsible for the Puerto Rican language gap? Journal of Development Economics 85(1), 105-128.

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

Arellano, M. (1987). Practitioners' corner: Computing robust standard errors for within-groups estimators. Oxford bulletin of Economics and Statistics 49(4), 431-434.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden (2007). Remedying education: Evidence from two randomized experiments in India. The Quarterly Journal of Economics 122(3), 1235-1264.

Banerjee, A. V. and E. Duflo (2011). Poor economics: A radical rethinking of the way to fight global poverty. PublicAffairs.

Benson, C. (2004). The importance of mother tongue-based schooling for educational quality. Commissioned study for EFA global monitoring report 2005.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-indifferences estimates? The Quarterly Journal of Economics 119(1), 249-275.

Blimpo, M. P., N. Lahire, and D. K. Evans (2011). School-based management and educational outcomes: Lessons from a randomized field experiment. Working paper.

Bokamba, E. G. (1984). French colonial language policy in Africa and its legacies. Studies in the Linguistic Sciences 14(2).

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics 90(3), 414-427.

Cameron, A. C. and P. K. Trivedi (2009). Microeconometrics using stata, Volume 5. Stata Press College Station, TX.

Chin, A., N. M. Daysal, and S. A. Imberman (2013). Impact of bilingual education programs on limited english proficient students and their peers: Regression discontinuity evidence from Texas. Journal of Public Economics 107, 63-78.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. Review of educational research 49(2), 222-251.

Duflo, E., R. Hanna, and S. P. Ryan (2012). Incentives work: Getting teachers to come to school. The American Economic Review, 1241-1278.

Dustmann, C., T. Frattini, and G. Lanzara (2012). Educational achievement of secondgeneration immigrants: An international comparison. Economic Policy 27(69), 143-185.

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

Eriksson, K. (2014). Does the language of instruction in primary school affect later labour market outcomes? Evidence from South Africa. Economic History of Developing Regions 29(2), 311-335.

Fabunmi, M. (2009). Historical analysis of educational policy formulation in Nigeria: Implications for educational planning and policy. International Journal of African \& AfricanAmerican Studies 4(2).

Glewwe, P., M. Kremer, and S. Moulin (2009). Many children left behind? Textbooks and test scores in Kenya. American Economic Journal: Applied Economics 1(1), 112-135.

Laitin, D. D. and R. Ramachandran (2015). Language policy and human development. American Political Science Review, Forthcoming.

Loury, L. D. (2006). All in the extended family: Effects of grandparents, aunts, and uncles on educational attainment. The American economic review, 275-278.

Mitchell, B. R. (2003). International historical statistics: Africa, Asia \& Oceania, 1750-2000, houndmills, basingstokes. Hampshire, New York: Palgrave Macmillan.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. The Review of Economics and Statistics, 334-338.

Moulton, J. (1994). How do teachers use textbooks and other print materials: A review of the literature. The Improving Educational Quality Project, South Africa.

Muralidharan, K. (2013). Priorities for primary education policy in India's 12th Five-year plan. In India Policy Forum, Number 9, pp. 1-46.

Ramachandran, R. (2015). Language use in education and human capital formation: Evidence from a natural experiment in Ethiopia. Mimeo.

Slavin, R. E., N. Madden, M. Calderón, A. Chamberlain, and M. Hennessy (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. Educational Evaluation and Policy Analysis 33(1), 47-58.

Taylor, S., M. Coetzee, et al. (2013). Estimating the impact of language of instruction in South African primary schools: A fixed effects approach.

Thomas, W. P. and V. P. Collier (2002). A national study of school effectiveness for language minority students' long-term academic achievement.

UNECOSOC (2011). Background paper - imperative for quality education for all in Africa: Ensuring equity and enhancing teaching quality. Technical report, United Nations Economic and Social Council.

Weber, E. (1976). Peasants into Frenchmen: The modernization of rural France, 1870-1914. Stanford University Press.

White, H. (2001). Asymptotic theory for econometricians. Academic press New York.

Whitehead, C. (2005). The historiography of British imperial education policy, Part II: Africa and the rest of the colonial empire. History of Education 34(4), 441-454.

Zhang, Y. (2006). Urban-rural literacy gaps in Sub-Saharan Africa: The roles of socioeconomic status and school quality. Comparative Education Review 50(4), 581-602.

Table A1: Summary Statistics on Outcomes by Treatment Status in Grades 1, 3 and 5

|  | Mean | SD | N | Min | Max |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Panel A: Treated Students - Grade 1 |  |  |  |  |  |
| Standardized values of overall score - Grade 1 | 0.73 | 0.87 | 323.00 | -1.35 | 2.56 |
| Standardized values of English score - Grade 1 | 0.34 | 1.03 | 325.00 | -1.71 | 2.19 |
| Standardized values of Math Score - Grade 1 | 0.57 | 1.05 | 325.00 | -0.86 | 3.55 |
| Raw Overall Score Grade 1 | 52.31 | 21.99 | 323.00 | 0.00 | 98.17 |
| Raw Score English Grade 1 | 52.52 | 26.25 | 325.00 | 0.00 | 100.00 |
| Raw Score Math Grade 1 | 44.80 | 32.58 | 325.00 | 0.00 | 100.00 |
| Panel B: Control Students - Grade 1 |  |  |  |  |  |
| Standardized values of overall score - Grade 1 | -0.71 | 0.46 | 335.00 | -1.35 | 1.56 |
| Standardized values of English score - Grade 1 | -0.33 | 0.86 | 336.00 | -1.71 | 2.19 |
| Standardized values of Math Score - Grade 1 | -0.55 | 0.54 | 336.00 | -0.86 | 1.59 |
| Raw Overall Score Grade 1 | 16.12 | 11.52 | 335.00 | 0.00 | 73.17 |
| Raw Score English Grade 1 | 35.38 | 21.91 | 336.00 | 0.00 | 100.00 |
| Raw Score Math Grade 1 | 9.84 | 16.98 | 336.00 | 0.00 | 76.92 |
| Panel C : Treated Students - Grade 3 |  |  |  |  |  |
| Standardized values of overall score - Grade 3 | 0.43 | 0.88 | 166.00 | -1.47 | 2.41 |
| Standardized values of English score - Grade 3 | 0.38 | 0.91 | 166.00 | -1.42 | 2.84 |
| Standardized values of Math score - Grade 3 | 0.42 | 0.90 | 166.00 | -1.28 | 2.28 |
| Raw Overall Score Grade 3 | 45.27 | 17.52 | 166.00 | 7.70 | 84.60 |
| Raw Score English Grade 3 | 41.76 | 17.18 | 166.00 | 8.00 | 88.00 |
| Raw Score Math Grade 3 (max. possible 100) | 51.55 | 23.60 | 166.00 | 7.10 | 100.00 |
| Panel D: Control Students - Grade 3 |  |  |  |  |  |
| Standardized values of overall score - Grade 3 | -0.68 | 0.77 | 100.00 | -1.85 | 1.90 |
| Standardized values of English score - Grade 3 | -0.61 | 0.81 | 100.00 | -1.84 | 2.20 |
| Standardized values of Math score - Grade 3 | -0.64 | 0.75 | 100.00 | -1.55 | 2.01 |
| Raw Overall Score Grade 3 | 23.36 | 15.20 | 100.00 | 0.00 | 74.40 |
| Raw Score English Grade 3 | 23.12 | 15.21 | 100.00 | 0.00 | 76.00 |
| Raw Score Math Grade 3 | 23.79 | 19.69 | 100.00 | 0.00 | 92.90 |
| Panel E: Treated Students - Grade 5 |  |  |  |  |  |
| Standardized values of overall score - Grade 5 | 0.18 | 0.96 | 85.00 | -2.00 | 2.72 |
| Standardized values of English score - Grade 5 | 0.20 | 0.98 | 85.00 | -1.98 | 2.71 |
| Standardized values of Math score - Grade 5 | 0.03 | 0.96 | 85.00 | -2.18 | 2.90 |
| Raw Overall Score Grade 5 | 27.78 | 7.81 | 85.00 | 10.00 | 48.57 |
| Raw Score English Grade 5 | 35.59 | 11.47 | 85.00 | 10.00 | 65.00 |
| Raw Score Math Grade 5 | 17.37 | 7.58 | 85.00 | 0.00 | 40.00 |
| Panel F: Control Students - Grade 5 |  |  |  |  |  |
| Standardized values of overall score - Grade 5 | -0.25 | 0.97 | 39.00 | -2.18 | 2.20 |
| Standardized values of English score - Grade 5 | -0.33 | 0.91 | 39.00 | -1.98 | 1.85 |
| Standardized values of Math score - Grade 5 | 0.06 | 1.06 | 39.00 | -2.18 | 1.63 |
| Raw Overall Score Grade 5 | 24.32 | 7.90 | 39.00 | 8.57 | 44.29 |
| Raw Score English Grade 5 | 29.36 | 10.68 | 39.00 | 10.00 | 55.00 |
| Raw Score Math Grade 5 | 17.61 | 8.34 | 39.00 | 0.00 | 30.00 |

The raw scores are expressed in percentages.


|  | Treated Students - Present Grade 1 |
| :---: | :---: |
| - - - - | Treated Students - Present Grade 1 and 3 |
|  | Treated Students - Present Grade 1, 3 and 5 |
| - - | Control Students - Present Grade 1 |
|  | Control Students - Present Grade 1 and 3 |
|  | Control Students - Present Grade 1, 3 and 5 |

kernel $=$ epanechnikov, bandwidth $=0.2985$

Figure A1: Determining direction of selection bias


[^0]:    *Department of Political Science, Stanford University, dlaitin@ stanford.edu.
    ${ }^{\dagger}$ Faculty of Economics and Business Administration, Goethe University, ramachandran@econ.uni-frankfurt.de.
    ${ }^{\ddagger}$ Graduate Institute of Applied Linguistics, steve_walter@ gial.edu.
    The authors have no conflict of interest to report in regard to this paper. The project was funded by SIL International and SIL Cameroon.

[^1]:    ${ }^{1}$ Southern and Eastern Africa Consortium for Monitoring Educational Quality. SACMEQ II Project 2000-2004 [dataset]. Version 4. Harare: SACMEQ [producer], 2004 and SACMEQ II Project 2005-2010 [dataset]. Version 1. Harare: SACMEQ [producer], 2012. Paris: International Institute for Educational Planning, UNESCO [distributor], 2010 and 2012.

[^2]:    ${ }^{5}$ Though Berber and Arabic both are from the Afro-Asiatic language family they share only one branch in common; this is like comparing Hindi to English, two languages which are very distinct.
    ${ }^{6}$ Official language is the language in which the primary affairs of the community - the government, the media, the courts and most important for the discussion at hand, the schools - are conducted.
    ${ }^{7}$ The differential levels of exposure faced by language minority students in developed countries, for instance Ghanaian migrants in the United States, compared to students learning in the colonial language in post-colonial states, for instance Ghanaian students learning in English in

[^3]:    ${ }^{8}$ Akin to value added models, primary school leaving tests scores can be considered to be the output arising from the combination of the entirety of provided schooling inputs, such as teacher quality and experience and schooling infrastructure.

[^4]:    ${ }^{9}$ Primary schooling in Cameroon since 2005 is for a duration of 6 years.
    ${ }^{10}$ In the first 3 years, language skill assessment in Kom was also carried out for students in experimental schools.

[^5]:    ${ }^{11}$ For details about the content of the assessments, refer to the document available on the following link:
    https://drive.google.com/file/d/0B079K2j40KG6TmFha1dGNUo0bkk/view?pref=2pli=1.
    ${ }^{12}$ The experiment was conducted by Stephen L. Walter, who designed the protocols for the SIL International.

[^6]:    ${ }^{13}$ Here exposure can be understood generically as the amount of usage of the medium of instruction in social interactions and also the teacher competence in the medium of instruction.

[^7]:    ${ }^{14}$ The theoretical distinction though is crucial to explain why studies on language minority students in developed countries cannot be used to make inferences about the role of language in settings such as Cameroon. The reason being that the level of exposure to the medium of instruction for language minority students, for instance Ghanaian migrants in the United States, would be much higher than for Ghanaian students learning in English in Ghana.
    ${ }^{15}$ The data on schools and students characteristics are collected post-intervention. However the information is mainly on categories that are determined prior to the intervention period and/or unlikely to be affected by the intervention.
    ${ }^{16}$ Note that we theoretically conceptualized individual, family and school level factors as being included in the $v_{i}$. The fact that available individual, family and school level factors (i.e. elements of $v_{i}$ ) are not correlated to the treatment suggest that it is indeed orthogonal to local language instruction.

[^8]:    ${ }^{18}$ The data are from UNESC0 Institute for Statistics. Retrieved at the url: http://data.uis.unesco.org/Index.aspx?queryid=1540
    ${ }^{19}$ For instance assuming a 10 percentage point lower survival and higher repetition rate in rural areas, as compared to the national average, would imply an attrition rate of 65 percent.

[^9]:    ${ }^{20}$ The estimated kernel density of the standardized overall tests score in Grade 1 for the various subgroups is shown in Figure A1 in the Appendix.

[^10]:    ${ }^{21}$ In Table VI only the 95 percent confidence intervals are shown, the 90 percent confidence interval is $0.007-0.84$.
    ${ }^{22}$ The primary school leaving exams scores are available only for 22 of the 24 schools in our sample.

[^11]:    ${ }^{23}$ The data on the years of education of father are not available for all the students present in Grade 5, as some of the students were not present in school on the day when the data was collected.

[^12]:    ${ }^{24}$ Surprisingly many randomized evaluations on the effect of school inputs on test scores tend to concentrate on normalized test scores, and often do not report raw scores (Banerjee et al., 2007; Duflo et al., 2012). In the absence of raw scores it is hard to evaluate absolute levels of learning and importance of the magnitude of treatment effects.

[^13]:    ${ }^{25}$ For instance in Namibia, where the language of instruction in all classrooms has been English for the past 20 years, a government conducted teacher language competency test showed that $98 \%$ of the southern African country's teachers are not proficient in the language. Worse, more than $70 \%$ of the teachers in senior secondary schools cannot properly read and write elementary English (Refer to The Guardian, $10^{\text {th }}$ January 2012, "Namibia's language policy is 'poisoning' its children"). Similarly the World Bank review of public expenditure (2009) notes that only $11 \%$ of the teachers in Tanzania have what education experts consider to be the minimum language skills in English required for teaching at the secondary school level. To be sure, a report in mid-19th century France revealed egregious gaps in knowledge of the state language by public school teachers in about $20 \%$ of its communes (Weber, 1976, 310-314). But, unlike Namibia, there was in France a majority that was fluent in the official language, French.

