Long-Lasting Effects of Exposure to Bible Translations: Evidence from Sub-Sahara Africa[1]

Vinicius Okada Da Silva
University of Illinois at Urbana-Champaign

M. Noelia Romero
University of Illinois at Urbana-Champaign

Abigail Stocker
University of Illinois at Urbana-Champaign

Rebecca Thornton
Baylor University

August 31, 2023

Abstract

This paper evaluates the impact of early life exposure to Bible translations on education. To estimate causal effects and avoid issues with selection into translation and mission locations, we compare educational outcomes across cohorts of individuals within language groups, with and without exposure to a Bible translation in their mother-tongue language, during their primary school years. We analyze data from a representative sample of ~77,000 adults in 13 sub-Saharan African countries using the Demographic and Health Surveys. Our difference-in-differences strategy accounts for the differential timing of Bible translations and the increase in educational outcomes over time within each language. Individuals born ten to fifteen years after the first Bible translation are 14 percentage points more likely to be literate later in life and attain 1.5 additional years of education than those born before the translation. Effects do not vary by proximity to missions (either Catholic or Protestant), distance to a printing press, urban area, or religious faith. We provide the first causal evidence of the impact of the Bible on education.

Keywords: literacy, Bible translations, languages, missions, ethnicity, cohort.

---

[1]*Acknowledgments*: We thank comments from Marieke Kleemans, Mary Arends-Kuenning, and the seminar participants at the Applied Micro Research Lunch Seminar, Graduate Seminar at the Department of Economics (University of Illinois at Urbana Champaign), the Inequality, Demography, Health, and Religion (IDHR) workshop at Baylor University, and the Collaborative for Econometrics and Integrated Development Studies (CEIDS) seminar. We also thank to attendees at sessions at the CIES 2023 Conference and seminar at the Department of Economic (Universidad de San Andres).

Contact information: University of Illinois at Urbana-Champaign. Department of Economics, 1407 W. Gregory Drive, David Kinley Hall Room 126, Urbana, Illinois 61801. Okada Da Silva: vo10@illinois.edu; Romero: marianr2@illinois.edu; Stocker: stocker3@illinois.edu; Thornton: Rebecca_Thornton@baylor.edu

Introduction

Sub-Saharan Africa experienced two significant changes in education and religious beliefs in the 19th century. First, between 1950 and 1990, the literacy rate for individuals with at least five years of schooling increased from 29.8 to 62.7 for women and 60.6 to 76.4 for men (Le Nestour et al., 2021). Second, more than half of the population identifies as Christian today, while only 5% of the population was Christian in 1900 (Jedwab et al., 2022; Johnson & Grim, 2020). Christian missions in Africa have an important role in Africa, providing the first formal education schools and institutions to evangelize (Bai & Kung, 2015; Cagé & Rueda, 2016). Additionally, Christianity stresses the importance of literacy, leading to investments in education and resulting in higher rates of literacy (Boppart et al., 2014; Calvi et al., 2022; D. Kim, 2020).[2] While many have claimed that the Bible itself led to increased literacy (Feld, 2022; Moilanen & Sommerseth, 2021; Mosher, 2016), despite various identification approaches to measure causal effects, previous papers have not been able to disentangle the effect of exposure to the Bible from the larger confounding set of missionary and Christian influences.[3]

We study the long-term impact of Bible translations on today's educational outcomes and Christianity in Africa. We construct a dataset of Bible translations over time in Sub-Saharan Africa, using information from the largest atlas of languages, the *Ethnologue* (Eberhard et al., 2023). To our knowledge, producing this data makes this paper the first to exploit the cohort variation within languages to measure the effect of a Bible translation on education and religious belief. We

[2] Measuring literacy as an outcome is important because it is relevant for many educational outcomes, such as memorization (Odén, 1975; Guttormsson, 1990), reading (Moilanen & Sommerseth, 2021), and overall cognitive skills (Boppart et al., 2014; Fernihough & Henderson, 2015)., suggesting a link between translation of texts and educational attainment. Additionally, Christian missions and pre-colonial contact from Europe are determinants of language survival (Buzasi, 2015).

[3] Out of the numerous papers studying the impact of religion or Christianity on educational outcomes, only two papers explicitly examine the impact of the number of Bibles or Bible schools and find no overall effect, in Africa and China (Bai & Kung, 2015; Nunn, 2010).

combine these Bible data with round 7 of the Demographic and Health Surveys (DHS), which includes reading assessments to measure adult literacy among a sample of men and women living across 13 Sub-Saharan African countries.

The reason why we cannot simply compare educational outcomes and Christianity in ethno-language groups with earlier and later Bible translations located in proximity to the same mission is because the choice of which languages to translate and when to translate them is not random. Historical accounts point to missionaries choosing to translate into the language of the largest ethno-language group first. For example, in Uganda, missionaries in the late 19th century decided to translate the Bible into Luganda, the language of the largest ethnic group, the Baganda people (Mutibwa, 2016; Tuma & Mutibwa, 1978). One of the first Protestant missionaries in Ethiopia in the 1840s, Johann Ludwig Krapf, translated the Gospels into the language of the Oromo people, who comprise the largest ethnic group in Ethiopia. He did this because he believed that converting the Oromo to Protestantism would be key to the conversion of the rest of northeast Africa (Pawlikova-Vilhanova, 2006). The same missionary started a Bible translation into Swahili, which was completed in 1891 would give another large ethnic group in East Africa access to the Scripture in their mother tongue (Pawlikova-Vilhanova, 2006).

We document evidence of positive selection into Bible translations. Using the Ethnoatlas dataset from George Murdock (Bahrami-Rad et al., 2021; Kirby et al., 2016), we show that ethno-language groups with a Bible translation, or an earlier Bible translation, are the groups that rely more on agriculture with extensive cultivation and rely less on gathering, hunting, and fishing. Given the positive selection of language translation, we turn to a novel identification strategy of measuring the impact of Bible translation exposure in primary school. Using an event-study difference in difference estimation strategy that accounts for differential timing of

Bible translation, we compare individuals within a language group, across cohorts. Assuming that the educational outcomes and Christianity is mainly during early elementary years between the ages of 5 and 15, we compare the adult outcomes of cohorts with and without exposure to a Bible translation in their mother-tongue language using the relative timing of the translation. Our specification relies on the variation across languages, years of Bible translation, and individuals' year of birth. Our estimates account for the differential timing of Bible translations to each language and the increasing literacy trend over time within each language (Callaway & Sant'Anna, 2020; Le Nestour et al., 2021).

For individuals who are approximately born ten years after the first Bible translation in their mother tongue, we find a positive effect of 14 percentage points on literacy and attain 1.5 additional years of education relative to those born before translation. The effects of exposure to a Bible translation are consistent across gender, different sample restrictions, and estimation method (TWFE instead of Callaway and Santa'Anna estimates). We discuss the use of geographic variation and the issues in the previous literature when using the mission location. The finding of the impact of Bible translation on literacy is robust and positive regardless of the metropolitan area location, proximity to missions (either Catholic or Protestant), and distance to printing press. We also find that exposure to Bible translations increases the likelihood of being a Christian by 11 percentage points.

We contribute to the literature on economic history regarding missions' long-term effects on education and religious beliefs. Missionary activities are associated with a package of inputs such as new ideas/beliefs, schools, hospitals, Bible translations, and printing presses. We build on previous findings regarding the role of missionaries in human capital accumulation (Acemoglu et al. 2014; Nunn 2014; Cagé and Rueda 2016; Baten and Cappelli 2017; Alesina et

al. 2019; Huillery 2009; Becker et al. 2021; Wantchekon et al. 2015; Okoye and Pongou 2014; Doyle et al. 2019; Cagé and Rueda 2019, Calvi and Mantovanelli 2018). Exposure to missions also increases Christian beliefs and attitudes in Africa (Nunn 2010; Jedwab et al. 2018; Becker et al. 2021; Woodberry 2012) as well as leading to inter-generational effects on trust in institutions (Wantchekon et al. 2015 in Benin, Alesina et al. 2019, Meier zu Selhausen et al. 2018 in Uganda, Ricart-Huguet 2019). Our paper adds to the evidence of the impact of access to a text in one's mother tongue on literacy as a first step in human capital accumulation.

We contribute to the literature studying the effect of Bible translation on Christian belief by documenting that the earlier translations of the Bible are positively correlated with ethnic groups having more dependency on hunting and agriculture and less gathering and nomadic behavior (Bai & Kung, 2015; Moilanen & Sommerseth, 2021; Nunn, 2010). Secondly, our unique empirical strategy exploits the Bible translation timing across languages and the cohort variation. Finally, we contribute with a novel approach that disentangles the mission from the Bible effects by using the theory of literacy during the schooling age.

The remainder of the paper is organized as follows. In Section 2, we describe our contemporaneous and historical data, summarizes the analytical sample, and provide the evidence of selection of languages for the first translation of the Bible. The empirical approach in Section 3 start with the treatment definition of exposure of languages to a Bible translation. Then, we present our difference-in-difference specification, discuss assumptions and threats to identification. In Section 4, we show our main finding of the impact of Bible translation on educational and Christianity outcomes, with a subsection that summarize the robustness checks. Finally, we discuss the implications and final remarks in Section 5.

2. Data

This section introduces our contemporary and historical data sources and then sets the stage with the background and context of African Bible translations. In our modern data, we measure our primary and secondary outcomes of interest: education and religious affiliation. The historical data sources cover information on the locations and characteristics of missions, precolonial characteristics of various ethnic groups, and years and languages of Bible translations. We finish this section with a summary the linked datasets and analytical samples.

2.1. Contemporaneous Data: Demographic and Health Survey (DHS)

To study the effect of exposure to a Bible translation in one's mother tongue on adult educational outcomes, we use the Demographic and Health Surveys (DHS) round 7. We use data from 13 countries in Sub-Saharan Africa for individuals born between 1948 and 2005. These data contain a representative sample of married women between 15 and 49 years old in each country. Additionally, we use the DHS men's survey for data on the male members of the households sampled in the women's survey. The surveydesign means that the sample of men is not nationally representative of men between 15 and 59 years old.[4]

The Demographic and Health Surveys include information on the geographic locations of households by sampled cluster. The DHS has two types of GPS locations: urban (2 kilometer buffer) and rural (5 kilometer radius). Each cluster contains, on average, 36 observations, ranging from a minimum of one to a maximum of 123 (see Figure A.1. in the Appendix).

Our primary educational outcome of interest is literacy. In the survey, interviewers assessed reading ability using a card with various language sentences. The respondents choose which language they want to take the literacy test in, and they are asked to read the sentence in that

[4] Finally, for clarification, the join individual data of women and men's survey is not a longitudinal dataset.

language. The literacy test in round 7 measures whether an individual can read a whole sentence, part of a sentence, or nothing.[5] We use two definitions of literacy: i) fully literate, in which an individual can read the entire sentence, and ii) partially literate, in which an individual can read only part of a sentence. We present the latter definition for robustness as a less conservative measure of literacy in our additional estimations.

The data contains literacy, ethnicity, religion, education, and age of the respondent. We use the age of the respondent and ethnicity to determine whether individuals had access to a Bible translation. This variation is explained in further detail in Section 2.3. Literacy, religion, and other measures of education are used as outcome variables, and we estimate the impact of access to a Bible translation on each of these outcomes. Respondents classify themselves into one of 138 ethnic groups. We later assign individuals to languages based on these ethnic groups.

In all, our sample includes 306,591 individuals in 13 countries where we have non-missing literacy, GPS data, and ethnicity.[6] We restrict our analysis to the 13 countries that contain data on ethnicity: Benin, Chad, Ethiopia, Gambia, Ghana, Guinea, Kenya, Malawi, Mali, Nigeria, Senegal, Sierra Leone, and Uganda. Figure 1, Panel A, shows the geographic distribution of the available DHS clusters in these 13 countries. To have a context of the level of Christianity in the countries in our DHS sample, Panel B shows the ratio of Muslim to Christian from Pew Forum (Pew Forum, 2010). From comparing both panels, we notice that most countries like Benin, Ethiopia, Ghana, Guinea, and Kenya have a lot of variation of this Muslim-to-Christian ratio within the country.

[5] Individuals choose in which language to take the literacy test, though their chosen language was not included in the data set.

[6] There are originally 23 Sub-Saharan African countries with literacy and GPS information, but 10 countries do not have information on ethnicity (Angola, Burundi, Cameroon, Lesotho, Liberia, Rwanda, South Africa, Tanzania, Zambia, and Zimbabwe).

2.2. Historical Data

2.2.1. Missions

We obtain data on mission locations from two seminal papers with public data: Nunn (2010) and Cagé and Rueda (2020). The information on the location of the mission in these two papers comes from three historical maps. Nunn (2010) digitalized the map from William R. M. Roome (1924)[7], which also includes information on Bible translations. Cage and Rueda (2020) uses mission location on Protestant mission from *Geography and Atlas of Christian Missions* [Beach (1903)]. The Catholic mission information comes from official Vatican sources with a map by Streit (1929) in the 1929 *Atlas Hierarchicus*.

The geographic variation in locations of Catholic and Protestant mission stations is illustrated in Figure 2. This figure shows the location of 665 missions from both sources. The data from Nunn (2010) provides the sites of 394 missions, and the data from Cage & Rueda (2020) provides the locations of an additional 271 missions. We have a total of 224 Catholic missions and 441 Protestant missions. The data from Cage & Rueda (2020) also includes information on whether or not each mission has a printing press on site. However, this data is very limited and is only included for Protestant missions in our dataset.

2.2.2. Precolonial Characteristics

We use a database of precolonial characteristics of ethnic groups provided by (Kirby et al., 2016). This dataset combines cultural data from the *Ethnographic Atlas*, the Binford Hunter-Gatherer Dataset, the Standard Cross-Cultural Sample, and the Western North American Indians datasets (Lowes, 2020). The data is available for 1291 societies across the globe, including

[7] Roome, William R. M. 1924. "Ethnographic Survey of Africa: Showing the Tribes and Languages; also the Stations of Missionary Societies [map]."

societies located in our countries of interest in Africa. This data has many different cultural traits for each society, but the ones that we are most interested in for this analysis include reliance on hunting, gathering, fishing, and agriculture; marital structure; agricultural intensity; major crop types; and settlement patterns.

2.2.3. Bible Translations

To obtain data on Bible translations, we use information from the 16th edition of the *Ethnologue: Languages of the World* by Eberhard et al. (2023). This atlas has information on the year of the first translation for each specific language, the portion of the Bible translated, whether the language is written, and the status of the language. Figure 3 shows the timing of translation into different languages in Sub-Saharan Africa. After the 1980s, we see a significant increase in the number of African languages translated over time. By the end of the 1990s, 100 languages in Sub-Saharan Africa had at least some portion of the Bible translated. Our analysis uses this variation in timing of Bible translations across languages.

2.3. Linking Datasets and Analytical Sample Summary

To examine the impact of exposure to a mother-tongue Bible translation, we link our contemporary data on educational outcomes and ethnicity data with mission locations, Bible translation year by language, and ethnic group characteristics before the arrival of colonial powers. We do this in a two-step process.

In step one, we assign a language to each individual in the DHS based on that individual's ethnicity, since the DHS does not directly report language. We use the package "Linking Ethnic Data from Africa" (LEDA) by Müller-Crepon et al. (2021) to match each ethnic group in the DHS to a language. The package offers several language matches for each ethnic group, prioritized based on how commonly the language is spoken. To obtain the one-to-one match

between ethnic groups and languages required for our analysis, we choose the most commonly spoken language, L1, as the matched language for each ethnic group. This results in a one-to-one match between ethnic groups in the DHS and languages in LEDA without losing observations from any ethnic group. For simplicity and as a result of this match, we use ethnic and language group synonymously in the rest of the paper.

In step two, we match each language to data on the year of Bible translation from the "*Ethnologue: Languages of the World*" (Eberhard et al., 2021), which allows us to combine the data on Bible translations with the DHS survey data.[8] Next, we match mission locations to the DHS data using the GPS data for both the mission locations and the clusters to assign the closest mission to each cluster.[9] Finally, we match each ethnic group in the DHS to the ethnic groups in the Ethnoatlas in order to obtain information on the characteristics of each group. These two steps result in 138 ethno-language groups with matched Bible translation years, mission locations, and ethnic group characteristics.

Table 1 summarizes our final analytical sample after linking the individual DHS dataset with the other data sources. All of our 13 Africa countries, are located in Central Africa except for Malawi. Column (1) shows the number of individuals surveyed by country, while Column (2) shows the percentage of those individuals who are literate. Column (3) shows the number of languages per country. Notably, Uganda has the most language variation, with representation across 52 out of the 214 languages in our sample. Finally, Column (4) displays the number of missions per country, ranging from only six missions in Chad to 189 missions in Nigeria.

[8] Figure A.2. shows the percentage of individuals in our DHS sample that speak those languages and have a Bible available in their mother tongue over time.
[9] It is possible that there are missions missing from our dataset, which could result in bias if the locations of the missing missions are correlated with translation timing.

2.4. Selection of Languages with Early Translations

To show the selection of earlier translations, we divide our language sample into languages translated before1970, languages translated after 1970, and languages that were never translated.[10] In our sample of 138 languages, 45 were translated before 1970 (early Bible translations), 86 were translated after 1970, and seven have never been translated. Table 2 provides information on the precolonial characteristics of the language groups in our sample by whether they received an early translation, received a later translation, or never received a translation. For each group of languages, we show the percentage of ethno-language groups in each category by activity panel. For example, of the groups that received a translation prior to 1970, 77.8 percent of the groups were in the lowest category for gathering dependence. The table shows that the ethno-language groups with earlier translations had lower dependence on gathering, hunting, and fishing. Additionally, agricultural dependence for ethno-language groups with earlier Bible translations is stronger, at around 46-55%, with a predominance of extensive cultivation and cereal crops. Compared to languages with later or no translations, the languages with the first Bible translations were more likely to have polygamous marital composition and village or town settlements.

Table 2 suggests that there are systematic differences between languages that were translated at different points in time (or never translated). These results provide evidence of the selection of languages into translation, meaning that a comparison between languages with early translations and those without early translations is likely to be biased. This bias implies that a simple comparison between adult literacy rates with translated languages and those without translated languages would overstate any causal effect of Bible translation. For this selection of the early

[10] We use the year 1970 because is almost the middle point of the range of year of birth in our DHS individual data. Additionally, as shown in Figure 3, after 1980s there is a clear change in the trend of translations.

Bible translation, we exploit the ethnic group cohort-level treatment to estimate the impact of exposure to a Bible during schooling on adult literacy.[11]

3. Empirical Approach: Difference-in-Differences strategy

The empirical approach estimates the impact of exposure to a Bible translation, exploiting two types of variation for the identification strategy. First, we use cross-sectional variation at the language group level, comparing across languages that receive a Bible translation in different years. Second, we use time variation across birth cohorts within each language group. In this section, we explain how we utilize both the differential timing of Bible translation and cohort level treatment. Then, we present the specification, assumptions, and threat to identification.

3.1. Treatment Definition: Exposure to Bible Translation

We use two time variables to define the treatment of exposure of an individual to a Bible. First, the language and year of birth of the respondents from the DHS. Second, the year of the Bible translation to the respondent's language. We combine these timing in the exposure to a Bible translation as follows:

$$Exposure_{ilt} = \mathrm{I}(BirthYear_{ilt} - BibleYear_{l}) \qquad (1)$$

In this equation, the indices indicate individual i, language l, and cohort t. $\mathrm{I}(.)$ is an indicator for the gap between the year of birth of individual i language l cohort t, $BirthYear_{icl}$, and the year of translation of language l, $BibleYear_{l}$. The treatment variable, $Exposure_{ilt}$, is a set of indicators for each cohort's access to a Bible translation. Cohorts are defined as 5-year intervals based on birth year relative to the year of the translation. For example, there is an indicator for

[11] In Appendix C, we present a simple exercise of exploiting only the language level and mission geographic variation to estimate the effect of the earlier Bible translations on literacy.

the individuals born 15 years before the Bible translation, an indicator for individuals born 10 years before the Bible translation, continuing every 5 years. The youngest cohort in our balanced sample is individuals born 15 years after the Bible translation. In an event study framework, these cohort indicators represent different times relative to the event time.

In order to construct the bins, we round both the year of Bible translation and the year of the individual's birth to the nearest 5, then take the difference between them. Thus, individuals are assigned into groups around the year of the Bible translation. In other words, $Exposure_{il} =$ I$(BirthYear_{il} - B\text{ible}Year_{l}) = \alpha$ for each 5-year bin ($\alpha = -15$ to $\alpha = 15$ for the balanced sample). Essentially, this variable indicates for how long an individual's ethno-language group was exposed to a Bible translation before that individual was born. To illustrate, if a language had a translation in 1956 and a respondent was born in 1959, we would round the year of Bible translation to 1955 and the year of birth to 1960. The difference between the two would be 5, indicating that the individual's ethno-language groups had access to a Bible translation for approximately 5 years before the individual was born. Thus, this individual would belong in the bin for $\alpha = 5$, indicating an average of 5 years of exposure to a Bible translation. In the regression in subsection 3.2, we include a balanced panel, which includes only ethnic groups that have individuals surveyed for all cohorts born 15 years before to 15 years after the Bible translation.

3.2. Specification

We evaluate a treatment of exposure to Bible translation within language across cohort in a reduced form specification. Providing an event study framework, the two-way fixed effects (TWFE) equation is as follows:

$$y_{ilt} = \sum_{t \neq -15} \beta_t \, Exposure_{ilt} + \delta X_i + \alpha_t + \lambda_l + \epsilon_{ilt} \qquad (2)$$

Where the indices indicate individual i, language l, and cohort t. Our outcome of interest y_{ilt} is educational measures (such as literacy discussed in previous section) and Christian beliefs. In this specification, we do include a fixed effect λ_l for the language group and a fixed effect α_t for each cohort. $X_i$includes individual-level characteristics, such as male and urban dummies. Notice that we do not include mission or country fixed effects or Ethnoatlas controls.

In our sample, we use the -15 bin as the reference group for comparison. This bin includes individuals whose ethno-language group received access to a Bible translation when the individual was 15 years old on average. We select this bin as the reference group because these individuals should already finished primary and elementary school around the time that the Bible translation is completed in their mother tongue. Thus, this group likely does not have access to a Bible translation when learning to read, since dispersion of the translation was likely not immediate.[12]

For this approach, we are interested in the coefficients β_a, which capture the effect of exposure to Bible translation, relative to the reference group. Intuitively, a way to rationalize how to estimate the effect of Bible translations on education is to think on the ideal field experiment. The ideal experiment of the effect of Bible translation during primary school is giving the Bible text to some languages randomly across schools. Then, one compares the literacy or any educational outcome of children with and without the text for each language across different schools. Since we do not have information on schools using Bible translation for

[12] Note that if these individuals did have access to a translation while learning to read, this would tend to attenuate our estimates.

reading, but we do have information on when the Bible is translated, then one can use the cohort exposed and not exposed. In other words, we present a reduce form that estimate in a difference-in-difference research design that estimates the effect of exposure of an individual language to a Bible translation depending on their year of birth.

Our identification comes from variation across cohorts and languages with different years of Bible translation. The individual's language, their year of birth, and the year of Bible translation into that language jointly determine their exposure to a Bible translation. In summary, all these comparison groups aim to account for the differential timing of each language treatment with a translation.

3.3. Assumptions and Threats to Identification

This approach uses an event-study estimation, and it is important to understand the assumptions that we are making in estimating the equation (2). Since the event study is a generalized difference-in-differences design, the main assumption for identification is the parallel trends assumption. We assume that literacy trends over time between the treated and control language are parallel. Intuitively, if there are compositional changes across cohorts, these should appear in both the treatment and the control groups.

The model in equation (2) is a naïve specification because it does not account for differential treatment timing, which is present in our model since the Bible was translated into different languages in different years. This differential treatment timing is has an additional underlying assumption to the parallel trend across cohort groups (Cunningham, 2021). We assume a constant treatment effects for our language groups over time. The overall problem with the TWFE in a context of differential treatment timing is the heterogeneity in time bias (Goodman-Bacon, 2021). However, given the recent literature on TWFE models with differential treatment timing,

our main results use the semiparametric method by the Callaway and Sant'Anna (2021) estimates. The Callaway and Sant-Anna (CS) estimates compare the treated and never-treated groups with semiparametric weights. These estimates correct for the differential timing of Bible translations across the treated ethno-language groups, ensuring that we are not improperly comparing treated and untreated groups at different points in time.[13] We show the TWFE results of the model in equation (2) in the Appendix tables.

4. Results

This section presents the results for the impact of exposure to Bible translations on education and Christian beliefs. Then, we present a series of robustness checks for the literacy outcomes and how positive effects of access to a mother-tounge Bible do not change significantly when we exploit the geographic variation and mission data.

4.1. Impact of Bible Translation on Education

We present the main results of the impact of exposure to a Bible translation on literacy in Figure 4. This figure shows the Callaway and Sant'Anna (2020) estimates for the within-ethnic-group effects of receiving a Bible translation on full literacy, which takes 1 when the respondent can read a full sentence. The estimates shows no evidence of violation of the parallel trends assumption in the periods before the treatment. We find an increasing effect of Bible translation on full literacy, which flattens out at around a 15 percentage-point increase. The results for partial literacy in Figure A.3 are similar (see Appendix A), with flat pre-trends and a treatment effect that increases to around 18 percentage points. These figures show that access to a Bible translation in one's mother tongue improves the probability of being literate by 15 to 18

[13] However, these results are not very different from the naïve TWFE estimations, which are shown in Appendix B, indicating that differential treatment timing is not causing a large amount of bias.

percentage points for cohorts whose ethno-language group has been exposed to the translation for 10-15 years at the time of the individual's birth. Table 3 shows the corresponding point estimates for the CS approach by cohort.[14]

By comparing across cohorts, these estimates account for the increase of literacy across all languages in Africa (Le Nestour et al., 2021). This means that our results are not driven by overall increases in literacy, and we separate out the effect of access to a Bible translation from the underlying trend. It is worth noting that the effect of a translation does not happen immediately but seems to be delayed until cohorts born around 10 years after the translation. One possible reason for this delay in the impact of exposure to a Bible translation is connected to the printing press and availability and distribution of the text, as indicated by (Cagé & Rueda, 2016a). While the missionaries translate the Bible in the year indicated, it takes time before they can reproduce those translations at a low cost for use in schools. Thus, it is reasonable that it takes several years from the completion of the translation to the Bible to a widespread increase in literacy.

We present the results of the impact of exposure to a Bible translation on other educational outcomes in Figure 5.[15] The two educational outcomes are the maximum year of educations (upper panel) and the completion of primary schools (lower panel). The first outcome is a continue variable expressed in year units, while the second outcome is a dummy that takes 1 when the DHS's respondent finished primary school. These panels display a similar pattern to the literacy outcome. For cohorts whose ethno-language group has been exposed to the translation for 10 years at the time of the individual's birth, the Bible translation increases

[14] Table B.1. shows the naïve TWFE estimates of the effect of the Biblle on full and partial literacy.
[15] For space purpose and to simplify the Figure, Table, and Appendix sections, we only include the figure of these and the rest of robustness estimates. The tables with the point estimates coefficient are available on request.

education by almost 1.5 years and the probability of finishing primary school by 15 percentage points. The positive effect of acces to a Bible are also similar when we use dummy variables for any education level and secondary or higher level completion (see Figure A.4).[16]

4.2. Effects of Bible Translation on Christian Beliefs

The missions in Africa have the overall goal to evangelize the people which has a direct effect on the Christian belief outcome with the translation of Bibles. In Figure 6, we present the effect of exposure to a Bible translation on Christian affiliation. The estimates showsome evidence of violation of the parallel trends assumption for respondents born 10 years before the Bible is translated. We find an increasing effect of Bible translation on Christian identity by 10 percentage-point increase for cohorts whose ethno-language group has been exposed to the translation for 15 years at the time of the individual's birth.

This result contrasts with the findings of no effect of Bible translation on Christianity by Nunn (2010). Our approach exploits the cohort variation within ethnic groups, while his approach exploits the geographic variation of ethnic groups close to a mission location. Since those missionary located strategically and the mission maps have biased location (Jedwab et al., 2022), the approach exploiting the cohort variation within language group shows the long lasting effect of access to a Bible on the Christian belief today,

4.3. Robustness Check

In this subsection, to provide evidence of the robustness of the findings and the heterogeneity across different characteristics, we summarize a series of estimations with different samples.

[16] The educational outcome of secondary education or Higher has two issues for the reader to be cautious. First, a younger sample is less likely to have finished secondary, so the coefficient might be downward biased. Second, the opportunity cost of continuing at school is higher for poor respondent than rich individuals. Thus, it is more likely they drop high school to start working affecting the measure of the completion outcome.

Then, we focus on the long-lasting effect of Bibles on literacy exploiting the geographic variation of the DHS observations and its proximity to the mission location.

4.3.1. Heterogeneity

While our main results are shown with the binned and balanced ethnic groups, we also show the un-binned results and unbalanced sample of all ethnic groups in Figure 7. The general trends remain the same. In the un-binned results in the upper panel, the estimates reach effect sizes of around 20 percentage points for full literacy for individuals born 10 years after the translation, which is similar to the main estimates. For the unbalanced sample in the lower panel, the point estimates are smaller and noisier, especially for later cohorts due to the smaller number of observations. Notice that this panel shows the estimates when we do not restrict the sample to languages with observations in each 5-year bin and for a longer event-study window. However, the general pattern remains the same, showing an improvement in literacy that increases for later cohorts.

We show the effect of access to Bible by gender in Figure 8. The upper and lower panel shows results on literacy for the women and men sample, respectively. There is no suggestive evidence of parallel trend assumption violation by gender group, and the positive effects are largely explained by the women sample. This findings was expected because the DHS survey is design to a representative sample of women in each country and men are survey if they are present in the household when women are selected for interviews. Additionally, the results by gender are consistent with the large positive effect of Protestant missions on female and small effect on male documented by Nunn (2014).

Due to concerns about Ethiopia and Mali being different from the rest of the sample due to having earlier access to written languages and a long history of Christianity, we also estimate the

effects without the observations for these two countries in Figure 9. This figure shows that the effects are nearly identical to our main results, meaning that observations from Ethiopia and Mali are not driving our results.

4.3.2. Exploiting the Geographic Variation

In this subsection, we present the findings when exploitiong the geographic varition in two ways. First, when we use the geographic information just from the DHS datat. Second, when we exploit the distance between the DHS cluster and the mission locations.

We compare the effects of access to the Bible translation on literacy by metropolitan area in Figure 10. The metropolitan area is information included in the DHS dataset on individuals who live in urban and rural areas. Both panel show that the effects are similar and follow similar trends and pre-trends, with slightly larger point estimates for urban areas compared to rural areas.

While the missions are the one that translate the Bible to the different ethnic group in Sub-Saharan Africa, we are caveat to interpret the results using the location of mission due to bias from mission maps (Acemoglu et al., 2014; Fourie & Swanepoel, 2015; Jedwab et al., 2022). Missions did not locate randomly in Africa and settle close to rivers and slavery route, where the ethnic groups had less diseases (like Malaria) and less hostile to the colonies (Jedwab et al., 2022). Yet, there is an informative value to present the effect of exposure to Bible translations on adult literacy for individuals that are in proximity to historical mission locations.

The previous subsection shows that our findings regarding the impacts of access to a mother tongue Bible translation on literacy are robust to many different specifications. It is important to consider what is driving these impacts on literacy. First, it is possible that missionaries' arrival and work around the time of the translation (such as providing increased access to schooling, literacy programs, etc.) may be driving the increase in literacy. Second, it is possible that

individuals who receive access to a written text in their mother tongue are more likely to go to school or learn better at school, meaning that the effects come through the channel of increased schooling. Finally, it is possible that individuals learn to read at home using the mother-tongue text of the Bible, independent of a corresponding increase in schooling. This section examines each of these potential channels.

First, is the effect that we are capturing truly an effect of Bible translations or an effect of missionaries' presence? If the presence of missionaries was driving the results, we would expect to see effects upon the arrival of these missionaries. It is important to note that we only see effects for individuals who are born at the time that the Bible is translated or afterwards. This is important because Bible translation takes a long time, from 7 years for a New Testament to 16 years for a full translation, on average. Thus, if the effects were driven by the arrival of the missionaries, we would expect to see effects on earlier cohorts. Since this is not the case, it is unlikely that this channel is the primary driver of our results.

However, it is possible that programs launched contemporaneously with the release of the translation may be driving our results. When considering this, we must keep in mind that treatment is defined at the ethno-language group level, regardless of geography. This means that, for a very spread-out group, some of these people will not even be close to any missions and will still be considered treated. Related to this, we can estimate the effects on literacy for individuals who did not live near to a mission location. Figure 11 shows the impacts for individuals who live near (within 10 km) of a mission and individuals who live farther (more than 10 km away) from a mission. The upper and lower panel shows results on literacy for the less and more than 10 km away from a mission, respectively.

We see that both groups experience similar increases in literacy, though the effects are slightly larger for individuals who live near a mission. This indicates that living in close proximity to a mission is not the main driver of literacy, since individuals who live farther away also experiences similar increases in literacy. This suggests that access to the services and programs that the mission has to offer, such as schooling and literacy programs, is not the main driver of the increase in literacy.

One might also believe that another possible channel for the impact of Bible translation availability is the proximity to a printing press to produce and distribute those translations. While there is some anecdotal evidence of cooperation between Catholic and Protestant missions to translate the Bible together in Africa, previous papers suggest a religious competition between the denominations (Gallego & Woodberry 2010; Larreguy & Schmidt-Padilla 2017). We present our estimates of the impact of Bible translation on literacy by proximity to the type of mission in Figure 12. Our estimates are probably explained by both type of missions, where the individual close to historical Catholic mission present larger coefficient than those close to the Protestan stations. This results contrast with other studies that suggest the positive and larger effect of Protestantism on development comes from knowledge diffusion through printing presses, schools, and hospitals (Bai & Kung, 2015; Cagé & Rueda, 2020; Dittmar, 2011). In Figure A.5., we show that the sample of DHS observation close to missions without printing press also present a positive effect on literacy for individuals born after 10 and 15 years of the Bible translation.

5. Conclusion

Using contemporany data on education in Africa and historical datat on Bible translation into vernacular languages and mission locations, we explore the long-term effects of access to a Bible

translation in one's mother tongue on education and Christian beliefs. First, we document the positive selection of missionary decision on which languages to translate first. We find that the early Bible translation (pre-1970) are languages of ethnicities with less dependance on gathering and nomadic behavior, and more dependance on hunting and agricultural activities. Second, we exploit the cross-language variation with and without Bible, the timing of the Bible translation, and the year of birth from our individual data. We show the positive impact of having access to a Bible in one's mother tongue for individuals born 10-15 years after the first translation increases the literacy rate in almost 14 percentage points and 1.5 years of additional education. Within-language-group estimates suggest positive effects of translation on education and Christian beliefs for individuals who have access to a translation while they are of school age.

Amidst the challenges of discord, social division, and the search for common ground, this study finds that Bible translations have played a pivotal role in improving literacy and education in sub-Saharan Africa. This finding highlights that efforts to translate religious texts into mother-tongue languages resulted in long-lasting positive educational outcomes among Christians and non-Christians alike, transcending religious boundaries. The results also find that making the Bible accessible in local languages resulted in an increased affiliation with the Christian faith. Our results point to the role of missionary and religious institutions' involvement in development activities and lasting effects in improving education for all.

Our findings align with evidence from other fields, such as history, religion, and education. First, human capital accumulation and higher educational achievement for individuals from different ethnic groups around the missions depend on the original language structures. For instance, Chen (2013), Galor et al. (2021), and Kim et al. (2017) find that languages with periphrastic future tense are associated with higher educational attainment, more saving, more

wealth, better health outcomes, and higher prevalence of earnings management. Unfortunately, we do not have information in our dataset collection on the structure of the languages that would allow us to predict the probability that a language is translated in the first place, let alone predicting whether Bible translation increases the likelihood of being literate. However, our analysis considers the precolonial characteristics because the ethnic group features are determinants of missionary contact and later survival of languages (Buzasi, 2015; Jedwab et al., 2022). Future research approaches can help understand which ethnic groups are more predisposed to learn to read from the Bible depending on their language structure and proximity between languages, as shown by Ginsburgh & Weber (2020).

Finally, ethnic groups may have incentives to adopt either the dominant language from the European colonial power or other languages to increase economic interactions with the developed world. The labor and migration literature shows that language distance, which quantifies the dissimilarity between two languages, can increase communication costs, lower workplace productivity, and decrease trade between the two societies (Dale-Olsen & Finseraas, 2020; Heller, 2014; Isphording & Otten, 2013). Thus, it is probable that African ethno-language groups had economic incentives to interact with the missionaries and learn from those Bible translations to improve their relationship with Europeans.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). When should you adjust standard errors for clustering? In *NBER Working Papers* (Vol. 24003). http://arxiv.org/abs/1710.02926

Acemoglu, D., Gallego, F. A., & Robinson, J. A. (2014). Institutions, human capital, and development. *Annual Review of Economics*, *6*, 875–912. https://doi.org/10.1146/annurev-economics-080213-041119

Bai, Y., & Kung, J. K. S. (2015). Diffusing knowledge while spreading god's message: Protestantism and economic prosperity in China, 1840-1920. *Journal of the European Economic Association*, *13*(4), 669–698. https://doi.org/10.1111/jeea.12113

Cagé, J., & Rueda, V. (2016). The Long-Term Effects of the Printing Press in Sub-Saharan Africa. *American Economic Journal: Applied Economics*, *8*(3), 1–31.

Cagé, J., & Rueda, V. (2020). Sex and the mission: The conflicting effects of early Christian missions on HIV in sub-Saharan Africa. *Journal of Demographic Economics*, *86*(3), 213–257. https://doi.org/10.1017/dem.2019.16

Callaway, B., & Sant'Anna, P. H. C. (2020). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *xxxx*, 1–31. https://doi.org/10.1016/j.jeconom.2020.12.001

Cunningham, S. (2021). *Causal inference: The mixtape.* Yale University Press. https://mixtape.scunning.com/potential-outcomes.html?q=random#methodology-of-fishers-sharp-null

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2023). *Ethnologue. Languages of the World. Twenty-sixth edition.* https://www.ethnologue.com

Fourie, J., & Swanepoel, C. (2015). When selection trumps persistence: The lasting effect ofmissionary education in South Africa. *Tijdschrift Voor Sociale En Economische Geschiedenis*, *12*(1), 1–30.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Jedwab, R., Meier zu Selhausen, F., & Moradi, A. (2022). The economics of missionary expansion: evidence from Africa and implications for development. In *Journal of Economic Growth* (Vol. 27, Issue 2). Springer US. https://doi.org/10.1007/s10887-022-09202-8

Johnson, W. R., & Grim, B. J. (2020). *World Religion Database*. Leiden/Boston: Brill.

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H. J., Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R., Leehr, D., Low, B. S., McCarter, J., Divale, W., & Gavin, M. C. (2016). D-PLACE: A global database of cultural, linguistic and environmental diversity. *PLoS ONE*, *11*(7), 1–14. https://doi.org/10.1371/journal.pone.0158391

Le Nestour, A., Moscoviz, L., & Sandefur, J. (2021). The long-term decline of school quality in the developing World. *Article Submitted to a Journal*, 1–60. https://www.cgdev.org/sites/default/files/consultation-draft-le-nestour-et-al-school-quality-decline.pdf

Lowes, S. (2020). *Ethnographic and Field Data in Historical Economics* (Issue Working Paper No. 27918).

Müller-Crepon, C., Pengl, Y., & Bormann, N. C. (2021). Linking Ethnic Data from Africa (LEDA). *Journal of Peace Research*, 1–11. https://doi.org/10.1177/00223433211016528

Nunn, N. (2010). Religious conversion in Colonial Africa. *American Economic Review*, *100*(2),

147–152. https://doi.org/10.1257/aer.100.2.147

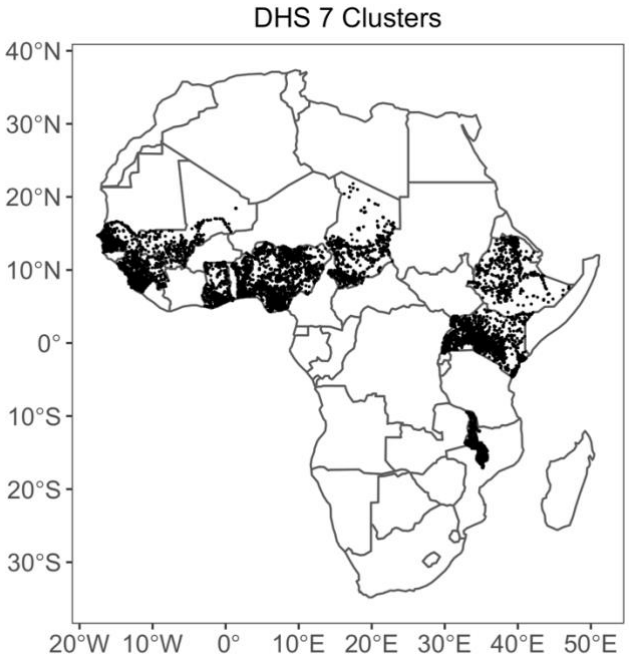
Nunn, N. (2014). Gender and missionary influence in Colonial Africa. *Africa's Development in Historical Perspective*, *February*, 489–512. https://doi.org/10.1017/CBO9781139644594.021

Pew Forum. (2010). *Tolerance & Tension: Islam and Christianity in Sub-Saharan Africa*. tp://features.pewforum.org/africa/%5Cnwww.pewforum.org

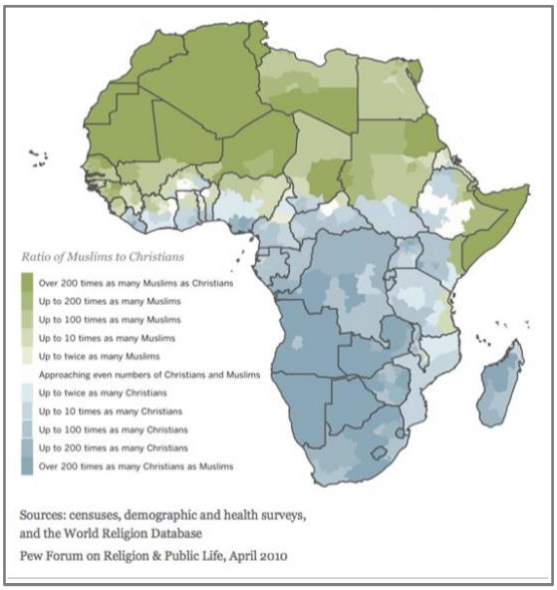
Figures

Figure 1. DHS sample and Christianity in Africa

Panel A. DHS 7 Clusters location

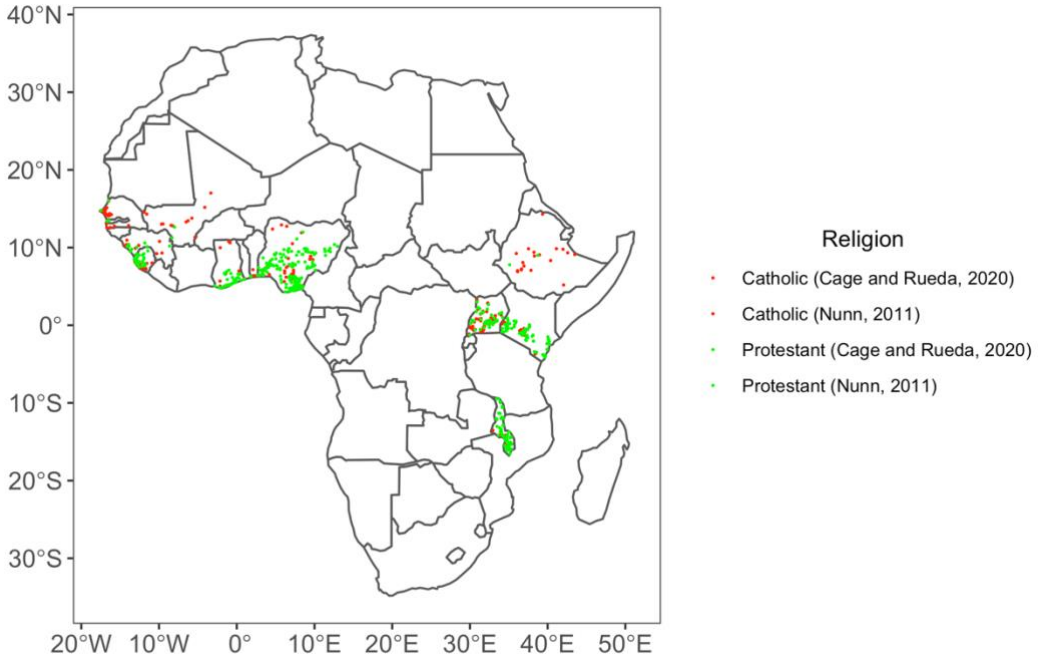


Panel B. Ratio of Muslims to Christian



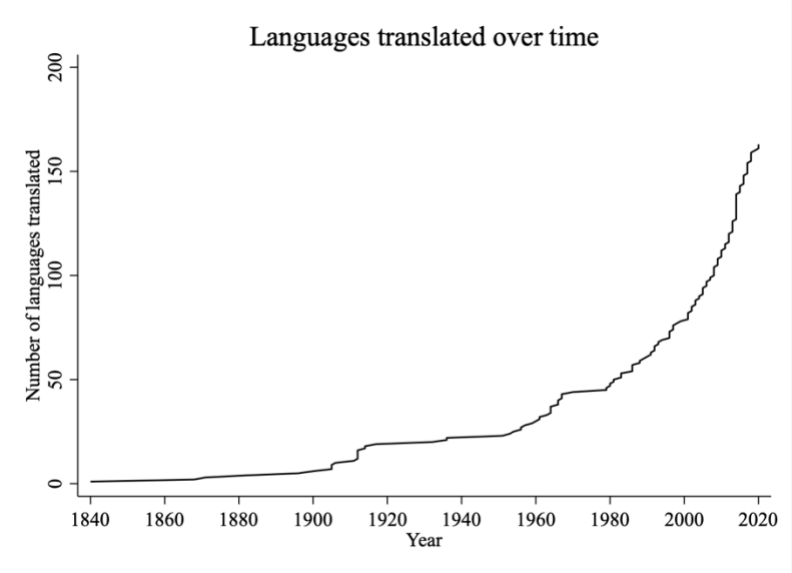
Note: In these figures, we present the geographic location of our individual data from the Demographic and Health Survey (DHS) round 7, Muslim to Christian distribution in Africa, and the DHS sample with a Bible translation . In Panel A, the individual data is grouped by cluster (see Subsection 2.A). Panel B shows the ratio of Muslims to Christians by Pew Forum on Religion & Public Life (2010).

Figure 2. Christian missions locations in Africa by type and data source



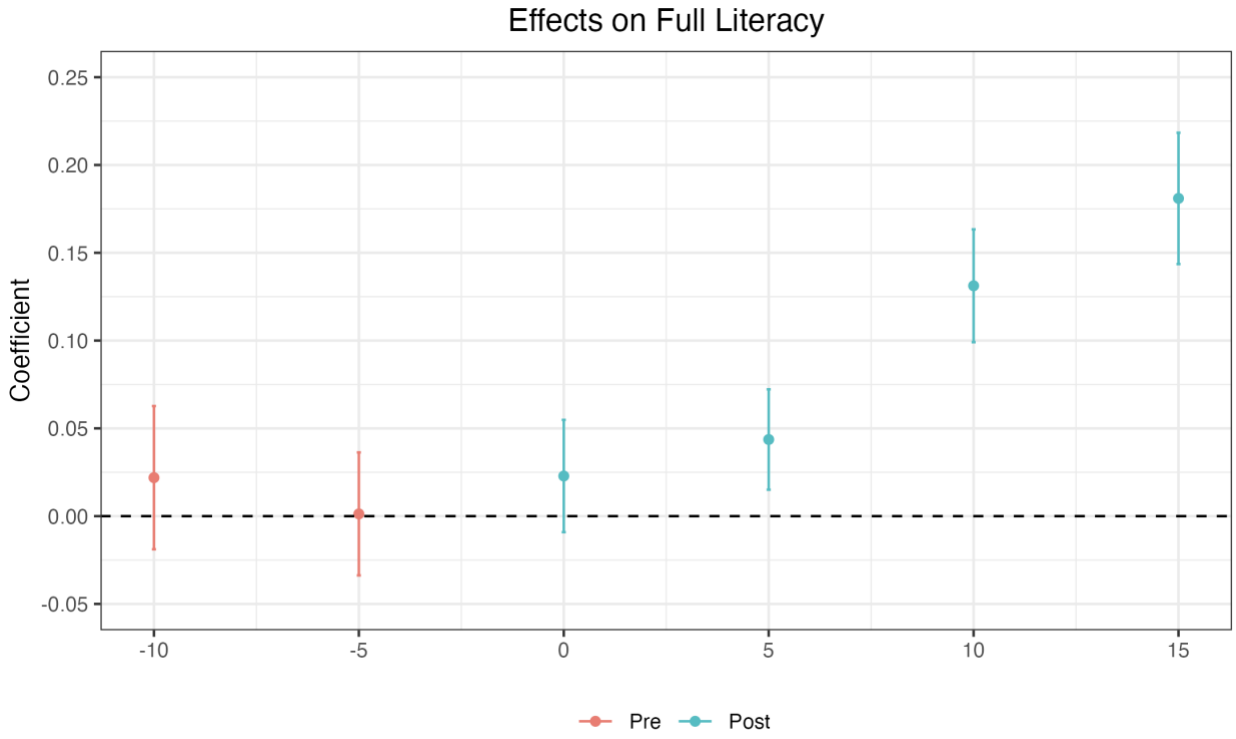
Note: This figure shows the location of missions by religion and source of information. There are 224 total Catholic missions, from those 126 missions are from (Cagé & Rueda, 2020) and 78 missions from (Nunn, 2010). Similarly, there are 441 Protestant missions in total, where 146 missions are from (Cagé & Rueda, 2020) and 316 missions from (Nunn, 2010).

Figure 3. Bible translations over time



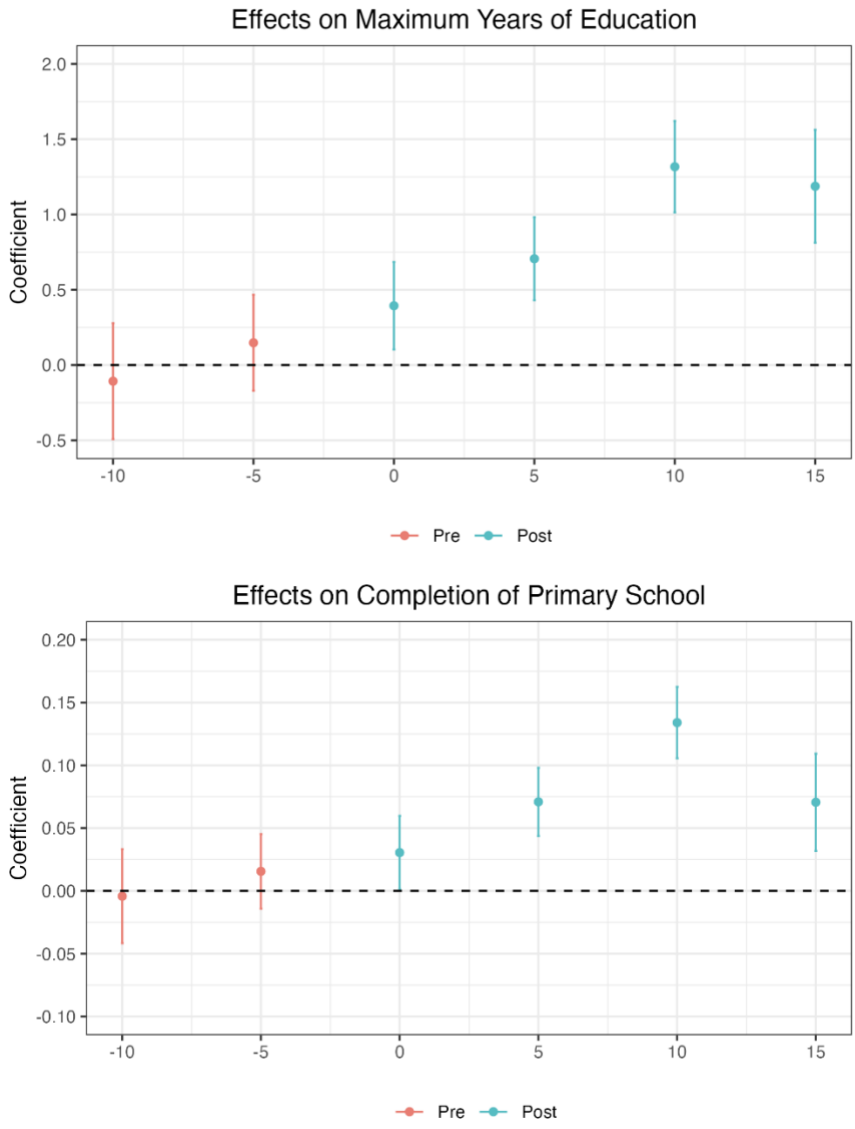
Note: This figure shows the timing and number of different African languages with a Bible translation from Eberhard et al. (2021).

Figure 4. Impact of Bible Translation on Full literacy



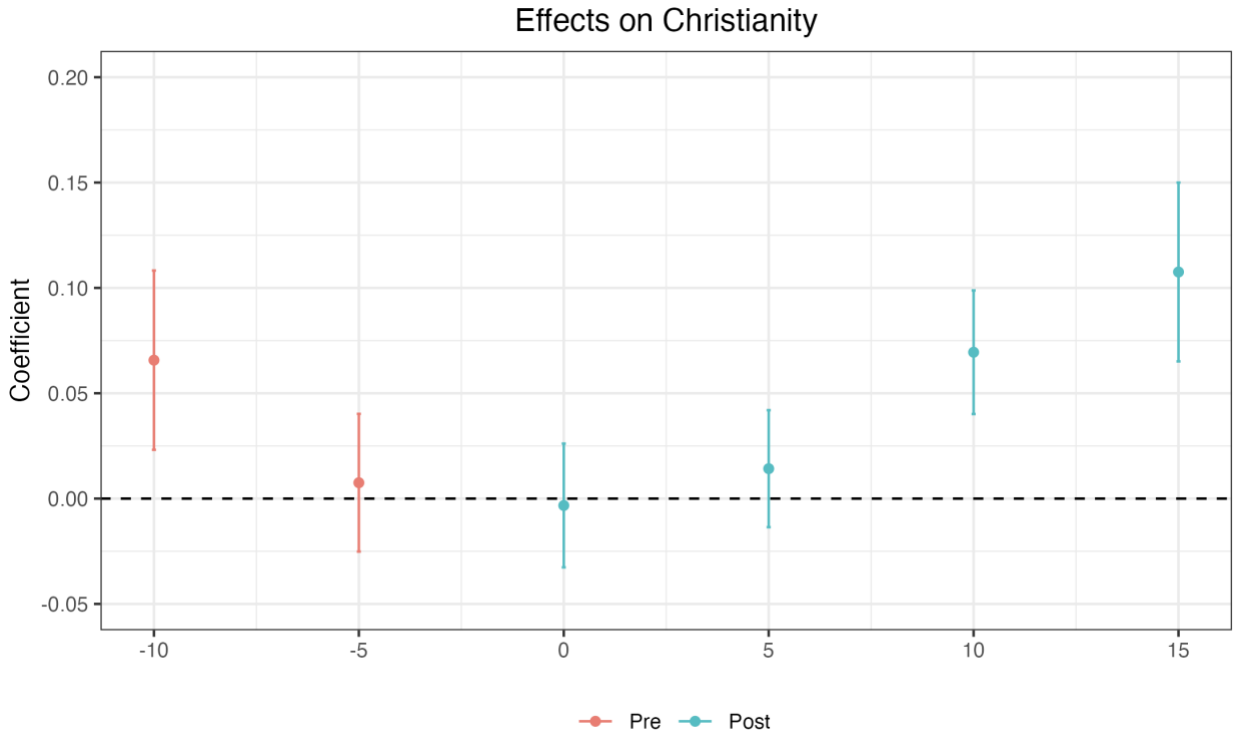
Note: This figure shows the effect of exposure to Bible translation on literacy when using the estimation method by Callaway & Sant'Anna (2020)**.** For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

Figure 5. Impact of Bible Translation on other Educational Outcomes



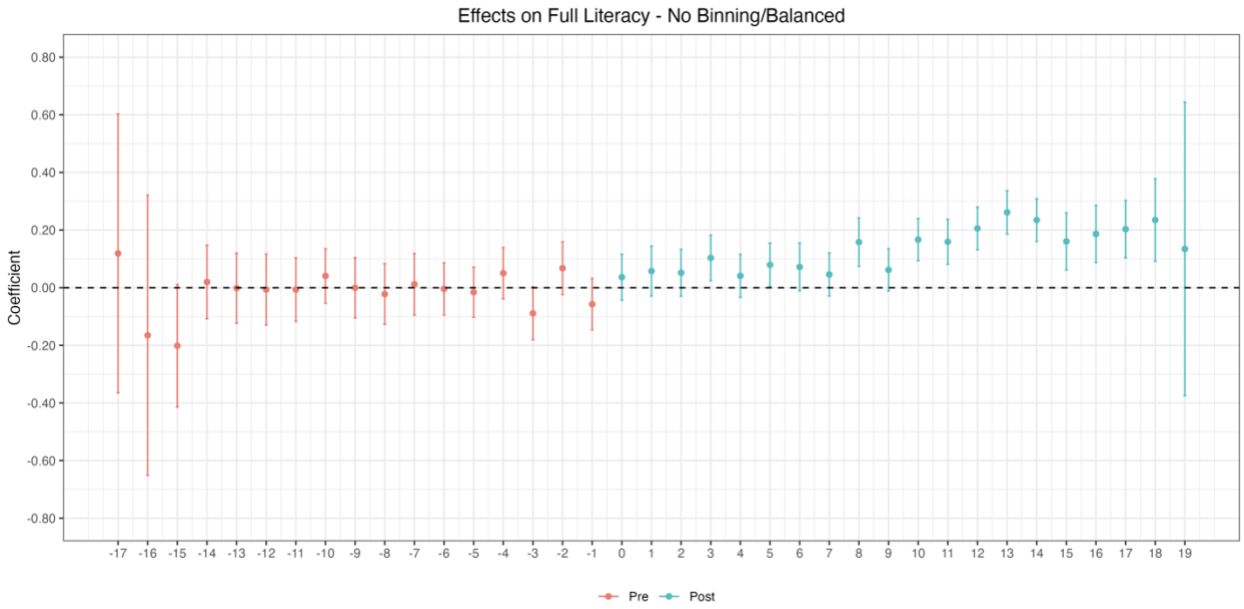
Note: These figures show the effect of exposure to Bible translation on two educational outcomes when using the estimation method by Callaway & Sant'Anna (2020)**.** The upper panel shows the maximum years of educations (continoues variable) and the completion of primary school (dummy variable that takes 1 when the individual finished primary school)**.** For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

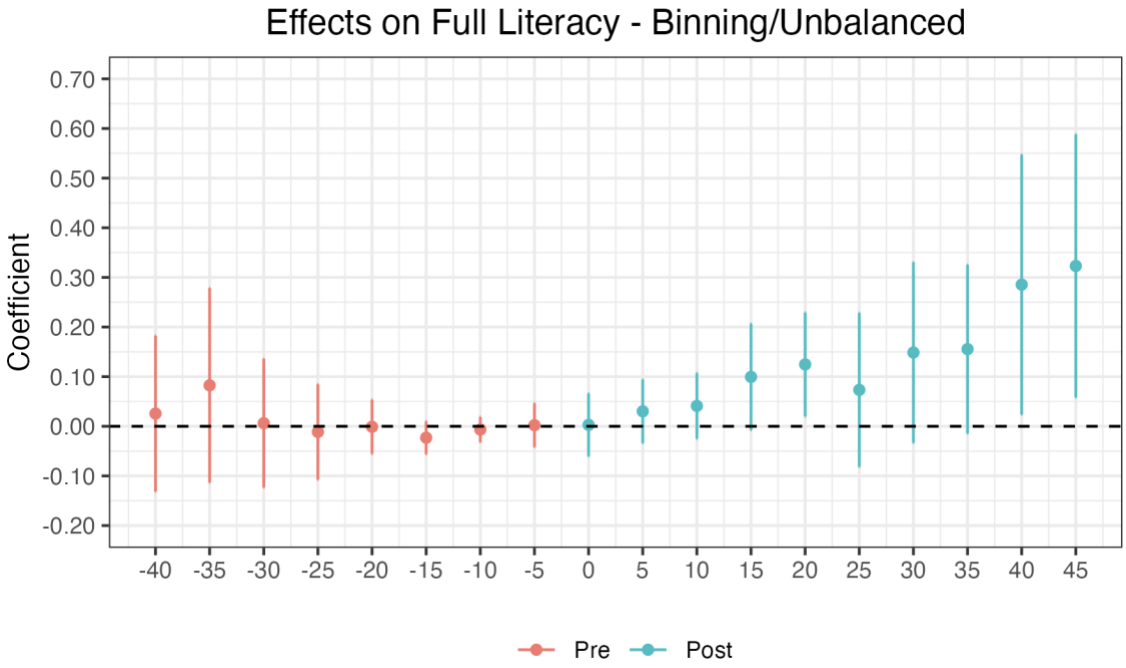
Figure 6. Impact of Bible Translation on Christian Belief



Note: This figure shows the effect of exposure to Bible translation on Christian belief when using the estimation method by Callaway & Sant'Anna (2020). For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

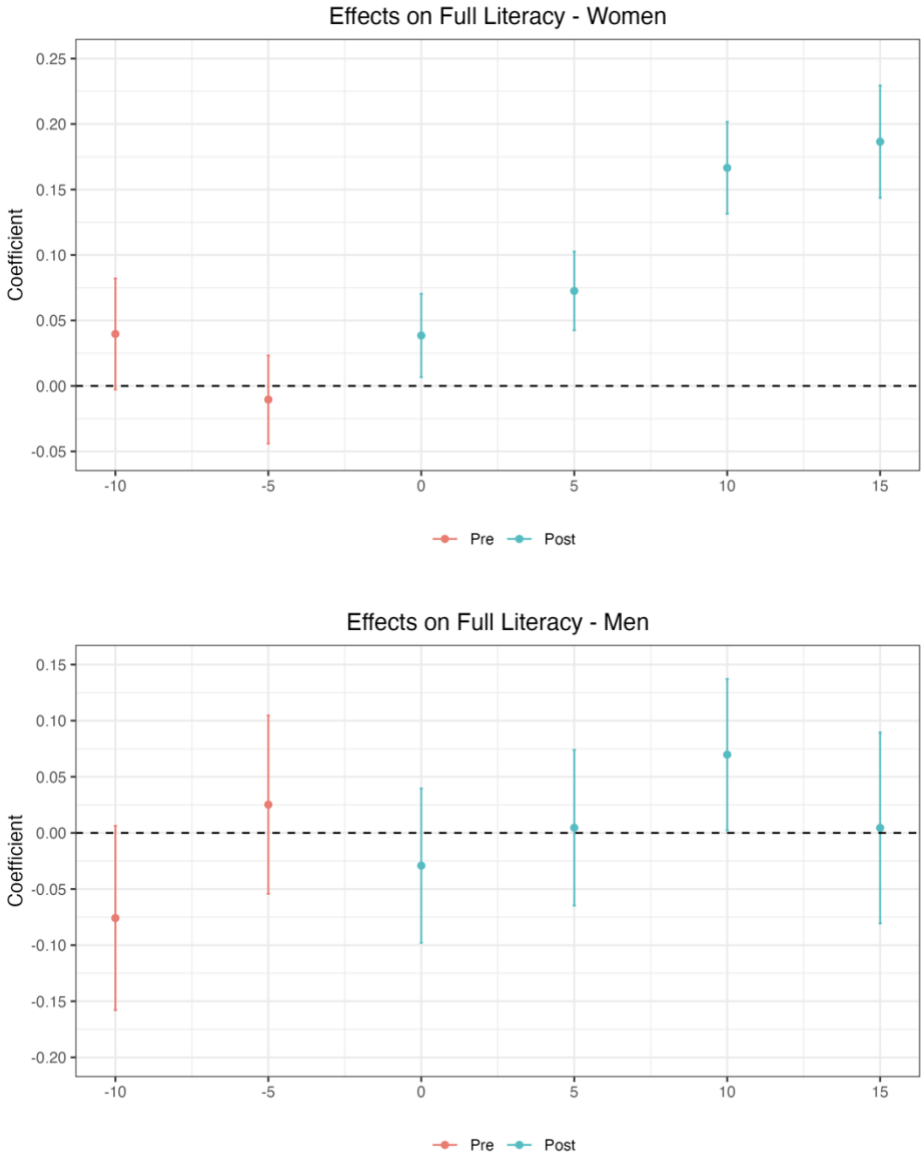
Figure 7. Effect of Bible Translation on Literacy Unbinned and Unbalanced sample





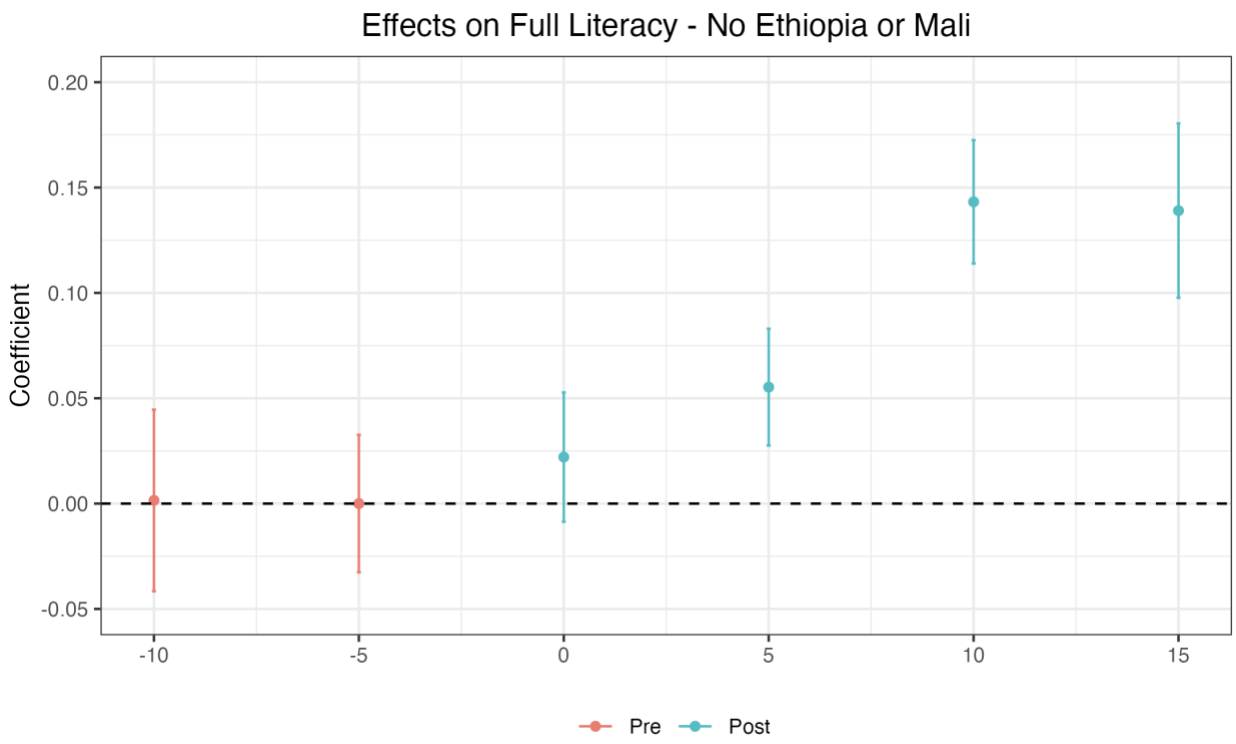
Note: These figures show the effect of exposure to Bible translation on literacy when using the estimation method by Callaway & Sant'Anna (2020)**.** The upper panel shows results on literacy when we do not use 5-year bin. The lower panel shows the estimates when we do not restrict the sample to languages with observations in each 5-year bin and for a longer event-study window For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

Figure 8. Effect of Bible Translation on Literacy by Gender



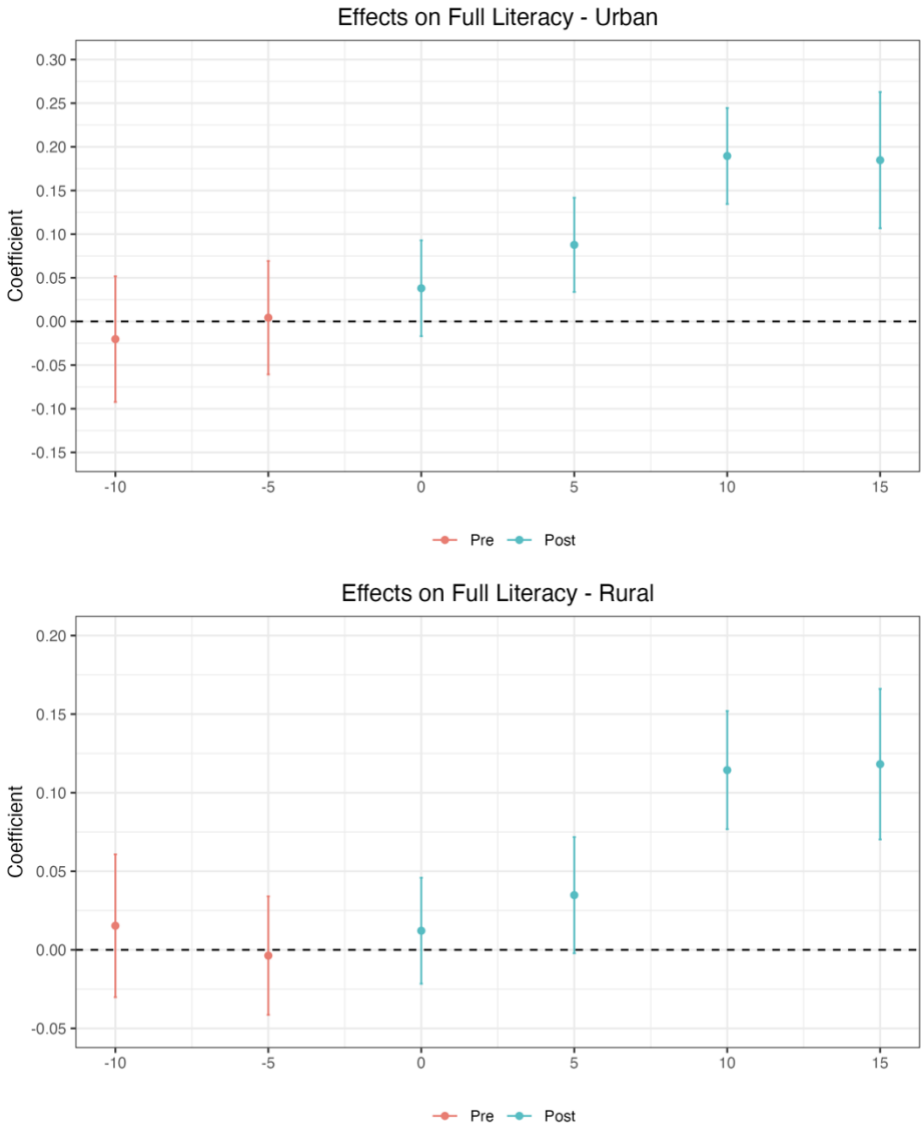
Note: These figures show the effect of exposure to Bible translation on literacy by gender using the estimation method by Callaway & Sant'Anna (2020)**.** The upper and lower panel shows results on literacy for the women and men sample, respectively. For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

Figure 9. Impact of Bible Translation on Full literacy without Ethiopia or Mali



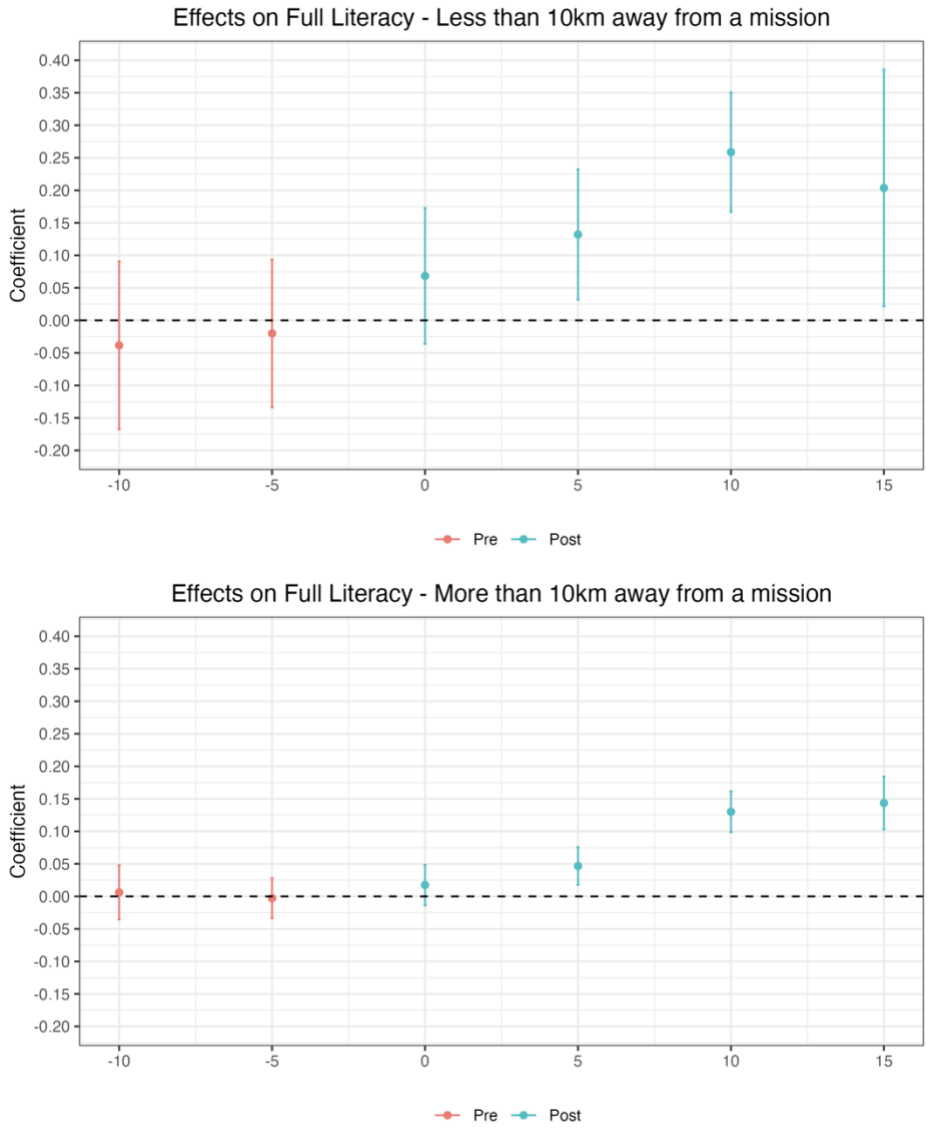
Note: This figure shows the effect of exposure to Bible translation on literacy when using the estimation method by Callaway & Sant'Anna (2020)**.** For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

Figure 10. Effect of Bible Translation on Literacy by Metropolitan Area



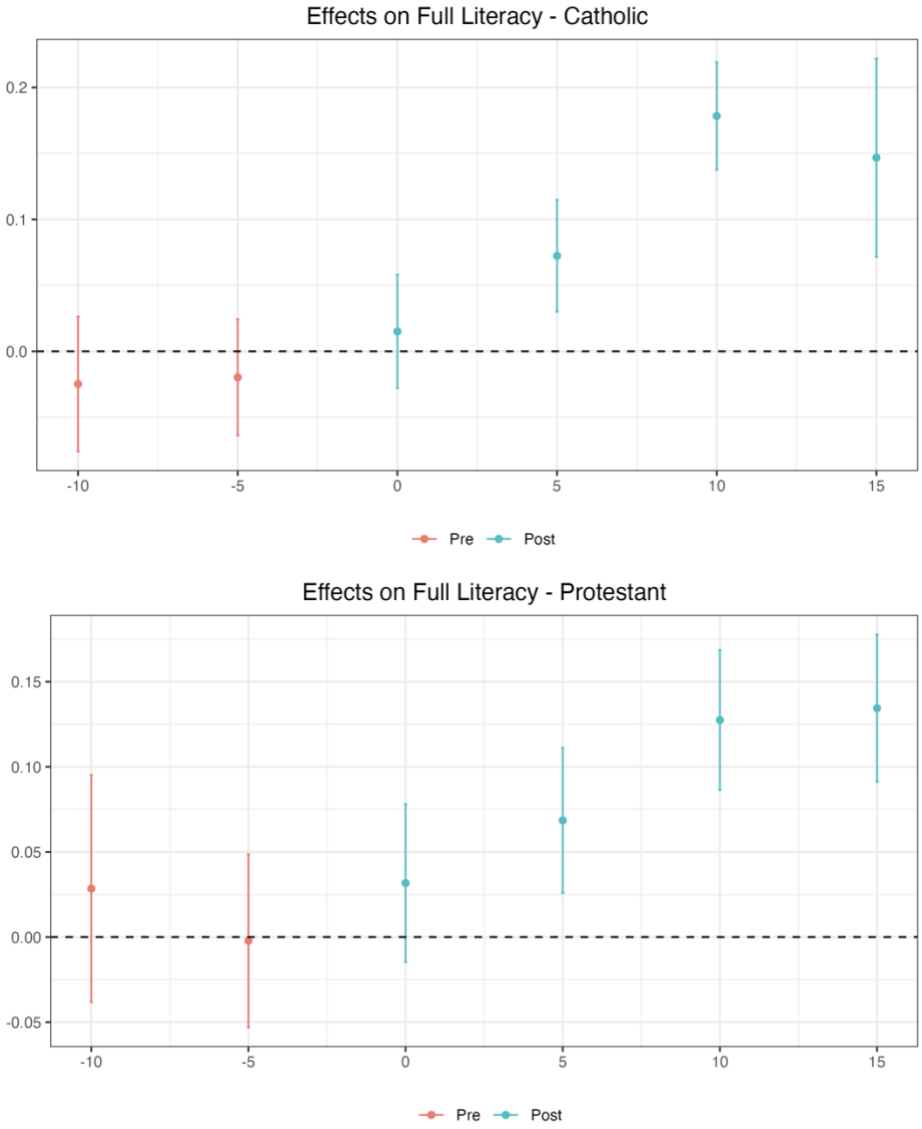
Note: These figures show the effect of exposure to Bible translation on literacy by metropolitan area using the estimation method by Callaway & Sant'Anna (2020)**.** The upper and lower panel shows results on literacy for the urban and rural sample, respectively. The metropolitan area is information included in the DHS dataset. For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

Figure 11. Effect of Bible Translation on Literacy by Proximity to Any Mission



Note: These figures show the effect of exposure to Bible translation on literacy by proximity to any mission location using the estimation method by Callaway & Sant'Anna (2020)**.** The upper and lower panel shows results on literacy for the less and more than 10km away from a mission, respectively. For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

Figure 12. Effect of Bible Translation on Literacy by Protestant and Catholic Mission proximity



Note: These figures show the effect of exposure to Bible translation on literacy by proximity to the type of mission using the estimation method by Callaway & Sant'Anna (2020)**.** The upper and lower panel shows results on literacy for the sample close to historical Catholic and Protestants missions, respectively. For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

Tables

Table 1. Descriptive data overview of countries, sample size, literacy, languages, and missions

DHS country	(1) DHS Sample size (N by Country)	(2) DHS countries' Fully literate rate (in %)	(3) Total Languages per country	(4) Total mission per country
Benin	21781	32%	8	22
Chad	22446	20%	23	6
Ethiopia	27471	42%	42	28
Gambia	13186	38%	8	8
Ghana	13362	48%	8	59
Guinea	14933	20%	6	28
Kenya	42839	76%	22	109
Malawi	31027	67%	10	70
Mali	14171	23%	9	17
Nigeria	39891	36%	10	189
Senegal	21785	39%	6	34
Sierra Leone	20723	27%	10	53
Uganda	22976	56%	52	102
Total	**306591**	**40%**	**214**	**725**

Note: In this table, we provide an overview of countries in the DHS, each country sample size, and characteristics of literacy rates, languages, and mission. This table uses the analytical sample described in subsection Linking Darasets and Analytical Sample.

Table 2. Pre-colonial characteristics across ethnic groups by time of Bible translation

	Translated Before 1970 (N=45)		Translated After 1970 (N=86)		Never Translated (N=7)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Gathering Dependence:						
Gathering 0-5%	77.8	42.0	72.1	45.1	57.1	53.5
Gathering 6-15%	20.0	40.5	25.6	43.9	42.9	53.5
Gathering 16-25%	2.2	14.9	2.3	15.2	0.0	0.0
Hunting Dependence:						
Hunting 0-5%	33.3	47.7	43.0	49.8	85.7	37.8
Hunting 6-15%	53.3	50.5	52.3	50.2	14.3	37.8
Hunting 16-25%	13.3	34.4	4.7	21.2	0.0	0.0
Fishing Dependence:						
Fishing 0-5%	35.6	48.4	48.8	50.3	14.3	37.8
Fishing 6-15%	35.6	48.4	36.0	48.3	57.1	53.5
Fishing 16-25%	22.2	42.0	11.6	32.2	28.6	48.8
Fishing 26-35%	4.4	20.8	3.5	18.5	0.0	0.0
Agricultural Dependence:						
Agriculture 26-35%	0.0	0.0	0.0	0.0	14.3	37.8
Agriculture 36-45%	6.7	25.2	7.0	25.6	0.0	0.0
Agriculture 46-55%	24.4	43.5	16.3	37.1	28.6	48.8
Agriculture 56-65%	42.2	49.9	44.2	50.0	57.1	53.5
Agriculture 66-75%	15.6	36.7	15.1	36.0	0.0	0.0
Agriculture 76-85%	11.1	31.8	9.3	29.2	0.0	0.0
Intensity of Cultivation:						
Extensive/shifting	66.7	47.7	59.3	49.4	28.6	48.8
Horticulture	4.4	20.8	3.5	18.5	14.3	37.8
Intensive	28.9	45.8	31.4	46.7	57.1	53.5
Principal type of crop cultivated:						
Tree-fruit	4.4	20.8	9.3	29.2	14.3	37.8
Roots/tubers	22.2	42.0	16.3	37.1	14.3	37.8
Cereals	73.3	44.7	72.1	45.1	71.4	48.8
Marital composition of family:						
Limited polygny	0.0	0.0	10.5	30.8	14.3	37.8
Polygyny, sororal separate	4.4	20.8	3.5	18.5	0.0	0.0
Polygyny, non-sororal cohabit	73.3	44.7	65.1	47.9	14.3	37.8
Polygyny, non-sororal separate	17.8	38.7	17.4	38.2	42.9	53.5
Prevailing type of settlement pattern:						
Seminomadic	0.0	0.0	2.3	15.2	14.3	37.8
Semisedentary	4.4	20.8	4.7	21.2	0.0	0.0
Dispersed homesteads	35.6	48.4	31.4	46.7	28.6	48.8
Hamlets	20.0	40.5	10.5	30.8	28.6	48.8
Villages/towns	31.1	46.8	41.9	49.6	28.6	48.8

Note: This table summarizes the precolonial reliance of each language group on different activities. We define earlier Bible translation as languages with a Bible translation before 1970, later translation as those ethno-languages with Bible translation after 1970, and languages without any Bible translations. We obtained information on precolony characteristics from Kirby et al. (2016).

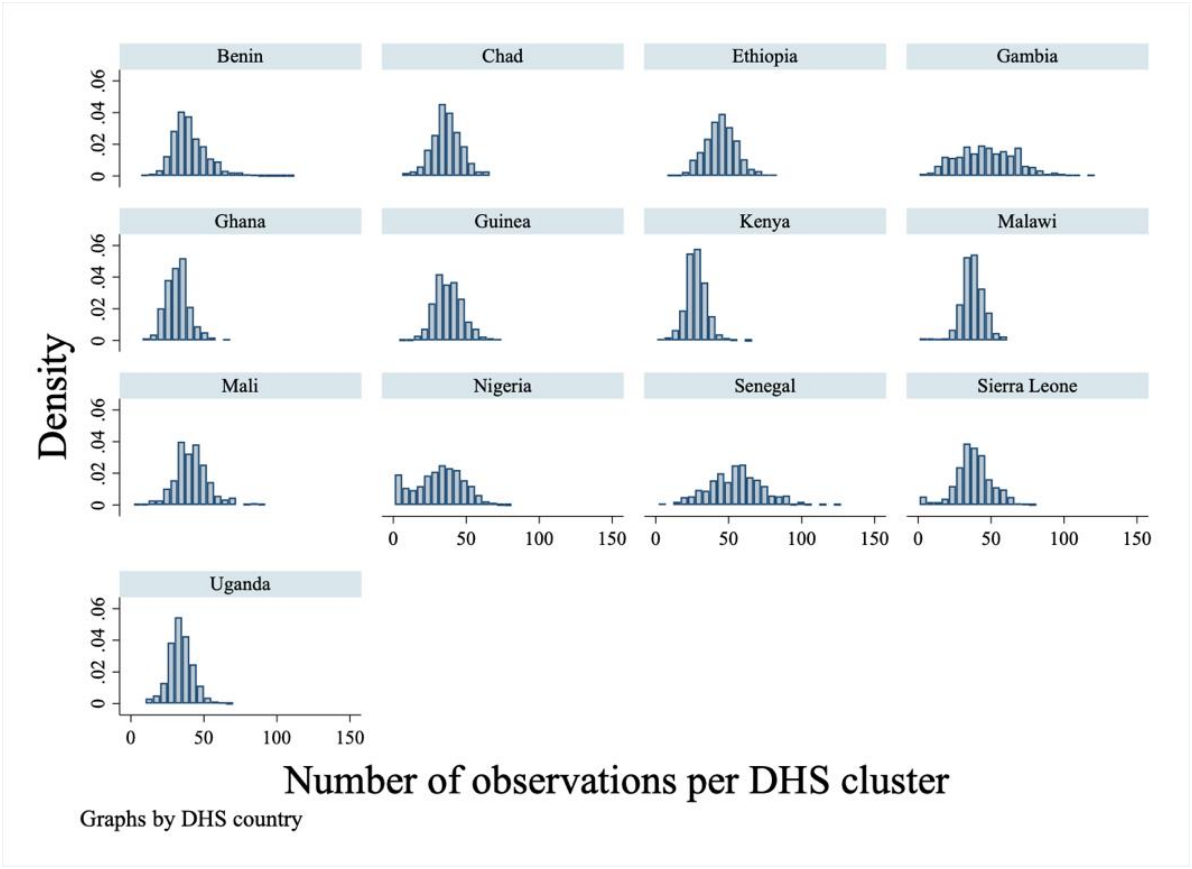
Table 3. Estimates of the Impact of Bible Exposure on Literacy

	Partial Literacy	Full Literacy
-10	0.022	0.002
	[-0.020, 0.064]	[-0.038, 0.041]
-5	0.001	0.000
	[-0.034, 0.036]	[-0.031, 0.031]
0	0.023	0.022
	[-0.009, 0.055]	[-0.008, 0.052]
5	0.044	0.055
	[0.012, 0.075]	[0.027, 0.084]
10	0.131	0.143
	[0.099, 0.163]	[0.116, 0.170]
15	0.181	0.139
	[0.142, 0.220]	[0.101, 0.177]
Num.Obs.	77088	77088

Note: This table shows the estimates of Bible exposure on the two literacy measures using Callaway & Sant'Anna (2020) estimates. The outcome *fully literate* takes 1 for respondents that can read the whole sentence, while the *partial literacy* takes 1 for individuals that can read the whole and part of a sentence (less concervative measure). DHS Controls include male and urban dummies. Estimations includes year of birth and language fixed effects. Formally, each column displays the estimates of β_a described in specification (2) in the Empirical Approach section.

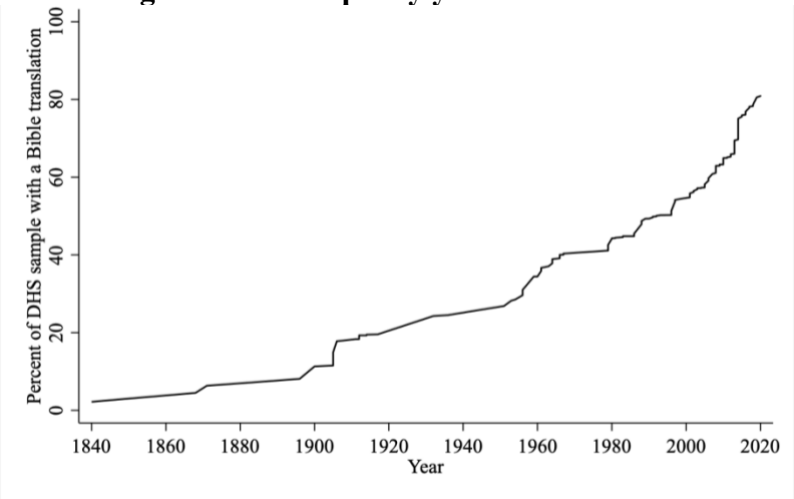
Appendix A: Figures

Figure A.1. Distribution of observations for DHS clusters by country



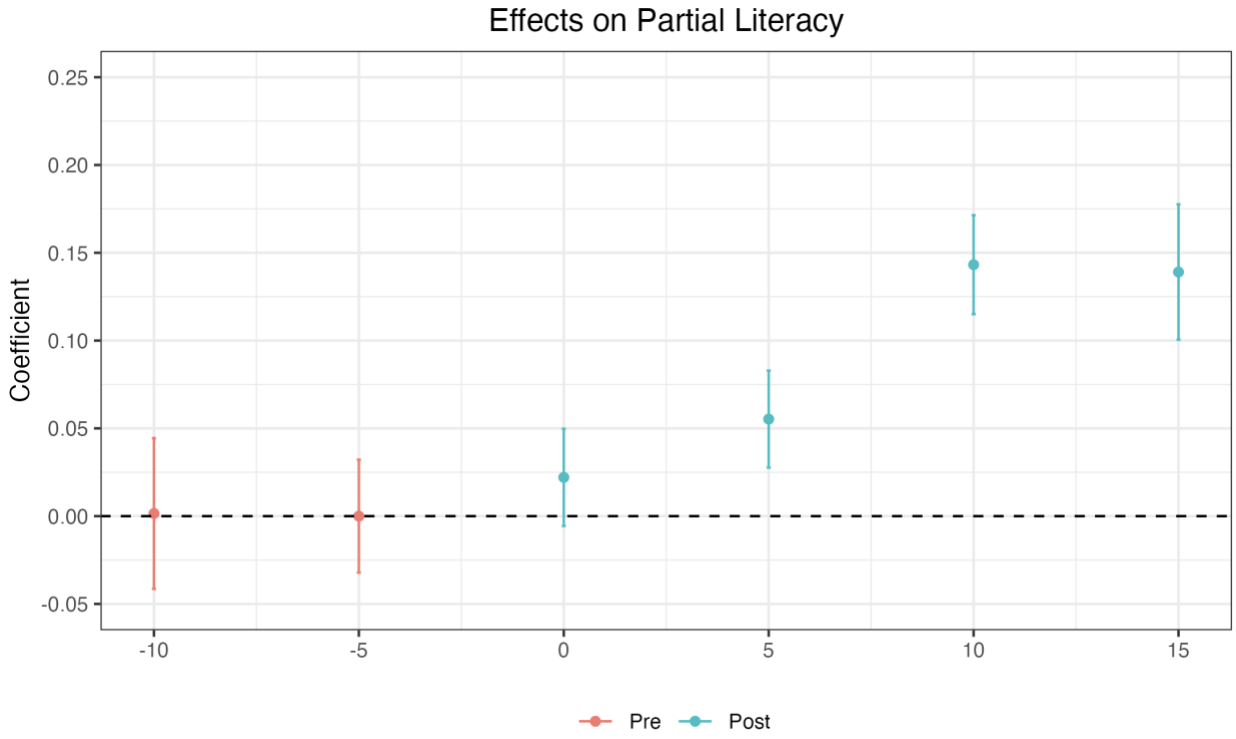
Note: This figure shows the distribution of observations in each DHS cluster by country.

Figure A.2. Percentage of DHS sample by year of birth with a Bible translation



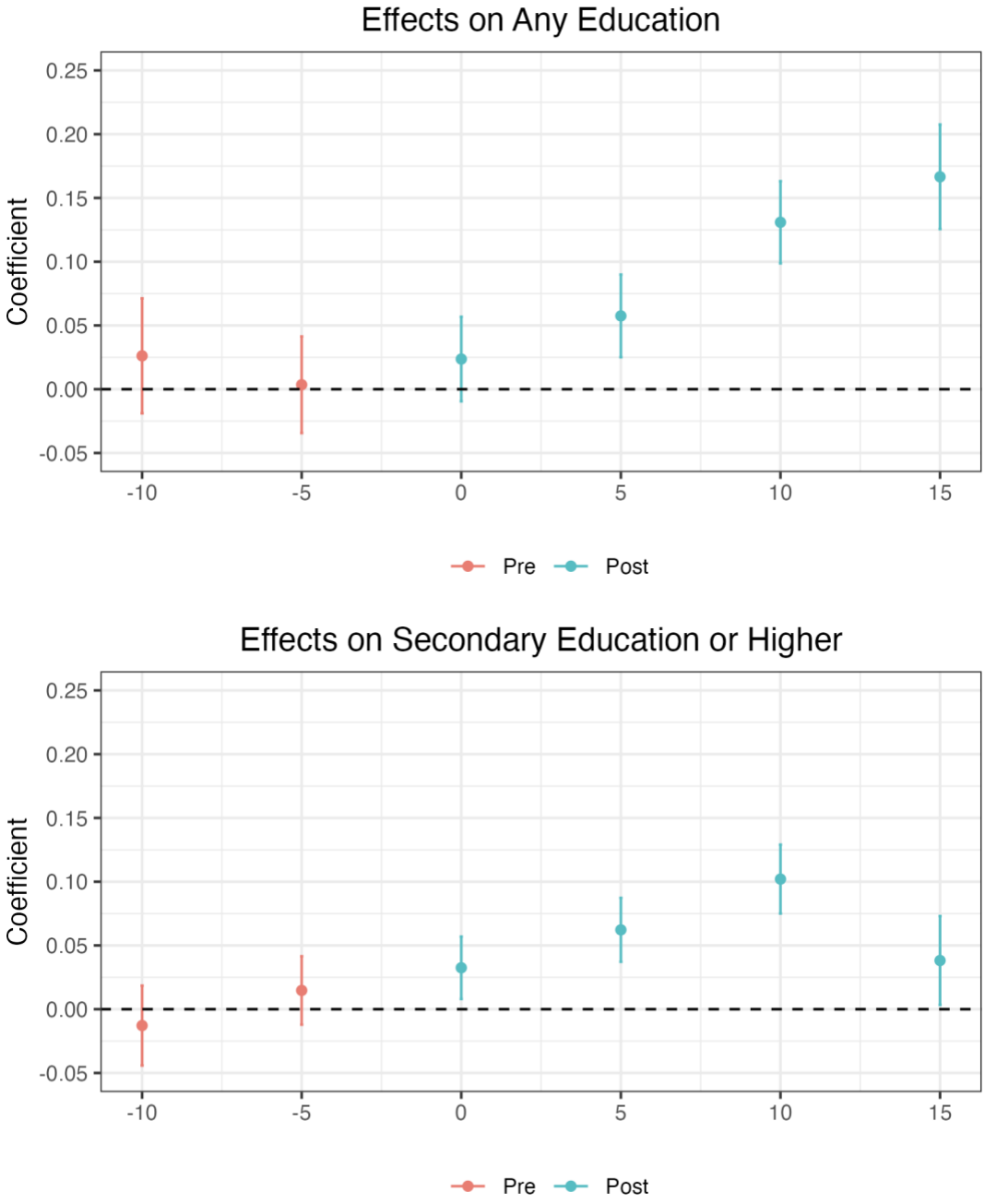
Note: This figure shows the percentage in our DHS sample that has a Bible available in their mother tongue after step one in linking our dataset. See subsection on linking datasets for further detail.

Figure A.3. Impact of Bible Translation on Partial Literacy



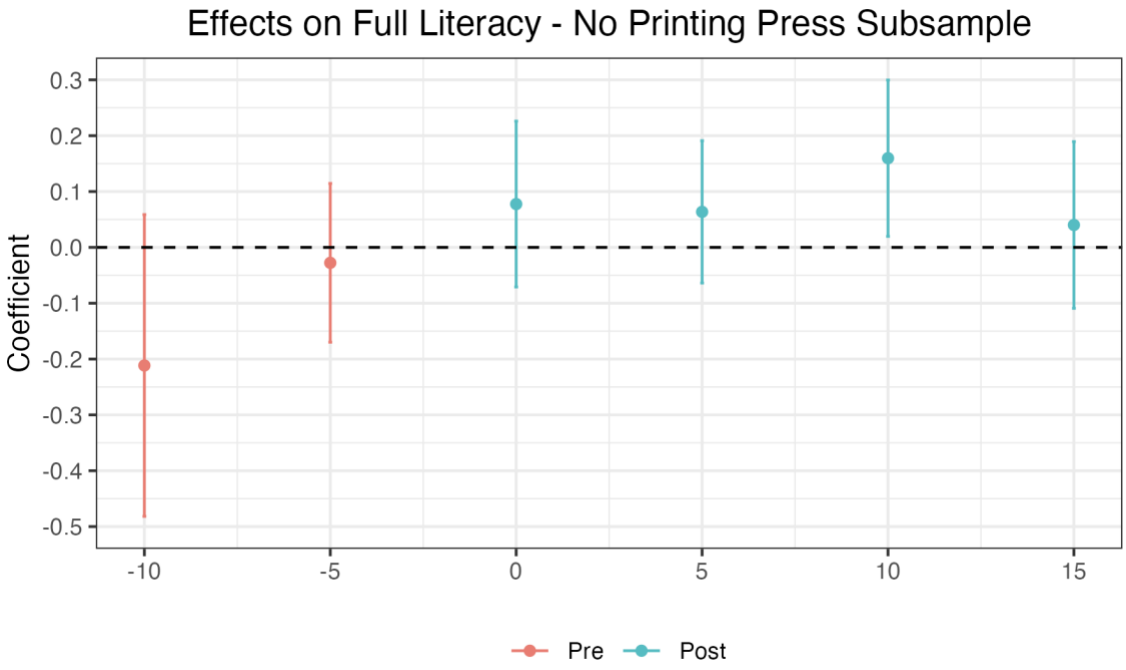
Note: This figure shows the effect of exposure to Bible translation on partial literacy when using the estimation method by Callaway & Sant'Anna (2020). For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

Figure A.4. Impact of Bible Translation on other Educational Outcomes



Note: These figures show the effect of exposure to Bible translation on two additional educational outcomes when using the estimation method by Callaway & Sant'Anna (2020)**.** The upper panel shows the any education (dummy variable that takes 1 when the respondent has 1 or more years of education) and the completion of secondary school or higher level (dummy variable that takes 1 when the individual finished secondary school)**.** For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

Figure A.5. Impact of Bible Translation on Literacy by Proximity to Missions without Printing Press



Note: This figure shows the effect of exposure to Bible translation on literacy when using the estimation method by Callaway & Sant'Anna (2020)**.** For further detail on the methodology, see the empirical approach section. Confidence intervals calculated using clustered standard errors at the ethnic group level and 5% significance level.

Appendix B: Tables

Table B.1. TWFE Estimates of Bible Exposure on Fully and Partial Literacy

	Fully Literate		
	(1)	(2)	(3)
-15	−0.036	0.020	0.018
	(0.033)	(0.018)	(0.017)
-10	0.007	0.012	0.011
	(0.028)	(0.013)	(0.011)
0	0.027	0.006	0.004
	(0.023)	(0.013)	(0.011)
5	0.044	0.020	0.020
	(0.043)	(0.022)	(0.018)
10	0.099	0.071***	0.070***
	(0.061)	(0.024)	(0.021)
15	0.168**	0.118***	0.116***
	(0.075)	(0.031)	(0.033)
Num.Obs.	19576	77231	192207
R2	0.251	0.376	0.352
Base Group	None	NT	NT + AT
Std. Errors Cluster	Ethnic group	Ethnic group	Ethnic group
DHS Controls	✓	✓	✓
Fixed Effects	✓	✓	✓

	Partially Literate		
	(1)	(2)	(3)
-15	−0.060**	0.023	0.027**
	(0.025)	(0.016)	(0.014)
-10	−0.016	0.006	0.009
	(0.028)	(0.016)	(0.013)
0	0.038**	0.006	−0.001
	(0.015)	(0.010)	(0.011)
5	0.069**	0.031*	0.025
	(0.028)	(0.018)	(0.018)
10	0.125***	0.083***	0.078***
	(0.039)	(0.020)	(0.021)
15	0.162***	0.118***	0.117***
	(0.038)	(0.023)	(0.023)
Num.Obs.	19576	77231	192207
R2	0.242	0.348	0.315
Base Group	None	NT	NT + AT
Std. Errors Cluster	Ethnic group	Ethnic group	Ethnic group
DHS Controls	✓	✓	✓
Fixed Effects	✓	✓	✓

Note: This table shows the TWFE estimates of Bible exposure on fully and partial literacy in Africa. The outcome *fully literate* takes 1 for respondents that can read the whole sentence. DHS Controls include male and urban dummies. Year of birth and language fixed effects. Formally, each column displays the estimates of β_a from specification (2) in section 4 using different sample restrictions in the control group. NT means "never treated," and AT denotes "always treated" see section 4 for further explanation on control group comparisons.

Appendix C

In this appendix, we show the relationship between earlier mother-tongue Bible translations and adult literacy, presenting the results for ethnic group-level treatment.

Effects of Earlier Translations on Literacy

Previous papers on the impact of Christian missions find selection regarding mission locations (Becker and Woessmann 2010; Kim 2020; Bai and Kung 2015; Waldinger 2017; Barro and McCleary 2017; Wantchekon et al. 2015). Since access to a Bible translation is likely associated with distance from a mission, we reduce the potential for bias by narrowing our sample to respondents living within 10 km of a mission.

In this approach, we exploit variation across ethno-language groups near a mission, comparing groups with and without access to mother-tongue Bible translations before 1970. Forty-five treated groups received a Bible translation before 1970, and 93 groups had no Bible translation available before 1970, as shown in Table C. 1. [17] We estimate the following equation:

$$literate_{ilmcdt} = \beta Bible1970_l + \delta X_i + \alpha Z_l + \theta_m + \gamma_c + \varphi_d + \alpha_t + \epsilon_{ilmcdt} \qquad (1)$$

where indices indicate individual i, language l, mission m, country c, DHS cluster d, and cohort t. Our outcome, $literacy_i$, is either full or partial literacy. The treatment variable is $Bible1970_l$, equal to 1 if a Bible translation for language l is available before 1970. Individual characteristics, X_i, include male and rural indicators. Ethno-language group characteristics, Z_l, include gathering dependence, hunting dependence, fishing dependence, agriculture intensity, marital composition, major crop types, and settlement patterns. We include fixed effects for each mission (θ_m), country (γ_c), DHS cluster (φ_d), and cohort (α_t). The mission fixed effects allow to capture the within-

[17] This table summarizes the individual sample lost and the ethno-language groups from the DHS cluster located within 10 km of each mission.

mission relationship between translations and literacy, accounting for time-invariant differences between missions. Finally, ϵ_{ilcmdt} is the error term. We are interested in β, which captures the correlation between of early Bible translations and literacy for individuals located near a mission. Finally, for correct statistical inference of estimates, we cluster our standard errors at the language group level (Abadie et al., 2017).

We show our findings for specification (1) in Table C.2. In column (1) of Panel A, access to a pre-1970 Bible translation is associated with an imprecisely estimated increase of 7.2 percentage points in the full literacy rate. In column (2), when we include indicators for male and rural and introduce fixed effects for DHS cluster, country, year of birth, and mission, the estimated coefficient becomes statistically significant but is smaller than in column (1). With these controls, access to an early Bible translation is associated with a three percentage point increase in literacy. Column (3) includes further controls for pre-colonial characteristics for each ethnic group, resulting a less precise estimate of a three percentage point increase in literacy.

The results are consistent with positive selection into Bible translation.[18] The estimates in column (1) are likely to be upward-biased due to differential pre-colonial characteristics and are driven mainly by largest the ethnic group, as shown in Table 2. While Jedwab et al. (2018) and Michalopoulos & Papaioannou (2013) argue that controlling for pre-colonial ethnic group characteristics as in column (3) corrects for bias, it is likely that this approach does not produce a causal estimate.[19]

[18] Another issue in the approaches that utilize mission locations come from nonrandom measurement error of historical maps (Jedwab et al. 2018; Fahs 1925)

[19] These results are consistent with the findings by Jedwab et al. (2018) and Michalopoulos & Papaioannou (2013), who argue that the characteristics at the ethnic group level before the missionaries' arrival correct bias in the estimates of the impacts of missions on various outcome.

Appendix C: Tables

Table C.1 Summary of ethnic groups in analytical sample for the simple approach to estimate the association between adult literacy and Bible translations before 1970

Percentage of observations from DHS individual data drop with the sample restriction <10km	81.3%
Total ethnic group in DHS before matching with missions	214
Number of ethnic groups drop with the sample restriction <10km	80
Final number of ethnic groups in DHS after the sample restriction <10k	138
Number of treated ethnolanguages with the sample restriction <10km	45
Number of control ethnolanguages with the sample restriction <10km	93

Note: This table summarize the final analyitcal sample of ethnic groups for Aproach I .The sample restriction "<10km" refers to the distance from the DHS cluster centroid to the mission geolocation.

Table C.2. Positive association of earlier translations on literacy in Africa

Panel A. Fully literate outcome

	Literate		
	(1)	(2)	(3)
Translation	0.072	0.030**	0.030*
	(0.047)	(0.012)	(0.015)
Mean Control	58.4	58.4	58.4
Std. Errors Cluster	Ethnic group	Ethnic group	Ethnic group
DHS Controls		✓	✓
Fixed Effects		✓	✓
Ethnolatlas Controls			✓
Num.Obs.	57337	57337	57337
R2	0.005	0.261	0.263

Panel B. Partially literate outcome

	Literate		
	(1)	(2)	(3)
Translation	0.110***	0.026**	0.026*
	(0.038)	(0.011)	(0.013)
Mean Control	69.2	69.2	69.2
Std. Errors Cluster	Ethnic group	Ethnic group	Ethnic group
DHS Controls		✓	✓
Fixed Effects		✓	✓
Ethnolatlas Controls			✓
Num.Obs.	57337	57337	57337
R2	0.016	0.267	0.270

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the effect of earlier Bible translations on fully and partially literate in Africa. Controls include male and urban dummies and the following "fixed effects": DHS cluster by country, country, year of birth, and closest mission. Ethnoatlas controls are dummies for each precolonial activity category, such as reliance on hunting, gathering, fishing, and agriculture; marital composition, agricultural intensity, major crop types, and settlement patterns (see Table 2 for further detail). The sample includes DHS clusters within 10km of a mission location.