School Clustering and Religious Competition: Persistence of Educational Inequality in Colonial and Post-colonial D.R. Congo*

Latest version

Pablo Álvarez-Aragón†, Catherine Guirkinger†, and Paola Villar‡

†*University of Namur*

‡*Université Paris Cité*

February 21, 2025

Abstract

This paper examines the persistence of educational inequalities in Africa, focusing on colonial and contemporary Congo. By analyzing historical and contemporary surveys, archives, and school administrative data, we document the continuing impact of colonial-era factors and explore mechanisms of path dependency. We find that contemporary educational inequalities are driven by the clustering of post-independence schools around historical Christian missions. This clustering dynamic has increased over time and is shaped by religious competition between schools of different denominations. This spatial pattern significantly affects girls' education more than boys' because girls are more sensitive to distance barriers, explaining a stronger colonial legacy for female education. Our analysis suggests that parental education and missions as agents of structural change have limited influence on regional educational outcomes in this context.

JEL Classification : I25, N37, F54

Keywords: Education, gender, school competition, religious competition, intergenerational transmission, Christian missions, persistence, Democratic Republic of the Congo.

---

*Corresponding Author: Catherine Guirkinger, catherine.guirkinger@unamur.be. P. Álvarez-Aragón and C. Guirkinger are affiliated to the University of Namur, Belgium. Emails: pablo.alvarez@unamur.be and catherine.guirkinger@unamur.be. P. Villar is affiliated to the University Paris Cité and to the University of Namur. Email: paola.villar@u-paris.fr. The authors would like to thank Jean-Marie Baland, Jean-Philippe Platteau, Augustin Bergeron, Etienne Le Rossignol, William Parienté, Denis Cogneau and Pauline Rossi for helpful comments, as well as participants of seminars and conferences at Namur, CY Cergy Paris University, University of Castilla-La-Mancha, the African School of Economics, Université Saint-Louis, Harvard University, University of Valladolid, the Bordeaux School of Economics, the Paris School of Economics, the University of Barcelona School of Economics, ENS Lyon, SAEe 2023, Université de Pau et des Pays de l'Adour, Bolzano, and the EUDN conference. Óscar Carpintero provided useful feedback at an early stage of the paper. C. Guirkinger gratefully acknowledges financial support from the European Research Council under the H2020 research and innovation programme/ERC grant agreement 759294. P. Álvarez-Aragón acknowledges financial support from an FNRS/FRESH doctoral fellowship.

1. Introduction

Regional inequalities in education and educational mobility are a persistent and worrying phenomenon observed in many countries. These disparities not only reflect stark inequalities of opportunity, but also threaten social cohesion by perpetuating cycles of disadvantage (Collier, 2024). Such inequalities are particularly pronounced across Africa (Alesina et al., 2021; Heath Milsom, 2023), with a wealth of literature linking them to patterns of colonial investment in education. In particular, regions that received greater colonial investment –whether through public or missionary schools– tend to have higher levels of human capital today, especially for women (Huillery, 2009; Nunn, 2014; Montgomery, 2017). However, the mechanisms behind this persistent inequality are complex and often difficult to identify due to limited data availability (Cagé and Rueda, 2016; Calvi et al., 2022; Okoye, 2022). Understanding these mechanisms is nevertheless crucial for designing effective policy interventions aimed at correcting these inequalities and preventing their reproduction. For example, if parental demand is the primary driver of these inequalities, conditional cash transfers could incentivize school enrollment in lagging regions. But, if the main problem is the unequal distribution of schools, expanding school infrastructure may be a more appropriate response.

In this paper, we offer new insights into the persistence of regional educational inequalities since colonial times by highlighting a factor neglected in the existing literature: the contemporary spatial clustering of post-independence schools around historical missions, which has increased over time, driven by religious competition between schools of different denominations. Focusing on the Democratic Republic of the Congo (DRC), we proceed in two steps. First, we examine regional inequalities in education today and then trace their roots back to the colonial period using unexplored individual-level data from a sample of about 250,000 individuals born during the colonial period or immediately after independence. Specifically, we assemble a comprehensive dataset of missionary posts and, using the state-of-the-art set of controls and robustness checks, find a strong correlation between missionary presence in 1948 and educational attainment today, especially for women. Importantly, this correlation does not weaken over time. To determine whether this correlation reflects the persistence of educational inequalities, we use a difference-in-differences analysis to compare educational outcomes

in the colonial period before and after the establishment of new mission posts between 1930 and 1948. Our results show that the opening of these missions significantly raised local educational levels during the colonial period (for both men and women).

The main contribution of our study lies in the second part of the analysis, where we explore the mechanisms of path dependency that have sustained these educational inequalities over time. First, we use administrative data on contemporary schools in the DRC and show that the current distribution of schools exhibits pronounced clustering around historical missions. Specifically, dividing the country into small 20km x 20km grid cells, we find that an additional historical Catholic (Protestant) mission in the grid cell is associated with a 119% (57%) increase in the average number of primary schools per 1000 inhabitants today, or with a 133% (68%) increase in the average number of contemporary secondary schools. Examining the dynamic pattern of secondary school openings reveals that the establishment of a school affiliated with a particular denomination leads to the opening of a school of a competing denomination in the same area (holding time-invariant characteristics constant at the grid-cell level), suggesting that religious competition through schools is a driver of these clustering dynamics. Notably, and perhaps worryingly, the spatial concentration of schools appears to increase over time.

These results are consistent with historical accounts of religious competition through (missionary) schools (Gallego and Woodberry, 2010) and with recent policy reports arguing that religious organizations open schools to attract new adherents (and possibly money), especially where other organizations are already active (Gauthier et al., 2021; Briand and Nicolai, 2021).[1] Yet, to the best of our knowledge, the impact of (religious) school competition on the geographic location of new schools today has not been formally demonstrated. This gap of research is of significance as the clustering of schools driven by competitive forces is likely to operate in many contexts where the private education sector is important and largely unregulated, so that private (religious) actors, competing with each other, choose where to open schools. We show that the clustering of schools is unlikely to be driven by other supply-side mechanisms such as the endogenous location of missions, or structural transformation (using indicators such as

[1]They also fit in well with the standard economic approach to religion based on the idea that denominations compete with each other in a religious market (Iannaccone, 1998).

nightlights or labor market structure).

Finally, we examine whether demand-side mechanisms, such as the intergenerational transmission of educational demand, are important drivers of educational inequality. We rely on a representative sample of urban migrants educated during the colonial period to isolate a "pure demand channel" in the transmission of education. Using a difference-in-differences specification, we examine how parental education (and thus exposure to colonial schools) affects children's educational outcomes. We find that in a context of significant educational mobility, parental education (as measured by their exposure to colonial missionary schools) has only a modest impact on children's education and hence on long-term educational inequality. In short, our results suggest that these factors are less influential than the dynamics of school clustering, which appears to be a key driver of the persistent educational inequalities observed today.

The remainder of this paper is organized as follows. Section 2 reviews the literature, highlighting the main contributions of this paper. Section 3 presents the data and measures. Section 4 provides evidence on the impacts of colonial investments in the DRC on the education of people born during the colonial period as well as on the education of subsequent generations, starting with a description of the context. Section 5 discusses and tests the supply-side mechanisms at work, while Section 6 presents those on the demand side. Section 7 concludes the paper with a discussion of the gendered aspect of persistence and the religious differences that our analysis reveals.

2. Contributions and literature review

Long-term effects of colonial investments, particularly Christian missions.–There is an extensive literature in economics on the long-term consequences of Christian missionaries on different outcomes in Africa (see, for example, Cagé and Rueda 2016, 2020; Valencia Caicedo 2019b; Calvi et al. 2022; Nunn 2014; Montgomery 2017 and the extensive reviews by Becker 2022 and Okoye 2022). In the specific context of the DRC, Bergeron (2020) highlights the influence of historical exposure to Christian missionaries on attitudes toward in-group versus out-group members, while Le Rossignol et al. (2023) examine how exposure to the historical missionary presence influences the *quality* of education and students' performance today.

One of the main shortcomings in the above literature is the lack of historical microdata

to explore the evolution of African societies as a result of missionary activities and document the mechanisms behind the long-run impact of missions (Okoye, 2022). Our paper addresses this limitation in two ways. First, we investigate the impact of colonial investments in education over several generations (mobilizing different sources of data) from 1930 to the present. Our novel colonial data allow us to estimate the effect of exposure to Christian missions on the first generations of individuals directly exposed to them, in line with Wantchekon et al. (2015).[2] Moreover, by exploiting the *timing of the opening of posts* in the spirit of difference-in-difference estimation, we rely on less demanding assumptions than the existing literature.[3] Second, we focus on understanding persistence mechanisms. In particular, determining whether the persistent effects work through the supply or the demand side has been shown to be challenging (Cagé and Rueda, 2016).

Several authors mention supply-side factors to explain the persistence of colonial investments (not only Christian missions) on educational outcomes. For example, Huillery (2009) argues that initial investments in education lead to continued funding in the post-independence in French Africa (in contrast to Ben Salah et al. (2022) for the case of Tunisia). In line with our findings, Okoye and Pongou (2023) show that people living near historic missions in Nigeria are now more likely to live in districts with existing schools. Using administrative data on the universe of secondary schools in contemporary DRC, we explore this question at a more granular level and propose a mechanism to rationalize this finding: religious competition, operating through school construction, shapes the geography of schooling supply.

School and religious competition.–While there has been considerable research on school competition, to the best of our knowledge, none has focused on its effects on school location.[4] Moreover, most studies focus on developed countries. Exceptions include Bau (2022), who examined the negative effects of competition on low-income students in Pakistan. Closer to our context, Gallego and Woodberry (2010) focus on the impact of religious competition at play *during* the colonial period on the quality of edu-

[2]In the same line and also in the context of colonial Congo, Guirkinger and Villar (2022) study the impacts of pro-natalist policies on women fertility during the colonial period.

[3]By constructing an original and comprehensive dataset of missions, we also address Jedwab et al. (2022) criticism of altas-based sources that disproportionately report a selection of the best missions.

[4]Spatial location models have been used to study "product differentiation" among religious organizations (see Iyer, 2016 and Iannaccone, 1998 for detailed reviews).

cation today in Africa. In particular, they hypothesize that the large differences in educational attainment between Catholic and non-Catholic states in Africa today are largely explained by market forces operating during the colonial era, and these institutional differences have persisted to the present because of their high degree of inertia. Because Catholic states favored the Catholic Church and restricted the entry of Protestant missions, Catholic schools in these states invested less in school quality than their counterparts in non-Catholic states, which feared competition from Protestants. In non-Catholic states, there was no regulation favoring a particular denomination, and both had to invest in school quality to attract converts. In this paper, we take a different approach by focusing on the post-independence competitive dynamics triggered by the initial location of missions and provide quantitative evidence on how this post-independence competition induces a clustering of schools from different religious networks.

Intergenerational transmission, educational mobility, and their determinants. Little is known about these issues in the context of developing countries (for recent reviews regarding developed countries, see Black and Devereux (2011) or Solon (2018)). Important exceptions are Alesina et al. (2021) and Heath Milsom (2023), who highlight the existence of a strong "regional effect" in educational mobility. To measure this effect, they rely on children of migrants to compare siblings exposed to different "regions" while holding parental education constant. In a sense, we do the opposite, comparing children of migrants who live in the same city (and are therefore exposed to the same local environment), but whose parents were differently exposed to colonial education. This allows us to isolate a "pure demand effect" from regional exposure effects.

3. Data and Variables

3.1. Data

Contemporary data.–To study persistence in education and some of its underlying mechanisms, we rely on contemporary demographic information.

Demographic and Health Surveys (DHS): We use all available rounds of the DRC DHS (2007-2013), a nationally representative survey that provides detailed information on education, literacy, occupation, and religion. In total, we have information on about 40,000 individuals living in 785 clusters. Since we have information on the latitude and

longitude of DHS clusters, we are able to combine this dataset with the historical colonial schools data detailed above.

Contemporaneous schools: We retrieved data on the universe of primary and secondary schools in the DRC in 2020. This dataset is provided by the Ministry of Education and includes school-level information such as exact GPS location, number of students and teachers, the date of opening (only for secondary schools) and school management regime. The education system in the DRC consists of two types of schools: publicly funded and private. Public schools are divided into official schools and schools under contract with the state. Official schools are run by the state and are secular, while schools under contract are run by the four religious organizations that have signed an agreement with the state (Catholics, Protestants, Kimbanguists, and Muslims). While they are managed most often by faith-based organizations, private schools do not receive any state subsidies (new evangelical churches typically operate private schools).

In our data, there are approximately 60,000 primary schools, of which nearly 20% are Catholic (public), 37% are Protestant (public), 17% are official schools, 12% are managed by other religious groups (public), and 14% are private (unfortunately these data do not provide information on whether these private schools are denominational or not). As for secondary education, there are 35,000 secondary schools, of which 15% are Catholic (public), 35% are Protestant (public), 18% are official schools, 15% are administered by other religious groups (public), and 17% are private.

Historical data.–We construct our main historical dataset by combining surveys containing information on individuals born during the colonial era with data on the location, type and opening of Christian missions.

Demographic Survey from the 1970s: These data provide original individual-level information on 250,000 individuals in 46,000 households surveyed between 1975 and 1977 in seven major cities in the DRC: Kinshasa, Matadi, Bandundu, Kikwit, Mbandaka, Kananga, and Bukavu (1/10 of the total number of households identified in each of these cities are surveyed). Demographic information was collected on age, sex, ethnicity, marital status, place of birth, and migration status, as well as socioeconomic characteristics such as years of education and occupation. Among respondents born before independence, 80% are migrants, and 85% of them migrated to the cities after independence (or shortly before) and were therefore educated in their place of birth. In fact, migration

prior to the last years of the colonial era was highly controlled and restricted. Importantly, this survey includes information on the position of individuals within the household, which allows us to link parents and children and to study educational mobility. We retrieved these data from digital archives at the University of Louvain-La-Neuve in Belgium. These data are considered high quality by demographers who rely on them to study demographic dynamics in the Congo in well-published academic papers (Tabutin, 1982; Shapiro, 1996).[5]

Missionary activities: During colonial times, virtually all education took place in mission schools. To measure colonial investment in education at the local level, we construct a database with comprehensive information on Christian mission posts opened in the DRC between 1885 and 1948 (see Figure A1 in Appendix A.1 for the geographic location of Catholic and Protestant missions in the DRC in 1948.). The data on Catholic missions come from information digitized from comprehensive yearbooks and maps published in 1924, 1935, and 1949 (Corman, 1924, 1935; Van Wing, 1949). As for the location and date of opening of Protestant missions, we obtain them from maps published in 1905, 1921, 1930, 1944, 1953, and 1960, and from a 1978 handbook (Irvine, 1978). The recorded Protestant missions cover the period 1879-1960, while the recorded Catholic missions cover the period 1885-1948. To make the results more comparable, we focus on the period 1885-1948. The final sample of missions includes a total of 697 missions for the DRC, of which 300 are Protestant and 397 are Catholic. Importantly, we know the date of arrival of Catholic nuns (if any).[6] On the other hand, Protestant missions almost always had female missionaries.[7] We will use this information when examining the gendered effects of exposure to missionary presence, since the presence of nuns within the Catholic school system was a necessary condition for the education of girls.[8]

For the DRC, these data on missions are far more complete than most of the existing

[5]These surveys were conducted during Mobutu's dictatorship, but both the design and implementation of the survey and data collection were managed jointly by a team of demographers based at the Congolese Institute for Research and Statistics and the Université Catholique de Louvain in Belgium. Data cleaning and statistical programming were carried out in Belgium.

[6]Catholic nuns mostly arrived at pre-existing missions run by male missionaries. In fact, the date of arrival of Catholic nuns is the same as the date of establishment of the mission post only 6% of the time.

[7]According to Irvine (1978), about 90% of Protestant posts hosted at least one female missionary.

[8]This was true until 1960. After independence, girls were also educated in missions without Catholic nuns.

data on missions used in the literature, enabling us to address one of the criticisms raised by Jedwab et al. (2022), namely that mission atlases exacerbate endogeneity in mission location by reflecting only the best missions (Appendix A.1 shows a comparison of our dataset with the missions reported for Congo in Cagé and Rueda (2020) and Nunn (2010), who use information provided by Streit (1929) (Catholic missions), Beach (1903) (Protestant missions), and Roome (1924)).[9]

Determinants of missions' location: Finally, since missionaries were likely to locate in better areas, it is crucial to consider the geographic and historical characteristics that influence this decision when we examine the lasting influence of missionary presence. Building on Jedwab et al. (2022), our control variables include: altitude, slope, rainfall, length of growing season, distance to the coast, distance to a colonial road, distance to a colonial railroad, distance to a navigable river, population density in 1900, cropland area in 1900 as estimated in the HYDE 3.1 database (Klein Goldewijk et al., 2011), malaria index from Cagé and Rueda (2016), tsetse suitability index from Alsan (2015) and number of slaves exported in the Indian and Atlantic trades (normalized by ethnic homeland area) from Nunn and Wantchekon (2011). We construct these historical and geographical controls at the DHS cluster level and the territory level.

3.2. Measures of exposure to missions

Since missionaries provided virtually all education during the colonial period, and since all missions operated schools during our study period, we use exposure to missionary presence as a measure of exposure to colonial investment in education. However, we distinguish between three types of missions - Catholic, Catholic with nuns, and Protestant - since the type of education they offered slightly differed (see section 4.1 for details). Our construction of exposure variables differs across datasets, as they provide different levels of precision regarding the location of the individual place of birth / location (in the DHS we know the GPS coordinates of the sampling cluster, while in the historical demographic survey we have information on the territories of birth). With DHS data,

[9]The comparison shown in Appendix A.1 refers to the final dataset used in both papers. Even if we restrict our data to missions opened before the year they use (1903 for Protestant missions and 1929 for Catholic in the case of Cagé and Rueda (2020), and 1924 in the case of Nunn (2010)), we have 62 Protestant missions and 199 Catholic, while Cagé and Rueda (2020) have 26 Protestant and 69 Catholic. If we compare the number of missions in the 1924 Roome map, we have 166 Protestant missions and 156 Catholic, compared to 120 Protestant and 93 Catholic in the Roome map.

we measure exposure to former missionary presence as the logarithm of the distance from the DHS cluster to the nearest mission post, multiplied by (-1).[10] We follow Cagé and Rueda (2016) and always restrict our sample to DHS clusters located less than 50km away from the closest mission, reducing the concern that DHS clusters are similar to distant places, and omitting regions with extreme values of the explanatory variable.

For the historical analysis, we construct an average exposure to missionary presence for each territory and year and assign this measure to each respondent according to their year of birth.[11] To do this, we follow Calvi and Mantovanelli (2018) and construct a continuous measure of proximity at the territory level that controls for mission density. Specifically, we generate 1000 random points within each territory and compute the distance from each random point to its closest mission before averaging over these distances. A major advantage of this method is that it takes into account missions in neighboring territories when calculating the exposure of a given territory. We repeat this process for each territory, each year between 1885 and 1948, and the three types of missions. Finally, based on these distances we construct a measure of proximity by multiplying the logarithm of the distance by (-1). Table A1 in Appendix A.2 provides descriptive statistics for these exposure measures.

4. Impacts of colonial educational investments through time

After a brief description of the historical context, we first examine the long-lasting impact of colonial schools on current educational outcomes in the DRC using DHS data (introducing two dimensions of heterogeneity: gender and religious affiliation of these schools at independence). We then turn to the question of whether missionary investments in education led to higher average educational levels near mission posts during the colonial period (which may have persisted). This is an important first step in understanding the mechanisms of persistence. However, while this is often an (implicit) claim, it is rarely formally tested.

[10] We use distances in logarithms to remove the high skewness in the distribution of distances and because we expect the influence of mission to be fundamentally nonlinear, with individuals located close to missions being disproportionately affected by the presence of the mission. We provide robustness estimates with alternative measures of exposure.

[11] We choose exposure at birth primarily because it allows us to include individuals in our sample who were born up to 1948. For example, if we instead compute exposure at age 6, we can only include individuals born before 1942, since our last mission opening occurs in 1948.

4.1. *Historical and present-day context*

Colonial Schooling.–In the Belgian Congo, both Catholic and Protestant missions invested heavily in mass education and played a crucial role in the development of the educational system. By 1960, at independence, 97% of all students were attending missionary schools (Yates, 1976).[12] Although the primary intention of missionaries was to convert Africans to Christianity, they saw the provision of formal education as the most effective way to do so (Meier zu Selhausen, 2019).[13] As in most African colonies, the colonial state encouraged mission investment in education by providing direct subsidies or indirect benefits (such as infrastructure or land concessions) to missions that operated schools. In the Congo, however, Catholic missions benefited from preferential treatment by the colonial authorities. Unlike Protestant missions (mostly from Britain, the U.S., or Sweden), they were typically "national missions" and were perceived as more loyal to Belgium and easier to work with or control. According to Boyle (1995), by the end of the 1950s, about 11% of the total number of students were enrolled in Protestant schools, while about 85% were enrolled in the Catholic network.

There were other important differences across denominations. Catholic missions followed an official program set by the government (to qualify for subsidies), while Protestant missions were free to design their own curricula. In Catholic missions, girls and boys were in separate classes, and only nuns could be responsible for girls' education (Masandi, 2004). Thus, opening classes for girls was only possible if nuns were present.[14]

At independence, the Congo did very well compared to other colonies in terms of overall primary school enrollment: as much as 71% of the appropriate age group was enrolled in primary school, the highest rate in Africa. However this overall figure

[12]In fact, the Belgian Congo had the highest number of Western missionaries in Tropical Africa, reaching almost 4.000 in the eve of the World War II, as compared to about 3.500 in the entire region of East Africa during the interwar years (Frankema, 2013; Bergeron, 2020).

[13]The focus on education and the reach of missionaries were global, leading to what some authors have called a "first global mass education wave" (Valencia Caicedo, 2019a).

[14]The late arrival of nuns in the territory (around the mid-1920s) could explain (in part) the delay in women's education in the colony (most Protestant posts had at least one female missionary at the opening of the post (Irvine, 1978)). The program focused on domestic skills such as cooking, child care, farming, and dressmaking. Depaepe and Kikumbi (2018) and Yates (1982), argue that the education provided by the Protestants was better in terms of quality, especially for girls who were in the same classes as boys: the curriculum was thus less focused on teaching practical, traditionally "female" skills. Protestant missionaries thereby promoted a different ideal of a Christian woman than the Catholics. In Appendix B, we provide additional evidence on the differences between Catholic and Protestant education.

masks significant gender disparities with girls far less likely to attend school than boys (Masandi, 2004). Also, very few students were enrolled in secondary schools, accounting for less than 2% of total enrollment, and only a few dozen university students had graduated.[15]

Post-independence Schooling.–After independence, the newly elected government committed to expanding its education system and subsidized schools that met minimum standards. In addition, the differences in education between boys and girls were officially abolished in 1962 (André and Poncelet, 2013).

In the 1960s, new secondary schools opened massively, mostly still run by Christian denominations, and especially by Catholic congregations (Lembagusala Kikumbi, 2018): in the early 1970s, 62% of primary school children were enrolled in Catholic schools, compared to 21% in Protestant and 14% in official public schools.[16]

In the mid-1980s, Mobutu nationalized schools but quickly a deep crisis hit the education sector and the management of the system was delegated to representatives of the country's four main religions (Catholicism, Protestantism, Islam and Kimbanguism) in return for grants and subsidies, while private actors were authorized to open schools. In the face of significant financial challenges, both private and public schools have sought parental contributions to cover a substantial portion of the schooling expenses, particularly teachers' salaries. These private contributions continue to represent about two-thirds of the countrys basic education costs (Briand and Nicolai, 2021).

Since the beginning of the 2000s (and with the end of the civil wars and economic turmoil), the situation regarding education has clearly improved, and today, DRC outperforms other income-comparable countries (such as Niger, Mali and Burkina Faso). Yet, in their detailed report on education in the country, Gauthier et al. (2021) show that geographical inequality remains high and public planning and regulation by the State are lacking. In many places, only some religious networks, in particular the Catholic and Protestant churches, are present. These religious schools are said to rely on school openings as a tool of evangelization to draw in more converts. In addition, evidence

[15]Gender disparities are also striking in higher education. Lauro (2020) notes that there was no woman at the university and only one among the 800 general high school graduates at independence.

[16]In secondary education, the state played a more active role: one third of students attended an official school, while 42% (resp. 17%) attended a Catholic (resp. Protestant) school in 1973 (MacGaffey, 1982). It is worth noting that even in the official secondary schools, the teachers were mostly from (Belgian) Catholic orders (Sheline et al., 1984).

in Briand and Nicolai (2021) suggests that faith-based organizations use a significant proportion of the households' school fees to finance their church activities.

4.2. Present-day education and missionary schools in the past

To investigate whether people living today cluster to historical missions are more educated, we follow the existing literature by estimating a linear regression model of contemporary outcomes on the proximity of individuals' cluster in the DHS survey to the nearest Christian mission, controlling for historical and geographic characteristics.[17] For a formal exposition of the regression model, see Appendix C.1.

Results are reported in Table 1. We find that, today, *(i)* proximity to a Christian colonial school matters for women's education but not for men's and *(ii)* Catholic and Protestant colonial schools have similar effects.[18] In terms of magnitude, for women, halving the distance to a former Catholic or Protestant school is associated with 0.2 additional years of schooling (a 4% increase relative to the female sample mean), a 4 pp increase in literacy (an 8% increase), and a 3 pp increase in the likelihood of attending post-primary school (an 8% increase).[19] These effects therefore contribute to reducing gender inequality in education. For example, halving the distance to a Catholic or Protestant mission reduces the gender gap in literacy by 15%. Interestingly, exposure to colonial schools has no effect on women's likelihood of completing secondary education. For men, on the other hand, it is only in this dimension (completion of secondary education) that we find an effect of proximity to Catholic schools: halving the distance to a Catholic school increases the probability of completing secondary education by 1 pp (or an increase of 9%).

Our results on contemporaneous education confirm the general findings in the litera-

[17] In addition, we follow the strategy introduced by Cagé and Rueda (2016) and restrict the analysis to those individuals living in DHS clusters with at least one mission within a 50 km radius. The idea behind this restriction is that areas located far from mission posts most likely present substantial differences from those closer to posts, not only in terms of geographical attributes but also institutional or cultural factors. In order to enable comparisons between similar areas on these dimensions, we remove these remote areas from the sample. The main threshold chosen here to define proximity is 50 km. Cagé and Rueda (2016) use 100km as their threshold, but their respondents' cities are on average 140km away from the nearest mission, while in our case, due to the higher availability of data, it is 22km for Catholics and 32km for Protestants. The choice of this threshold does not affect our results. Very similar results are obtained by choosing different thresholds (i.e. 30km, 40km or no threshold).

[18] Using the notation introduced in Section C.1, we cannot reject the null hypothesis that $\beta_C + \theta_C = \beta_P + \theta_P$.

[19] The final effect of halving distances is obtained by multiplying the coefficient reported in the table by $ln(2)$, since $[-ln(d)] - [-ln(\frac{d}{2})] = ln(2)$.

ture regarding the persistent effects of missionary presence on education today: women living in places close to colonial schools are, on average, more educated than women living in more distant places. However, we also find a persistent effect of both historical Protestant and Catholic missions, which is consistent with the results in Baten et al. (2021) or Becker and Meier zu Selhausen (2023), where no differences in outcomes are found between these two religions.[20]

Robustness. – In Appendix C.2, we provide further discussion and analysis in order to rule out a spurious correlation between mission location and current educational outcomes. Specifically, we first show the (lack of) sensitivity of our results to the inclusion of different combinations of relevant control variables (historical determinants of missionary location). Then, we show that our results hold when we control for contemporary urbanization, or when we restrict the analysis to rural DHS clusters only (the latter two analyses help control for the fact that missions may have located in more densely populated areas). We also show that our results are robust to alternative definitions of exposure (defining discrete buffers of exposure intensity) and are not explained by spatial autocorrelation. Finally, we show that missions left a lasting impact on the main dimension they targeted: religion. We show that proximity to historical Catholic missions increases the probability of being Catholic, while proximity to historical Protestant missions increases the probability of being Protestant.

[20] In contrast, Nunn (2014) finds a significant effect only for Protestant missions.

Table 1: Education in 2000 and exposure to former missionary presence

	(1) Years of education	(2) Literacy	(3) Secondary or more	(4) $>$ 12 years of educ
Exposure to Catholics	0.0720	-0.0118	-0.00398	0.0197**
	(0.119)	(0.0108)	(0.0117)	(0.00928)
Exp to Catholics x Female	0.273***	0.0576***	0.0475***	-0.0127*
	(0.0712)	(0.0102)	(0.0103)	(0.00649)
Exposure to Protestants	-0.00320	-0.0143	-0.0107	-0.000604
	(0.121)	(0.0119)	(0.0136)	(0.00923)
Exp to Protestants x Female	0.348***	0.0564***	0.0525***	0.0110
	(0.0689)	(0.00946)	(0.0106)	(0.00722)
FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Mean Y	6.933	0.637	0.566	0.165
Mean Y female	5.678	0.505	0.435	0.102
R-squared	0.304	0.202	0.232	0.165
N	34654	34654	34654	34654

Note. Data: Pooled 2007 and 2013 waves of the DHS. The sample includes women aged 15-49 and men aged 15-59. The sample is restricted to people living in DHS clusters located 50km or closer to a mission. The table reports OLS estimates. We use $-log(distance)$ to measure exposure. Outcome variables are defined as follows: single years of education in column (1), ability to read a whole sentence in column (2), probability of having post-primary education in column (3) and probability to have 12 or more years of education in column (4). Standard errors () are clustered at DHS cluster level. Geographical, historical and individual controls included (see Section 4 for details). *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

4.3. Persistence is not fading!

The DHS samples include generations of adults educated in the first decades after independence as well as adults born in the early 1990s, so it is possible to examine whether the persistent effects highlighted above fade over time. Examining fading is informative because different mechanisms lead to different predictions regarding the fading of initial inequalities over time. If the mechanism of persistence is the intergenerational transmission of education from parents to children, it is expected to fade over a small number of generations.[21] If, instead, there are persistent regional inequalities in the supply of schools, fading would arguably be slower (Arenas and Hindriks (2020) develops a model to formalize this idea and argue for school equalization and desegregation policies in the US). Consistent with this argument, Ben Salah et al. (2022) find that in Tunisia, the effect of exposure to colonial public education on literacy declines over time because of the strong expansion of public education.

In our case, the correlation between exposure to missions and education is remarkably

[21] If the correlation between parental and child education is α, it is only α^2 between grandparents and grandchildren and α^3 over three generations. Even if it is as high as 0.6 for one generation, it is only about 0.2 after three generations.

stable over time: we find no heterogeneous effect of exposure to missionary presence by decade of birth (results are reported in Table C6 in Appendix C.3).[22] Taken at face value, this would point to persistent inequalities in the supply of education. In Section 5.1, we turn to an examination of supply factors and explore the above questions in more details.

4.4. The impact of early missionary schools on the expansion of colonial education

To systematically examine the relationship between the opening of mission schools and education *during colonial times*, we draw on microdata containing information on the education of individuals directly exposed to colonial schools (using individual-level data from the 1970s Urban Demographic Survey) and combine it with information on the timing of the opening of new mission posts during the colonial era in the territory of birth of the individuals. This allows us to estimate a difference-in-difference model, leveraging changes in exposure across cohorts born in the same territory. In practice, we measure exposure to missions (of different types) at birth and include territory of birth fixed effects in our regressions.

Our identification strategy helps to overcome a major weakness of the existing literature on the impact of missions (or other colonial investments): any time-invariant characteristics (e.g., related to geography or the environment) correlated with mission placement are absorbed by the territory-fixed effects and cannot bias our estimates. We distinguish between Catholic and Protestant missions and we add a variable to capture the effect of exposure to Catholic missions with nuns. This is because the opening of schools for girls within the Catholic system was conditional on the presence of at least one nun.[23] [24] For a formal presentation of the regression model, see Appendix D.1.

[22]Using the DHS personal record files (PR), which contain information on household members under the age of 18, we even observe persistent effects for children aged 12 to 17. However, we do not find an effect on children aged 6 to 11, suggesting that persistence today is driven by access to secondary schooling (see Table C7 in Appendix C.4).

[23]In the previous analysis, based on DHS data, we did not include exposure to Catholic missions with nuns. This is because here we only observe missions at one point in time (1948) and by that time a large proportion of Catholic posts counted with at least one nun, leading to potential multicollinearity ($\rho = 0.70$). In addition, many more Catholic posts began to include nuns after 1948, so that the distinction likely became less relevant after 1948. Nevertheless, when we distinguish between exposure to Catholic missions with nuns and Catholic missions without nuns, our results are left unchanged.

[24]When using the 1970s survey, we focus on individuals born between 1930 and 1948, since the data include few migrants born earlier (and heterogeneity in education is small for them).

Table 2 displays the results. Two main findings stand out. First, the effect of exposure to missionary presence is only detectable for Catholic missions. In terms of magnitude, Column (4) shows that halving the distance to a Catholic mission increases years of education by 0.5 (11% of the sample mean) or, in Column 6, the probability of completing primary education by 6 percentage points (13% of the sample mean). The effect of Protestant missions is not only statistically insignificant but also small in size. The contrast between Catholic and Protestant is in line with historical accounts: as mentioned in Section 4.1 the majority of pupils were in Catholic schools in the 1950s. Yet, given that proximity to former Protestant missions has today the same effect as proximity to Catholic ones, our results suggest that the mechanisms behind persistence may differ across denominations.

Second, we confirm that the presence of nuns in Catholic missions was important for female education: halving the distance to a Catholic mission with at least one nun is associated with an 8 percentage point increase in the probability of having any schooling (column 5). This latter result represents a large impact compared to the mean of the variable for the female sample (16%). These results are particularly striking given that our measure of exposure is an average exposure within a given territory. As such, it tends to attenuate the true effect of mission exposure and provides lower bound estimates.[25] These results are consistent with the historical record on the functioning of the Catholic school system, where the presence of nuns was a necessary condition for opening schools for girls, and suggests that ignoring this distinction in similar contexts may be masking important heterogeneity.

Threats to identification.–We now briefly discuss (and refer to detailed analyses in appendix) two important threats to our identification of the impact of mission on education outcomes during colonization: selective migration and the endogenous opening of missions.

Selective migration.- One potential concern is related to the nature of the 1970s sample composed of people living in seven major cities, most of whom are migrants from

[25]This interpretation of the coefficients (in which we interpret only the interaction coefficient) is also conservative. It assumes that nuns' missions were the farthest away from the simulated points, and remain the farthest away among Catholic posts, so that the average distance to "general" missions remains unchanged. If we assume that nun missions were the closest missions to the simulated points (and therefore remain the closest missions), then the distance to "general" Catholic missions is also halved. Summing up the coefficients, we obtain an increase of 13 points.

Table 2: Education in colonial times and missionary presence

	(1) Years educ	(2) Any educ	(3) Primary or more	(4) Years educ	(5) Any educ	(6) Primary or more
Catholic missions	0.587***	0.0344	0.0546**	0.759***	0.0907*	0.0831***
	(0.209)	(0.0278)	(0.0251)	(0.248)	(0.0460)	(0.0318)
Catholic missions x female	0.0415	-0.0158	-0.00337	-0.416	-0.146	-0.0699
	(0.255)	(0.0379)	(0.0312)	(0.393)	(0.0909)	(0.0540)
Protestant missions	-0.0690	0.0364	0.0161	-0.0963	0.0302	0.0129
	(0.275)	(0.0299)	(0.0315)	(0.280)	(0.0301)	(0.0318)
Protestant missions x female	-0.159	0.0185	-0.0100	-0.115	0.0310	-0.00367
	(0.240)	(0.0285)	(0.0310)	(0.245)	(0.0287)	(0.0320)
Catholic with nuns				0.0405	-0.0290	-0.0139
				(0.228)	(0.0303)	(0.0278)
Catholic with nuns x female				0.423	0.121*	0.0615
				(0.320)	(0.0635)	(0.0399)
FEs	Yes	Yes	Yes	Yes	Yes	Yes
Territory FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	4.995	0.734	0.435	4.995	0.734	0.435
R-squared	0.442	0.292	0.338	0.442	0.293	0.338
N	41655	41655	41655	41655	41655	41655

Note. Data: Demographic Survey of the 1970s. The sample is restricted to people born after 1930. The table reports OLS estimates. Exposure to missionary presence is measured as $-log(distance)$. Outcome variables are defined as follows: single years of education in columns (1) and (4), a binary variable that equals one if the respondent has at least one year of education in columns (2) and (5) and the probability of completing primary education (>5 years of education) in columns (3) and (6). Standard errors () are clustered at the territory of birth level. Fixed-effects include city of residence, territory of birth, 5yr age cohorts and whether the individual was born in an urban area. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

rural areas. Therefore, selection into migration may bias our estimates.[26] To examine this question, we use data from a demographic survey conducted in the 1950s, at a time when migratory flows were strictly controlled (see Appendix A.3), to confirm our results on this sample and to quantify the extent of selective migration. The empirical strategies and results, presented in the appendix D.2, confirm that our estimated effects of missionary exposure on (migrant) education are not driven by a composition effect (where the effect of education on the propensity to migrate would be a function of exposure). Thus, we interpret these effects as the direct influence of missions on the educational level of the local population.

Endogenous opening of missions.–Our main identifying assumption is that in the absence of new mission posts, education trends would have been similar across places where these posts opened and where they did not. Given that missionaries were the sole providers of education during colonial times, this assumption is quite plausible.

[26]First, the returns to education may differ between Protestant and Catholic schooling. Second, the presence of mission posts may have created economic opportunities even for people who did not attend Christian schools, affecting their likelihood of migrating.

Yet, a potential concern in our setting is related to the endogenous timing of new post openings: we might be worried that the intensity of educational investments varied with different (time varying) returns to education (and that these returns were heterogeneous along our key dimensions of analysis: gender and religious denominations).[27] A first remark is that time-invariant factors that would affect these returns are absorbed by the territory fixed effects. We conduct two other analyses that suggest that an endogenous timing in school openings in unlikely to drive our results. First, in Table D3 in Appendix D.3.1, we check decade by decade that the characteristics of territories where Catholic vs Protestant missions settled are not systematically different.[28] Second, we add time variant controls that capture differences in the economic dynamism of different places across time and our results barely change (Appendix D.3.2).[29]

5. School clustering today and the persistence of education supply: the role of religious competition

Why does the initial location of colonial schools still influence education outcomes more than 80 years after their openings? Why do we observe today a positive effect only on women's years of education? Why do we observe a similar effect for Protestant and Catholic missions in the long run, while the effect for Catholics was significantly stronger during the colonial period? To answer these questions, we now turn to transmission mechanisms, distinguishing between two broad categories: supply factors and demand factors. We start by investigating the patterns of school location over time.

5.1. Patterns of school clustering

We first examine whether the supply of contemporary schools is correlated with the location of colonial mission posts. To do this, we use data on the universe of contempo-

[27]For instance, if, in contrast with Protestant schools, Catholic schools opened strategically in places where other colonial investments were made (e.g increased employment opportunities), our coefficients would not capture the causal effect of mission opening on education but rather the differential religious strategies deployed.

[28]For readability purposes, we chose to show only three decades.

[29]To do so, we collected information at the district level from yearly reports on the administration of the colony presented to the Belgium parliament. Specifically, we include data on the size of the white Belgian and white non-Belgium population, the size of the indigenous population and the amount of per-capita tax levied from the indigenous population. This information covers the entire period of analysis (1930-1948).

rary primary and secondary schools (as of 2020) and their characteristics provided by the Ministry of Education. First, we divide the DRC into small grid cells of 20 km x 20 km. Then, we calculate the total number of schools and missions per grid cell (see Figure E1 in Appendix E.1 for an example). Since very few grid cells have more than one mission, we rely on a binary indicator of the presence of a mission. Thus, the unit of observation for this analysis is a cell. In each specification, we include *collectivité* fixed effects, which are smaller administrative units than territories, to control for time-invariant location-specific confounders.[30] In addition, we control for grid-cell level factors that may have influenced the location of missions *within collectivités*, including: the presence of a navigable river, the presence of a colonial railroad, the presence of a pre-colonial exploration route, population density in 1900, cropland area in 1900, mean altitude, ruggedness and the distance from the centroid of each grid cell to Kinshasa.[31]

A strong correlation between historical missions and the number of schools may simply reflect that the grid-cells that hosted missions are (and likely were) more populated. To absorb this confounding factor we express our main dependent variable as the number of schools *per thousand inhabitants*[32] and we systematically control for the population density in 1900 (our results also hold even when we control for population density in 2020 at the cell level, and when we restrict the sample to rural schools, for more details see Appendix E.2).[34]

The results are reported in Table 3. We find a strong and significant effect of past

[30]Today, the national territory of the DRC is divided into 26 provinces, the province into territories/cities, the territory/city into *collectivités*, the *collectivité* into groups/neighborhoods, and the group/neighborhood into villages/streets. The *collectivité* is the smallest administrative unit for which we have the shapefile.

[31]Some variables used in Section 4 to control for mission placements are not included here because they are absorbed by the *collectivité* fixed effects. Indeed malaria suitability, tse-tse suitability, or the number of slaves exported in the Indian and Atlantic trades, are defined at the level of ancestral ethnic territory and are therefore (almost) invariant within a *collectivité* (while there are about 60 ethnic groups in the Ethnographic Atlas within the boundaries of the DRC, we count more than 700 *collectivités*).

[32]We measure total population at the grid-cell level in 2020.[33] To do this, we fist compute the average population density in the grid-cell and then multiply it by the cell's area. We obtain estimates that, at the country level, are close to recent estimations. For example, while recent estimates suggest that the total population in the DRC in 2010 was about 69 millions, our estimates give 72M. In 1984, year of the only scientific census ever conducted in the DRC, the population was recorded at about 30.7M, while our method gives for the year 1980 an estimate of 28.5M. We only keep grid-cells with at least 100 estimated total population in its 20x20km area to remove extreme outliers due to inaccuracies in population density estimates.

[34]Furthermore, the results are robust to using the number of missions per 1000 inhabitants as an explanatory variable.

Table 3: Historical Missions and Contemporary Schools

	Panel A: Primary Education		Panel B: Secondary Education	
	Schools per 1000 people	Proportion of Girls	Schools per 1000 people	Proportion of Girls
	(1)	(2)	(3)	(4)
Catholic mission	0.831***	0.0613***	0.495***	0.0691***
	(0.155)	(0.0101)	(0.159)	(0.0994)
Protestant mission	0.399***	0.0401***	0.254***	0.0492***
	(0.138)	(0.0129)	(0.0911)	(0.0929)
Mean Y	0.701	0.288	0.371	0.189
Grid-cell controls	Yes	Yes	Yes	Yes
Collectivite FE	Yes	Yes	Yes	Yes
R-squared	0.224	0.390	0.220	0.433
N	5786	6010	5786	5984

Note. Data: Universe of primary and secondary schools in the DRC in 2020, from Ministry of education. The table reports OLS estimates. The unit of observation is a 20kmx20km grid cell. Outcome variables are defined as follows: primary (secondary) schools per 1000 population in 2020 in column 1 (column 3) and average share of girls in primary (secondary) school by grid cell in column 2 (column 4). Controls at the grid-cell level are always included: presence of navigable river, presence of colonial railroad, presence of precolonial explorer route, population density in 1900, cropland area in 1900, mean altitude, ruggedness, and distance from the centroid of the grid-cell to Kinshasa. Sector fixed-effects (collectivités) included in all specifications. The number of observations is lower is column 1 because grid-cells with very low population are excluded to obtain meaningful values. Robust standard errors in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

missionary presence on the number of primary and secondary schools a given cell has today. When a historical Catholic (Protestant) mission was present in the grid cell, there is 0.8 (0.4) additional primary schools per 1000 inhabitants today (Column 1). The same holds true for secondary schools: a Catholic (Protestant) mission in the grid cell is associated with 0.5 (0.25) additional secondary schools per 1000 inhabitants. These effects are large. Having hosted a Catholic (Protestant) mission in the past is associated with a 119% (57%) increase in the mean number of primary schools per 1000 inhabitants. Similarly, having hosted a Catholic (Protestant) mission in the past is associated with a 133% (68%) increase in the mean number of secondary schools per 1000 inhabitants. Interestingly, the proportion of girls in these secondary schools is also higher near former missions (we return to this point below).

5.2. School clustering is not fading!

Table 4 reports the correlation between the presence of a mission (within a grid cell) and the stock of secondary schools per 1,000 inhabitants by decade, using the same control variables as in Table 3.[35] The results suggest that the influence of colonial missions

[35]The drop in the number of grid cells in the last two columns is due to the different data sources for population density. While historical population density estimates come from the HYDE 3.1 database,

on school density has strengthened over time. Notably, the presence of a Protestant mission only becomes significantly correlated with secondary school density starting in the 1980s. Additionally, the coefficients for both "Catholic mission" and "Protestant mission" increase over time. Even relative to the average school density (reported in the table's lower panel), the effect of missions has not be fading over the last four decades (in a grid-cell where a Catholic mission operated at independence, there are about twice as many secondary schools per 1000 inhabitants today than in an average grid-cell). This is further supported by the rising R-squared values over time: while dependent variables explained 12% of the variation between grid cells in 1970, they account for 22% today.

The stock of secondary schools by decade is derived from the recorded opening dates of schools active in 2020. As a result, schools that have since closed are not included. One concern is that school closures may be more frequent in areas farther from missions, potentially leading to an overestimation of the correlation between missions and school density. However, two elements mitigate this concern. First, if school closures were more common in remote areas, we would expect to see weaker clustering effects in the recent past - when we are mostly observing newer schools that have not yet closed - but the data show the opposite. Second, school closures typically indicate disruptions in the supply of education and worse learning outcomes. Thus, the stock of "surviving" schools remains a meaningful measure of continuous access to schooling. Finally, our results are robust to removing grid cells with a significant amount of conflict, which could be correlated with school closures (see Appendix E.3).

---

recent estimates (2010 and 2020) come from the NASAs Gridded Population of the World (GPW) collection.

Table 4: Former missionary presence and supply of secondary education, by decade

	(1) N 1960/1000	(2) N 1970/1000	(3) N 1980/1000	(4) N 1990/1000	(5) N 2000/1000	(6) N 2010/1000	(7) N 2020/1000
Catholic mission	0.0637***	0.0815***	0.150***	0.157***	0.215***	0.471***	0.495***
	(0.0211)	(0.0250)	(0.0285)	(0.0313)	(0.0585)	(0.0874)	(0.0994)
Protestant mission	0.0240	0.0577	0.129***	0.138***	0.175**	0.197***	0.254***
	(0.0298)	(0.0361)	(0.0408)	(0.0464)	(0.0841)	(0.0722)	(0.0911)
Mean Y	0.0174	0.0218	0.0677	0.0890	0.173	0.317	0.371
Collectivite FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grid-cell controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.116	0.118	0.138	0.166	0.186	0.211	0.220
N	5852	5872	5891	5916	5935	5780	5786

NOTE. Data: Universe of secondary schools in the DRC in 2020, from Ministry of education. The table reports OLS estimates. The unit of observation is a 20kmx20km grid cell. Outcome variables are defined as follows: number of schools per 1000 inhabitants in 1960 in column 1, number of schools per 1000 inhabitants in 1970 in column 2, number of schools per 1000 inhabitants in 1980 in column 3, number of schools per 1000 inhabitants in 1990 in column 4, number of schools per 1000 inhabitants in 2000 in column 5, number of schools per 1000 inhabitants in 2010 in column 6, and number of schools per 1000 inhabitants in 2020 in column 7. Controls at the grid-cell level are always included: presence of navigable river, presence of colonial railroad, presence of precolonial explorer route, population density in 1900, cropland area in 1900, mean altitude, ruggedness, and distance from the centroid of the grid-cell to Kinshasa. Sector fixed-effects (collectivités) included in all specifications. Robust standard errors in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

Why do post-independence schools open near former missions (and schools)? The existing literature suggests two main mechanisms: 1) missions were located in areas where demand/returns to education would be higher (Jedwab et al., 2017); and 2) missions triggered structural change leading to higher demand/returns to education (Rocha et al., 2017; Valencia Caicedo, 2019a). We explore here a third alternative, namely that religious schools compete and choose to locate next to each other *à la Hotelling*. In what follows, we first examine school competition and its role in school clustering. We then turn to an examination of mission placement and structural transformation arguments, looking at the correlation between the structure of occupations and proximity to historic missions.

5.3. The dynamics of schools expansion and religious competition

Competition among education providers may lead to the clustering of schools in areas where they first emerge, as suggested by Hotelling-style models of spatial competition. However, other models predict dispersion rather than clustering (for a review of the forces driving clustering, see Fujita and Thisse (1996)), and it is not immediately clear that competing schools would choose to locate in the same areas. Nonetheless, strategic school placement driven by competition is highly likely in contexts where entry into the education sector is relatively easy and schools are often affiliated with religious organi-

zations. In such cases, religious competition may strongly influence location decisions.[36]

To investigate clustering effects and competition among schools, we use information on the dynamics of school openings and examine whether the opening of a secondary school of a given denomination triggers - or deters - the opening of a school of the same or another denomination in the same place (grid-cell).[37]. We are agnostic as to whether competition triggers clustering or dispersion, yet we expect competition to be stronger between schools of *different* denominations than between schools of the *same* denomination, leading to systematic differences in the estimated coefficient on schools of the same versus other denominations.

More specifically, we estimate the following regression, where $S_{i,t,r}$ takes value 1 if a school of denomination r opened at time t in grid-cell i (with $r = c$ corresponding to Catholic and $r = p$ to Protestant school). Finally, λ_t and σ_i are year and grid-cell fixed effects, respectively. We show the results using different lags for the opening of a school in the past (during the previous three or five years). Note that our explanatory variables include a lag of our dependent variable. However, as T is large in our context (the median grid-cell has $T = 40$), dynamic panel bias becomes insignificant and the standard within-groups estimator is appropriate (Roodman, 2009; Blundell and Bond, 2023).[38]

$$S_{i,t,r} = \beta_0 + \sum_{j=1}^{3,5} \beta_c S_{i,t-j,c} + \sum_{j=1}^{3,5} \beta_p S_{i,t-j,p} + \lambda_t + \sigma_i + \varepsilon_{i,t,r} \tag{1}$$

To the extent that competition is stronger across religious denominations than within, competition would imply systematic differences between β_c and β_p. In particular with a Hotelling-like model, we expect that the opening of a Catholic (Protestant) school is more likely after the recent opening of a Protestant (Catholic) school than of a Catholic (Protestant) school. If school openings are unrelated to religious competition, we would

[36]In several contexts, religious competition has been shown to induce churches to provide services (such as education) to attract or retain converts (Iyer, 2016).

[37]We focus on secondary schools that opened after 1949, the last year for which we have information on the arrival of missionaries

[38]Moreover, the number of instruments would be far too large if we implemented the kind of estimators proposed by Arellano and Bond (1991). For example, if we include three yearly lags of our dependent variable, we would have about 3000 instruments. We would still reach 150 instruments if we limit ourselves to a single lag. In any case, our main conclusions remain valid if we implement the estimator proposed by Arellano and Bond (1991) (see Appendix E.4).

Table 5: Competition between religious affiliations

	(1) Catholic school	(2) Protestant school	(3) Catholic school	(4) Protestant school
Catholic between t and t-3	-0.0317***	0.0215***		
	(0.00328)	(0.00443)		
Protestant between t and t-3	0.0108***	-0.00843**		
	(0.00267)	(0.00406)		
Catholic between t and t-5			-0.0366***	0.0172***
			(0.00289)	(0.00388)
Protestant between t and t-5			0.00690***	-0.0218***
			(0.00245)	(0.00360)
Time FEs	Yes	Yes	Yes	Yes
Grid-cell FEs	Yes	Yes	Yes	Yes
Mean Y	0.0581	0.120	0.0581	0.120
Catholic=Protestant (p-val)	0.000	0.000	0.000	0.000
R-squared	0.0948	0.167	0.0959	0.168
N	72132	72132	72132	72132

NOTE. The dependent variable is a dummy that equals one if a new school of type x opens in cell c at year t. The independent variables are dummies that equal one if a new school of the opposite type opened in cell c in year t-1, t-2 or t-3, or in cell c in year t-1, t-2, t-3, t-4 or t-5, respectively. Standard errors () clustered at the 20x20km grid cell level are reported in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

not expect any difference between β_c and β_p.

We estimate this model for the whole period (Table 5) and by decade (Figure 1). Table 5 reveals that both types of religious schools respond very differently to the opening of a school of the other denomination than to the opening of a school of the same denomination. Column 1 shows that the opening of a Catholic school in the same area in the last 3 years reduces the probability of a new Catholic school opening by about 3 percentage points. In contrast, the opening of a Protestant school in the last 3 years increases the probability of a new Catholic school opening by 1.1 percentage points, which is an increase of about 20% from the mean. Protestant schools also respond to the opening of a Catholic school: the opening of a Catholic school in the last three years increases the likelihood of a Protestant school opening by 2.2 percentage points (or about 20% of the mean) (column 2). These results are consistent with competitive forces (stronger across than within denomination) shaping school clustering in specific places.

Figure 1 plots the estimated coefficients by decades, with Catholic schools (panel 1a) or Protestant schools (panel 1b) as the dependent variable. Each plotted coefficient is the sum of β_c or β_p and its interaction with decade. The standard errors almost never overlap, suggesting that, since independence, religious schools react differently to the open-

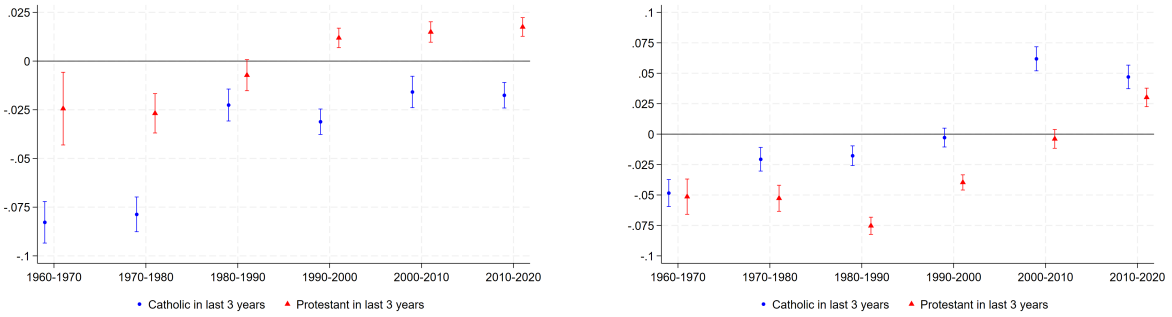
ing of a school of the other denomination than of their own (in-line with competition-driven positioning). Yet the competition-driven *clustering* appears to be relatively recent: before the 1990s, new schools avoided places where a competitor just opened a school (this dispersion effect was however weaker than for a school of the same denomination).

These *differential* effects by denomination after the opening of a new school are difficult to explain in the absence of competitive forces, even if clustering more generally may also be driven by economic factors. To account for the later possibility, we show in Appendix E.5, Table E7, that results remain unchanged when we control for education trends at the province level (capturing aggregate trends in the demand of education for example). Furthermore, this competition is not a purely urban phenomenon that would accompany a processes of structural transformation: if we restrict attention to rural areas, results are similar (results are displayed in Appendix E.6, using two different definitions of rural grid-cells). Moreover, we show that our results are not sensitive to the use of Conley standard errors, and therefore are not driven by the existence of spatial autocorrelation (see Table E8).[39] Finally, we conduct a robustness test where we restrict attention to grid cells with low exposure to conflicts where school closures have been less likely.[40] Appendix E.3 confirms that the results presented in sections 5.1 and 5.3 are robust to this sample restriction (Tables E4 and E5, respectively).

[39] A plausible interpretation of the negative coefficients for the opening of a school of the same denomination is that the capacity of the existing school is increased instead of building a new school very close to the existing one. However, this mechanism does not explain why Protestant schools are more likely to open where a new Catholic school has just opened (and vice versa) than elsewhere.

[40] We use data on the number of conflicts that ocurred from 1997-2020 from the Armed Conflict Location and Event Data Project (ACLED). Conflicts include events such as fighting, explosions or remote violence, protests, riots, strategic developments, and violence against civilians.

Figure 1: Competition dynamics of secondary schools by decade



(a) Probability of opening a Catholic school (b) Probability of opening a Catholic school

Notes: Authors' calculations using data on the universe of secondary schools in the DRC provided by the Congolese Ministry of Education. The Figure shows the plotted coefficients of estimating equation 1 interacted with decade. Each plotted coefficient is the sum of of β_c or β_p and its interaction with decade. Bars around point estimates represent standard errors, clustered at the grid-cell level.

In short, religious competition appears to have lead to a concentration of schools in areas close to historic missions in recent years. While this possibility has been raised in the literature, it has never been formally tested.[41] Historians insist on the importance of religious competition in colonial Africa and argue that it helps to understand the geography of mission settlements. In the case of the Congo, both Catholic and Protestant authorities carefully monitored each other's progress and tried to occupy as many "virgin" territories as possible while preventing the other religion from "conquering" large areas. Recent reports on education in the DRC indicate that today, as in colonial times, schools remain a vector of proselytization for the churches, so that church competition can lead to school competition (Gauthier et al., 2021).[42]

5.4. Alternative supply-side mechanisms

In this section we review two common explanations for the persistence in school supply: the endogeneity of mission placement and structural transformation.

If missions located in more favorable places (better connected to markets, with bet-

[41] Interestingly, in their investigation of the impacts of colonial public education on labor and marriage market outcomes in Cameroon, André and Dupraz (2023) note that Christian schools tend to locate where other Christian schools were already operating. While they do not distinguish between Protestant and Catholic schools, this finding could be the result of competition across schools.

[42] Although studying the impact of school clustering and competition on the quality of the education provided is beyond the scope of this paper, simple descriptive statistics show that Catholic schools are slightly larger (in number of students and classrooms) than Protestant ones, but have comparable proxy indicators of quality: same number of teachers per classroom, students per classroom and per teacher (Tables upon request).

ter natural endowments, etc.) where returns to education were higher, these places (in particular urban areas) may continue to enjoy higher returns and attract more schools, simply because they are better endowed (and not because they had missions).[43] The endogenous placement of missions is unlikely to drive our results on clustering for several reasons. First, the above analysis of clustering includes *collectivité* fixed effects to remove any time-invariant heterogeneity at a granular level. Second, it also includes grid-cell level state-of-the-art controls to account for the potential endogeneity in the location of missions within *collectivités*.[44] Furthermore, we show in Sections 4 and C.2 that our results on persistence are robust to *(i)* the inclusion of the determinants of mission location, *(ii)* the inclusion of historical and current population density as controls, *(iii)* a restriction to rural places. A related argument is that places that attracted colonial investment in education may continue to attract more public funding due to the permanent nature of school buildings. This mechanism is highlighted by Huillery (2009) in the context of French Africa. In the DRC, where public subsidies are dismal, we do not expect an important contribution from this channel.

Another explanation for the concentration of schools closer to historic missions relates to structural change and returns to education. If the initial educational advantage allowed mission-exposed areas to take advantage of economic opportunities (development of manufacturing or service sectors), then these areas may have enjoyed higher levels of returns to education and economic development, leading to a more dynamic supply of education and generating educational persistence.[45]. We already argued above that our results on competition are unlikely to be explained by structural transformation (they are not restricted to urban areas and they hold when controlling for province-level trends in education). Here we provide suggestive evidence based on contemporaneous occupational patterns in the DHS data, and on the intensity of nightlights in the area around each DHS cluster (as in Jedwab et al. (2022)). While DHS surveys do not include details about occupation or income, they provide basic information about labour mar-

[43]This issue is more problematic when mission data are limited to the oldest missions or when mission data come from atlases, (Jedwab et al., 2017), which is not our case (we have information on the universe of missions operating in 1948 from several yearbooks and colonial maps).

[44]There are more than 700 collectivités in the country.

[45]Occupational specialization is the main factor of transmission generating economic and educational persistence in Rocha et al. (2017); Valencia Caicedo (2019a)

ket participation and the broad sector of occupation. Nightlight is used as a proxy for economic activity in the area.[46]

In practice, we re-estimate the same equation as in Section 4.2 long-term effects on education using DHS data) using these outcomes as dependent variables. The results are shown in Table E11, Appendix E.7. They reveal that halving the distance to former Catholic or Protestant missions *reduces* the propensity to work in the formal sector by 1-2 percentage points (even when considering only individuals over 20 in column 2). These results speak against faster structural transformation closer to mission, since higher-productivity activities tend to concentrate in the formal sector (McMillan et al., 2014). Yet, conditional on working, the sector of employment is slightly different in the vicinity of Catholic missions: halving the distance reduces the probability of working in agriculture by four percentage points in favor of services (+2 pp) and manufacturing (+1.2 pp). In contrast, we see no clear pattern of occupation closer to historic Protestant missions. Finally, we find no effect of nighttime luminosity, suggesting that places closer to historic missions are not more economically dynamic, in line with the results for Ghana of Jedwab et al. (2022). In short, we do not find clear evidence of structural transformation near historical missions. At most, the occupational structure shows some modest signs of structural transformation near historical Catholic posts (but not Protestant), suggesting that these differences in occupational specialization (and associated structural transformation) are unlikely to explain the patterns of persistence.

6. Persistence in demand for education: empirical investigation

Demand for education may be persistently higher near former missions if educated parents have a higher demand for their children's education than uneducated parents, but also if missions directly affect parental demand for education, *regardless of parental education*. The latter mechanism would lead to higher educational mobility in the vicinity of former missions, with uneducated parents more likely to invest in their children's education if they were exposed to missions (Alesina et al., 2021). Such effects have been highlighted by Wantchekon et al. (2015). The former mechanism (whereby educated

[46]This variable is provided by the DHS. It is constructed as the average nighttime luminosity of the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location. It removes flares from petroleum extraction and other short-duration lights and is corrected for moonlight.

parents have a higher demand for education), termed the intergenerational *transmission* of education, has been the focus of a large literature in sociology and economics and may be driven by several factors.[47] First, there may be an intergenerational correlation in ability or other family characteristics (such as values, parenting skills or information). Second, parental education may affect the (perceived) returns to schooling. Finally, in the presence of liquidity constraints, uneducated parents may be poorer and therefore under-invest in their children's education.

An important confounding factor in trying to isolate these demand channels relates to supply: the educational levels of parents and children may be correlated simply because they are more likely to grow up in the same environment (with the same access to schools) than random individuals. This supply channel is central to Alesina et al. (2021)'s recent study of intergenerational mobility in sub-Saharan Africa, where they compare intergenerational transmission among siblings who grew up in different places.[48] Our historical data provide an ideal setting to isolate pure demand factors from this supply channel. Indeed, it consists of a representative sample of second-generation urban migrants in the 1970s who share the same supply environment but whose parents, educated in missions, have different levels of education (and exposure to missions). This allows us, in the spirit of Alesina et al. (2011)'s earlier work on the persistence of gender roles, to compare the educational outcomes of children who share the same access to school but whose parents had different levels of exposure to missions in their childhood (and different levels of education).

6.1. Intergenerational transmission of education

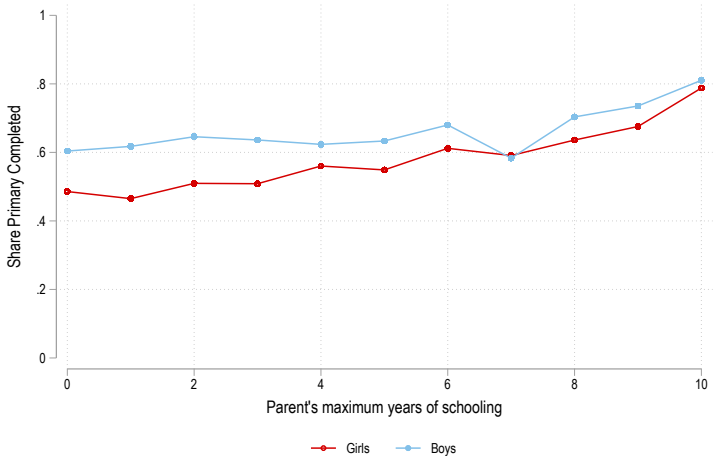
We begin by estimating the raw correlation between parents' and children's education. Figure 2 plots the share of children with completed primary education as a function of the maximum year of education of their parents. There is a positive relationship, and the slope is steeper for girls than for boys: while 60% (50%) of boys (girls) whose

[47]Seminal papers include Becker (1964) and Coleman (1968). For a recent review, see Bjorklund and Salvanes (2011).

[48]They find that "regional exposure" has a significant impact on the level of intergenerational correlation (or mobility). The classification of this "regional exposure" effect as a supply-side effect is debatable in that it captures not only school supply but also local returns to education. However, it differs from a "pure" demand channel, where educated parents would have a higher demand for education, holding the environment (for example, school suply or the local returns to education) constant.

parents have no education manage to complete primary education, this percentage rises to 80% when the maximum year of parental education is 10. This suggests that there is positive intergenerational correlation, yet the graph also reveals a *remarkable degree of upward educational mobility* immediately after independence: a majority of urban children (boys and girls) of uneducated parents finish primary school.[49]

Figure 2: Correlation between parents' and children education



Note: Data: Demographic Survey of the 1970s. The figure shows the share of children aged 14-19 with primary school completed (6 years of education) given the maximum year of education of their parents.

We show in Appendix F.1 that the positive, but very modest, correlation between children and parents education is confirmed in a regression framework, controlling for an extensive set of demographic and geographic characteristics.[50] Our estimates also suggest that the relationship between parents and children education is slightly stronger for girls, suggesting that parental education is a stronger determinant of girls' education than of boys'.[51] We come back to this point in the discussion section below.

[49]We follow Card et al. (2022) or Alesina et al. (2021) and focus on children aged 14-19 because they should have already completed primary school and cohabitation rates are very high (0.6% of people aged 14-19 in our sample are heads of household, while about 6% are spouses of the head of household).

[50]We obtain consistent results if we replace parents' years of education with a dummy indicating whether parents have any education. Table upon request.

[51]This is consistent with recent findings for other developing countries, both in Africa or in Asia (Emran and Sun, 2015; Emran and Shilpi, 2015; Azomahou and Yitbarek, 2016; Torche, 2019).

6.2. *Missions and education mobility*

The high degree of educational mobility shown above may itself be influenced by exposure to missions (maybe missions changes the demand for education of both educated and uneducated individuals). As mentioned above, Alesina et al. (2021) show that upward mobility in Africa is higher in areas exposed to missions. With our urban migrant sample, we can formally test for this positive effect of missions on educational mobility, again holding supply condition constant.

We exploit the variation in parental exposure to missions. Thus, our coefficients are again identified by time variations in parental exposure to missions in a given territory, holding constant the supply of education in the city. In addition, with the same estimation strategy, we explore the role of exposure to mission on education *expenditure*, taking advantage of an expenditure survey including 1/50 of the total number of households identified in the same seven cities of the 1970s Demographic Survey (see Appendix E.7 for more details).

Table 6: Parents' exposure to missions and children's education

	(1)	(2)	(3)	(4)	(5)	(6)
	Years educ	$\geq$ Primary	Post-primary	Years educ	$\geq$ Primary	Post-primary
Father's exposure to Cath	-0.0264	-0.0143	0.00179	0.0167	-0.0135	0.00228
	(0.116)	(0.0228)	(0.0196)	(0.158)	(0.0270)	(0.0225)
Father's exposure to Nuns	-0.118*	-0.0201	-0.0115	-0.133*	-0.00924	-0.00795
	(0.0643)	(0.0142)	(0.0119)	(0.0753)	(0.0164)	(0.0133)
Father's exposure to Prot	0.277***	0.0433**	0.0397**	0.193*	0.0281	0.0336*
	(0.102)	(0.0207)	(0.0185)	(0.116)	(0.0223)	(0.0198)
Exposure to Cath x Female				-0.0910	-0.00246	-0.00133
				(0.152)	(0.0196)	(0.0180)
Exposure to Nuns x Female				0.0232	-0.0229	-0.00764
				(0.0797)	(0.0154)	(0.0145)
Exposure to Prot x Female				0.179	0.0324**	0.0130
				(0.111)	(0.0136)	(0.0131)
FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	6.066	0.632	0.437	6.066	0.632	0.437
R-squared	0.284	0.201	0.253	0.285	0.202	0.253
N	20443	20443	20443	20443	20443	20443

NOTE. Data: Demographic Survey of the 1970s. The sample is restricted to children aged 14-19 at the time of the survey. The table reports OLS estimates. Father's exposure to missions is described Section 3.2. Outcome variables are defined as follows: single years of education in columns (1) and (4), the probability of completing primary education (>5 years of education) in columns (2) and (5), and the probability of having post-primary education (>6 years of education) in columns (3) and (6). Standard errors () are clustered at father's territory of birth level. Fixed-effects include city of residence and father's territory of birth. Controls include sex, year of installation in the current city, age and age of parents. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

Table 6 shows the results and reveals that parental exposure to Protestant missions

directly increases their demand for education, especially for their daughters. In contrast, exposure to Catholic missions has no significant effect. In terms of magnitude, these effects are modest. Halving the distance (at birth) of the father to a Protestant mission increases children's years of education by 3% or the probability of completing primary education by 3 pp (5%). The same conclusion is drawn from the analysis of educational expenditures (see Table F2 in Appendix F.2): exposure to Protestant, but not Catholic missions, increases educational expenditures.

These results come somewhat as a surprise, since we showed above that *Catholic but not Protestant* missions had a sizable effect on the *education of the first generation*. Taken together, these findings suggest that parental exposure to Catholic missions has an effect on children education *only to the extent that parents themselves are educated* (so that the results is far less strong in reduced form when we directly estimate the effects of missions on the education of the next generation). In contrast, exposure to Protestant missions appears to directly increase parents' demand for education, suggesting that exposure to Protestant missions is associated with more investment in education, even for uneducated parents (exposure to Protestant missions has a stronger impact on mobility). In Appendix F.3 we explore this conjecture and confirm that Protestant missions have a stronger effect on mobility than Catholic missions. We come back to this contrast across religion in the next section.

Overall, our examination of demand channels suggests that the influence of parental education and mission exposure on educational outcomes in urban post-independence Congo is modest. Parents, whether educated or not, invested heavily in the education of their children, the majority of whom completed primary school. However, interesting gender and religious contrasts emerge, with a positive correlation between children's education and parental exposure to Protestantism, and greater intergenerational persistence for daughters.

7. Concluding remarks and reflections on gender and religious disparities

We begin by summarizing our main findings. Our analysis reveals a persistent correlation between the historical presence of Catholic and Protestant missions and contemporary educational outcomes, particularly for women. This correlation has not diminished over time - if anything, it has strengthened. While we do not find differences in educa-

tional attainment today based on the religion of the mission, we show that Catholic missions had a stronger initial impact on the education of earlier generations than Protestant missions. This suggests that the mechanisms driving persistence may differ between the two denominations.

Our examination of these mechanisms suggests that supply-side factors, rather than demand-side factors, play a fundamental role in shaping educational inequality. Specifically, regions closer to historical missions (regardless of denomination) now have a higher concentration of primary and secondary schools. A key driver of this clustering is religious competition: schools of different religious affiliations tend to be located close to each other, with new schools often emerging in response to those of rival denominations. Meanwhile, our examination of demand-side factors suggests that their overall influence is limited, although interesting gender and religious contrasts emerge. In particular, there is a stronger intergenerational correlation for daughters and a positive association between parental exposure to Protestant schools and children's education.

Why is the persistence of spatial inequalities stronger for women?– Although school supply might appear gender-neutral (since most schools are mixed) the spatial distribution of schools has a disproportionate impact on female education. Distance to school is a well-documented barrier to enrollment, particularly for girls (King and Hill, 1993; Burde and Linden, 2013; World Bank, 2021). Therefore, persistent school clustering may be a critical factor in explaining spatial inequalities in female educational outcomes.

Data from the 2013 National Survey on the Situation of Out-of-School Children and Adolescents in the DRC (OOSC) confirm the strong correlation between distance to schools and attendance rates. This survey is representative of children aged 6-17 and includes detailed information on education and on the distance between the respondent's residence and the nearest primary and secondary school for approximately 20,000 children.[52] We show in Appendix E.8, Table E12, that for secondary school attendance, gender differences emerge: girls living 35 km from a secondary school are 7.4 percentage points less likely to attend than those within 1 kmequivalent to nearly 30% of the mean attendance ratewhile for boys, distance has no discernible effect.[53]

[52]The survey was managed by the Ministry of Primary, Secondary and Vocational Education of the DRC and implemented by the Higher Institute of Population Sciences of the University of Ouagadougou.

[53]For primary schools, the effect is substantial and similar for boys and girls: children living 510 km

Regarding the demand for education, we highlight a stronger intergenerational correlation for girls. The literature identifies three primary explanations for intergenerational transmission of education: (1) liquidity constraints, (2) family characteristics and values, and (3) differences in returns to education based on parental background. Our regressions, presented in Appendix F.4, include parental occupation as a proxy for liquidity constraints. Strikingly, the coefficient on parental education remains virtually unchanged when controlling for this variable, suggesting that liquidity constraints play a limited role.[54]

Could differences in family values or returns to education explain the stronger intergenerational persistence for daughters? Several factors suggest that they might. Colonial-era schooling inculcated gendered values that likely shaped parental attitudes toward female education. In addition, labor market returns were historically higher for men because educated women had limited opportunities for skilled employment until the end of the colonial period. However, women may have experienced higher marriage market returns to education. If marriage market returns depend more on parental background than on labor market returns (for example, if family networks play a larger role in matchmaking than in employment), then the intergenerational transmission of education may be stronger for daughters than for sons.[55] In sum, both the strong spatial clustering of schools and the modest intergenerational persistence of education appear to have a greater impact on girls' education than on boys'.

Why do Protestant and Catholic missions have similar effects on education today but not in the past? The clustering of schools around missions provides a straightforward explanation: both Catholic and Protestant missions shaped the educational landscape in ways that continue to matter today.

Furthermore, on the demand side, we detect a greater intergenerational impact of

from a school are 19 percentage points less likely to have attended school than those within 1 km, and this figure rises to 48 percentage points for distances greater than 10 km. Note that the mean of the dependent variable is significantly higher in columns 1-3 than in columns 4-6. This is because columns 1-3 focus on the age range 6-11 and 25% of the children in our sample start primary school later, when they are older than six years old.

[54]While parental occupation is a "bad control" in that it is influenced by parental education itself, the stability of our education coefficient is reassuring.

[55]On intergenerational transmission of family characteristics, Wantchekon et al. (2015) argues that ability likely played only a minor role in colonial school selection, as missionary education was rarely merit-based.

Protestant missions compared to Catholic ones. This aligns with a large body of research indicating that Protestantism has historically fostered a stronger attachment to education. A core Protestant principle – sola scriptura – emphasizes that every individual, male or female, should be able to read the Bible. Studies by Becker and Woessmann (2008, 2009, 2010) demonstrate that Protestantism in early 19th-century Prussia led to a significant increase in literacy demand even before industrialization.

Determining whether Protestant missionaries in the Congo placed greater emphasis on literacy than their Catholic counterparts is challenging. However, two key differences stand out. First, Protestant missionaries were, on average, more educated than Catholic missionaries – particularly female missionaries (Irvine, 1978). Second, Protestant missions took a more gender-equitable approach to schooling (Yates, 1982). These factors may explain why exposure to Protestant missions correlates more strongly with contemporary educational attainment than it did in the past.

Broader implications.–The persistence of colonial-era schooling patterns in the DRC underscores the long-term impact of historical institutions on gender and regional disparities. Religious competition has shaped the geography of education, while distance to schools remains a critical determinant of educational inequality, particularly for girls. These findings have broader implications beyond the DRC.

The growing involvement of private actors in education, both in Africa and globally, highlights the need to better understand how competition influences the quantity and quality of education across different regions (Panin, 2021). Effective regulatory frameworks are essential to ensure that competition enhances, rather than exacerbates, inequalities in educational access. Religious institutions, which play a major role in education and public service provision, also raise important policy questions. Believers may be reluctant to switch providers, potentially becoming "captive" to religious schools even when quality is low. Furthermore, when religious leaders are directly involved in service provision, they may shape policy outcomes based on their own political and economic interests. Emerging research demonstrates that they can significantly influence program uptake, either facilitating or obstructing public initiatives (Vyborny, 2024; Stoop et al., 2019).

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1):1–35.

Administration coloniale (1950). Rapport sur l'administration du congo belge: Présentation aux chambres législatives.

Alesina, A., Giuliano, P., and Nunn, N. (2011). Fertility and the plough. *American Economic Review*, 101(3):499–503.

Alesina, A., Hohmann, S., Michalopoulos, S., and Papaioannou, E. (2021). Intergenerational Mobility in Africa. *Econometrica*, 89:1–35.

Alsan, M. (2015). The Effect of the TseTse Fly on African Development. *American Economic Review*, 105:382–410.

André, G. and Poncelet, M. (2013). Héritage colonial et appropriation du ń pouvoir déduquer. *Cahiers de la recherche sur léducation et les savoirs*.

André, P. and Dupraz, Y. (2023). Education and polygamy: Evidence from cameroon. *Journal of Development Economics*, 162:103068.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricists Companion*. Princeton: Princeton University Press.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277–297.

Arenas, A. and Hindriks, J. (2020). Intergenerational Mobility and Unequal School Opportunity. *The Economic Journal*, 131(635):1027–1050.

Azomahou, T. and Yitbarek, E. (2016). ntergenerational Education Mobility in Africa: Has Progress Been Inclusive?. *Policy Research Working Paper 7865. Washington, DC: World Bank.*

Baten, J., De Haas, M., Kempter, E., and Meier Zu Selhausen, F. (2021). Educational Gender Inequality in Sub-Saharan Africa: A Long-Term Perspective. *Population and Development Review*, 47:813–849.

Bau, N. (2022). Estimating an equilibrium model of horizontal competition in education. *Journal of Political Economy*, 130(7):1717–1764.

Beach, H. (1903). *A Geography and Atlas of Protestant Missions: Their Environment, Forces, Distribution, Methods, Problems, Results and Prospects at the Opening of the Twentieth Century*. New York: Student Volunteer Movement for Foreign Missions.

Becker, B. (2022). The empire within: Longitudinal evidence on the expansion of christian missions in colonial africa. *Journal of Historical Political Economy*, 2:333–362.

Becker, B. and Meier zu Selhausen, F. (2023). *Women on a Mission: Protestant Legacies of Gender Equality in Africa?* AEHN, African Economic History Network.

Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.

Becker, S. O. and Woessmann, L. (2008). Luther and the girls: Religious denomination and the female education gap in nineteenth-century prussia. *The Scandinavian Journal of Economics*, 110(4):777–805.

Becker, S. O. and Woessmann, L. (2009). Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *The Quarterly Journal of Economics*, 124:531–596.

Becker, S. O. and Woessmann, L. (2010). The effect of protestantism on education before the industrialization: Evidence from 1816 prussia. *Economics Letters*, 107(2):224–228.

Ben Salah, M., Chambru, C., and Fourati, M. (2022). The colonial legacy of education: Evidence from of tunisia. *Working Paper, No. 411. University of Zurich, Department of Economics*.

Bergeron, A. (2020). Religion and the Scope of Morality: Evidence From Exposure to Missions in the DRC. *Working Paper*.

Bjorklund, A. and Salvanes, K. G. (2011). Education and family background: Mechanisms and policies. In Hanushek, E., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 3, chapter 03, pages 201–247. Elsevier, 1 edition.

Black, S. E. and Devereux, P. J. (2011). Recent developments in intergenerational mobility. *Handbook of labor economics*, 4:1487–1541.

Blundell, R. and Bond, S. (2023). Initial conditions and blundellbond estimators. *Journal of Econometrics*, 234:101–110.

Boyle, P. (1995). School Wars: Church, State, and the Death of the Congo. *The Journal of Modern African Studies*, pages 451–468.

Briand, V. and Nicolai, M. (2021). Religious networks in the drc: help or hindrance to education? *World Bank report*, (161509).

Burde, D. and Linden, L. L. (2013). Bringing education to afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics*, 5(3):27–40.

Cagé, J. and Rueda, V. (2016). The Long-Term Effects of the Printing Press in Sub-Saharan Africa. *American Economic Journal: Applied Economics*, 8:69–99.

Cagé, J. and Rueda, V. (2020). Sex and the mission: The conflicting effects of early Christian missions on HIV in Sub-Saharan Africa. *Journal of Demographic Economics*, 86:213–257.

Calvi, R., Hoehn-Velasco, L., and Mantovanelli, F. G. (2022). The protestant legacy: missions, gender, and human capital in india. *Journal of Human Resources*, 57(6):1946–1980.

Calvi, R. and Mantovanelli, F. G. (2018). Long-term effects of access to health care: Medical missions in colonial India. *Journal of Development Economics*, 135:285–303.

Card, D., Domnisoru, C., and Taylor, L. (2022). The intergenerational transmission of human capital: Evidence from the golden age of upward mobility. *Journal of Labor Economics*, 40(S1):S39–S95.

Coleman, J. S. (1968). Equality of educational opportunity. *Equity & Excellence in Education*, 6(5):19–28.

Collier, P. (2024). *Left Behind. A New Economics for Neglected Places*. Penguin UK.

Congo Missionary Conference (1911). A report to the sixth general conference of missionaries of the protestant missionary societies working in congoland.

Congo Missionary Conference (1924). Organisation de lenseignement libre subsidié pour indigènes avec le concours des sociétés de missions chrétiennes.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92:1–45.

Corman, A. (1924). Annuaire des Missions Catholiques au Congo Belge. *Lédition universelle, Bruxelles*.

Corman, A. (1935). Annuaire des Missions Catholiques au Congo Belge. *Lédition universelle, Bruxelles*.

Depaepe, M. and Kikumbi, A. L. (2018). Educating girls in Congo: An unsolved pedagogical paradox since colonial times? *Policy Futures in Education*, 16:936952.

Emran, M. and Shilpi, F. (2015). Gender, Geography, and Generations: Intergenerational Educational Mobility in Post-Reform India. *World Development*, 72:36280.

Emran, S. and Sun, Y. (2015). Magical Transition? Intergenerational Educational and Occupational Mobility in Rural China: 19882002. *Policy Research Working Paper 7316. Washington, DC: World Bank.*

Frankema, E. (2013). Colonial Education and Postcolonial Governance in the Congo and Indonesia. *in: E. Frankema and F. Buelens (eds.), Colonial Exploitation and Economic Development. The Belgian Congo and the Netherlands Indies Compared, London and New York: Routledge*, pages 153–177.

Fujita, M. and Thisse, J.-F. (1996). Economics of agglomeration. *Journal of the Japanese and international economies*, 10(4):339–378.

Gallego, F. A. and Woodberry, R. (2010). Christian Missionaries and Education in Former African Colonies: How Competition Mattered. *Journal of African Economies*, 19:294–329.

Gauthier, B. P., Abouly, Y., Yi Chang, A., and Adelman, M. A. (2021). Service delivery indicators report: Drc. *UNESCO report*, (ED/GEMR/MRT/2021/P1/14).

Guinnane, T. W. (2023). We do not know the population of every country in the world for the past two thousand years. *The Journal of Economic History*, 83(3):912938.

Guirkinger, C. and Villar, P. (2022). Pro-birth policies, missions and fertility : historical evidence from Congo. *DeFiPP Working Paper 04*.

Heath Milsom, L. (2023). Spatial inequality of opportunity in west africa. *Journal of Public Economics*, 225:104944.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., and Verashchagina, A. (2008). The inheritance of educational inequality: International comparisons and fifty-year trends. *The BE Journal of Economic Analysis & Policy*, 7(2).

Huillery, E. (2009). History Matters: The Long-Term Impact of Colonial Public Investments in French West Africa. *American Economic Journal: Applied Economics*, 1:176–215.

Iannaccone, L. R. (1998). Introduction to the economics of religion. *Journal of economic literature*, 36(3):1465–1495.

Irvine, C. (1978). The Church of Christ in Zaire : a handbook of Protestant churches, missions and communities, 1878-1978. *Indianapolis : Dept. of Africa, Division of Overseas Ministries, Christian Church*.

Iyer, S. (2016). The new economics of religion. *Journal of Economic Literature*, 54(2):395–441.

Jedwab, R., Kerby, E., and Moradi, A. (2017). History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities In Kenya. *The Economic Journal*, 127(603):1467–1494.

Jedwab, R., Meier zu Selhausen, F., and Moradi, A. (2022). The Economics of Missionary Expansion: Evidence from Africa and Implications for Development. *Journal of Economic Growth*, 27:149192.

Kelly, M. (2019). *The Standard Error of Persistence*. UCD Working Paper.

King, E. M. and Hill, M. A. (1993). *Women's education in developing countries*. The World Bank.

Klein Goldewijk, K., Beusen, A., van Drecht, G., and de Vos, M. (2011). The hyde 3.1 spatially explicit database of human induced land use change over the past 12,000 years. *Global Ecology and Biogeography*, 20:73–86.

Lauro, A. (2020). Women in the Democratic Republic of Congo. In Spear, T., editor, *The Oxford Research Encyclopedia of African History*. Oxford: Oxford University Press.(Oxford Research Encyclopedias).

Le Rossignol, E., Donat-Bouillud, P., and Hirschi, C. (2023). Christian missions and education quality: Evidence from the democratic republic of the congo. *Working paper*.

Lembagusala Kikumbi, A. (2018). L'education des filles dand les écoles secondaires du diocèse de Kikwit en R.D.Congo (1960-1997). *PhD Dissertation*.

MacGaffey, W. (1982). Education, religion, and social structure in zaire. *Anthropology & Education Quarterly*, 13(3):238–250.

Masandi, P. K. (2004). L'Éducation féminine au congo belge. *Paedagogica Historica*, 40(4):479–508.

McMillan, M., Rodrik, D., and Íñigo Verduzco-Gallo (2014). Globalization, structural change, and productivity growth, with an update on africa. *World Development*, 63:11–32. Economic Transformation in Africa.

Meier zu Selhausen, F. (2019). Missions, Education and Conversion in Colonial Africa. In Mitch, D. and Cappelli, G., editors, *Globalization and the Rise of Mass Education*, pages 25–59. Palgrave, Macmillan.

Montgomery, M. (2017). Colonial Legacy of Gender Inequality: Christian Missionaries in German East Africa. *Politics Society*, 45:225–268.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Nunn, N. (2010). Religious Conversion in Colonial Africa. *American Economic Review Papers and Proceedings*, 100:147–152.

Nunn, N. (2014). Gender and Missionary Influence in Colonial Africa. *Africa's Development in Historical Perspective. New York: Cambridge University Press*, pages 489–512.

Nunn, N. and Wantchekon, L. (2011). The slave trade and the origins of mistrust in Africa. *American Economic Review*, 101:322152.

Okoye, D. (2022). Historical christian missions and african societies today: Perspectives from economic history. *Economic History of Developing Regions*, 37(3):315–332.

Okoye, D. and Pongou, R. (2023). Missions, fertility transition, and the reversal of fortunes: evidence from border discontinuities in the emirates of Nigeria. *J Econ Growth*.

Panin, A. (2021). Four reasons why analysis of economic policy and religion go hand-in-hand in sub-saharan africa.

Rocha, R., Ferraz, C., and Soares, R. R. (2017). Human capital persistence and development. *American Economic Journal: Applied Economics*, 9(4):105–36.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system gmm in stata. *The Stata Journal*, 9(1):86–136.

Roome, W. (1924). *Ethnographic Survey of Africa: Showing the Tribes and Languages; Also the Stations of Missionary Societies [map]*.

Service de l'Enseignement (1952). Organisation de lenseignement libre subsidié pour indigènes avec le concours des sociétés de missions chrétiennes.

Shapiro, D. (1996). Fertility decline in Kinshasa. *Population Studies*, 50:89103.

Sheline, Y. E., Papagiannis, G. J., and Grant, S. R. (1984). The effect of school sponsorship on academic achievement: a comparison of catholic, protestant and government secondary schools in zaire. *Comparative Education*, 20(2):223–236.

Solon, G. (2018). What do we know so far about multigenerational mobility? *The Economic Journal*, 128(612):F340–F352.

Stoop, N., Verpoorten, M., and Deconinck, K. (2019). Voodoo, vaccines, and bed nets. *Economic Development and Cultural Change*, 67(3):493–535.

Streit, K. (1929). *Atlas Hierarchicus: Descriptio geographica et statistica Sanctae Romanae Ecclesiae tum occidentis tum orientis juxta statum praesentum accedunt nonnullae notae historicae necnon ethnographicae*. Paderbornae : sumptibus typographiae Bonifacianae.

Tabutin, D. (1982). Evolution régionale de la fécondité danslouest du Zaïre. *Population (french edition)*, page 2950.

Torche, F. (2019). Educational mobility in developing countries. WIDER Working Paper Series wp-2019-88, World Institute for Development Economic Research (UNU-WIDER).

Valencia Caicedo, F. (2019a). Missionaries in Latin America and Asia: A First Global Mass Education Wave. In Mitch, D. and Cappelli, G., editors, *Globalization and the Rise of Mass Education*, pages 61–98. Palgrave, Macmillan.

Valencia Caicedo, F. (2019b). The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America. *The Quarterly Journal of Economics*, 134:507556.

Van Wing, J. (1949). Annuaire des Missions Catholiques au Congo Belge et au Ruanda. *Lédition universelle, Bruxelles*.

Vinck, H., Briffaerts, J., Herman, F., and Depaepe, M. (2006). Expériences scolaires au congo belge: Etude exploratoire. In *Annales Aequatoria*, pages 5–101. JSTOR.

Voth, H.-J. (2020). Persistence - Myth and Mystery. *CEPR Discussion Papers, Centre for Economic Policy Research*, 15417.

Vyborny, K. (2024). Religious LeadersÄô Compliance with State Authority: Experimental Evidence from COVID-19 in Pakistan. *The World Bank Economic Review*, 38(3):514–534.

Wantchekon, L., Klanja, M., and Novta, N. (2015). Education and Human Capital Externalities: Evidence from Colonial Benin. *The Quarterly Journal of Economics*, 30:703–757.

World Bank (2021). Empowering girls and enhancing learning in drc. note 2: Supply and demand side determinants of the gender gap.

Yates, B. A. (1976). The Triumph and Failure of Mission Vocational Education in Zaïre 1879-1908. *Comparative Education Review*, pages 193–208.

Yates, B. A. (1982). Colonialism, education and work: Sex differentiation in Colonial Zaïre. *in Edna G. Bay, ed., Women and Work in Africa, Westview Press, Boulder*, pages 127–52.

Online Appendix

School Clustering and Religious Competition: Persistence of Educational Inequality in Colonial and Post-Colonial D.R. Congo

February 21, 2025

Pablo Álvarez-Aragón Catherine Guirkinger Paola Villar

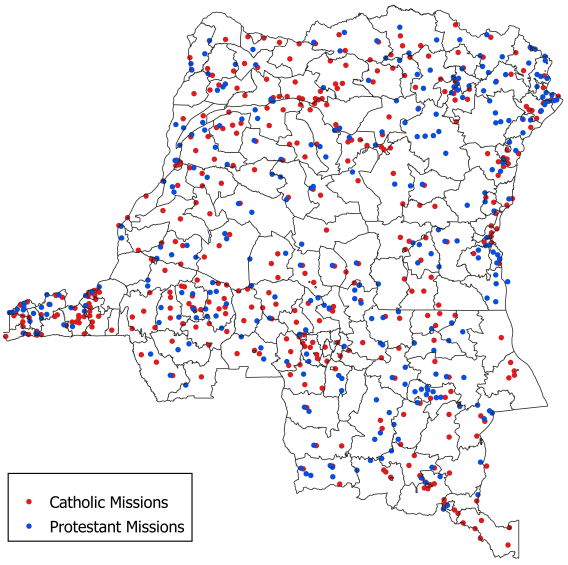
University of Namur *University of Namur* *Université Paris Cité*

A. Data and variables

A.1. Information on Christian missions, 1885-1948

A.1.1. Location of Christian Missions in 1948

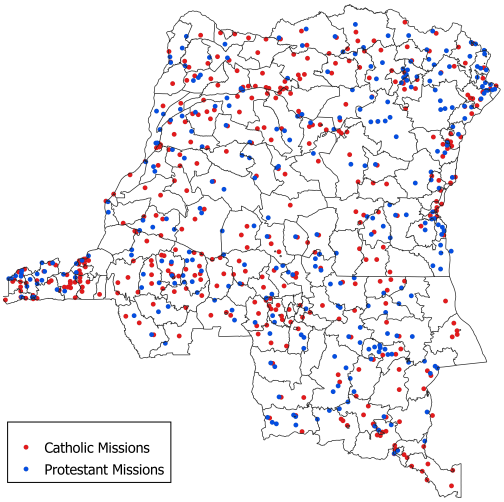
Figure A1: Location of Christian Missions in 1948



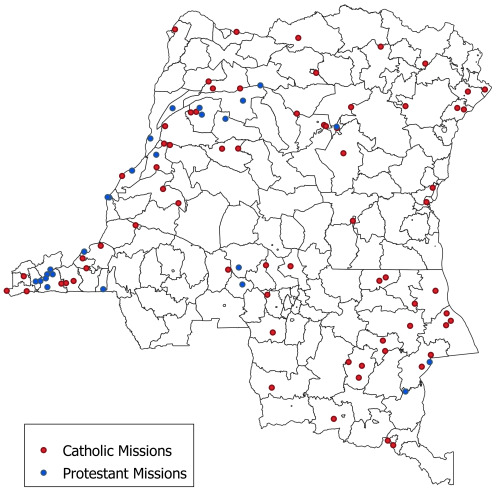
Note: The map shows the exact location of Catholic and Protestant missions in the DRC in 1948, along with territory level boundaries. The digitization has been done by the authors.

A.1.2. Data comparison with Cagé and Rueda (2020), Nunn (2010)

Figure A2: Data comparison with Cagé and Rueda (2020)

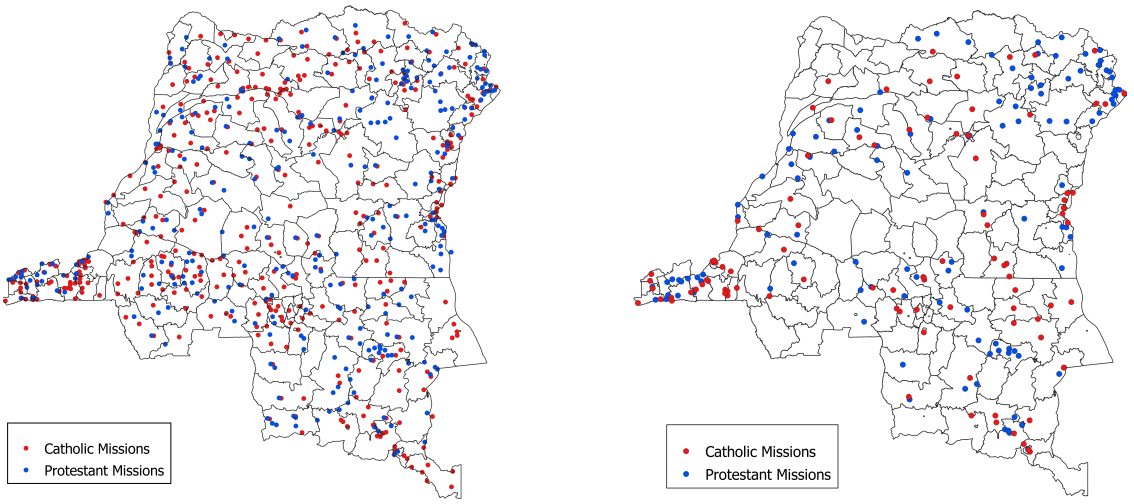


(a) Our data



(b) Cagé and Rueda (2020)

Figure A3: Data comparison with Nunn (2010)



(a) Our data (b) Nunn (2010)

A.2. Measures of exposure to missions

Table A1: Exposure to missionary presence in 1900,1948 and DHS

	N	Mean	Standard deviation	Max	Min
Exposure to Catholic missions in 1900	148	-230.14	145.61	-592.28	-4.52
Exposure to Catholic missions in 1948	148	-37.81	19.76	-114.50	-2.98
Exposure to Catholic missions in DHS	785	-21.64	20.83	-126.20	-0.37
Exposure to Catholic with nuns in 1900	148	-306.26	187.33	-790.68	-4.52
Exposure to Catholic with nuns in 1948	148	-54.43	31.47	-158.61	-2.98
Exposure to Catholic with nuns in DHS	785	-35.40	35.88	-189.21	-0.37
Exposure to Protestant missions in 1900	148	-142.64	87.24	-362.27	-6.43
Exposure to Protestant missions in 1948	148	-45.61	23.83	-161.64	-6.43
Exposure to Protestant missions in DHS	785	-32.16	28.33	-191.05	-0.33

NOTE. Exposure to mission of type X in 1900 and in 1948 is our measure of average exposure computed at the territory level. It is computed as the average (negative) distance in kilometres between 1000 random points generated within the territory in the corresponding year and the nearest mission of type X. With the DHS data, exposure is measured as the (negative) distance in kilometres from each DHS cluster to the closest mission in 1948. N corresponds to the number of clusters in the DHS, while it corresponds to the number of territories for our historical distances.

A.3. Demographic Survey from the 1950s

Another demographic survey was conducted in the 1950s throughout the country. The data from this survey have been published in the form of aggregates at the territory

level and by age category (5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-44, 45-54 and +55).[56] The survey covered about 11% of the total population and its sampling strategy was designed to be representative of both rural and urban areas. It includes information on the number of children enrolled in primary and secondary schools, the number of people who can read and write, and the number of people with no education. We digitized these data from the published booklets that we found in various libraries in Belgium. Since the information provided by the 1950s Demographic Survey is aggregated at the *Territory x Age-cohort* level, we mostly rely on the 1970s Demographic Survey in our main specifications, and use the 1950s data for descriptive purposes and robustness analyses.

[56] At the time of the survey, the territory was the smallest administrative unit in the Democratic Republic of the Congo, smaller than the district and than the province. There were 138 territories in 1955.

B. Catholic and Protestant Educational Programs in colonial times

Different views on girls' education in Catholic and Protestant schools.- The strong conservative view existing in the Catholic system regarding girls' education is well summarized in the following paragraph, which is an excerpt of the official (governmental) program of 1952 to be followed by subsidized schools in Congo.

> "The ideal would be to be able to pursue the moral and intellectual uplift of boys and girls at the same pace. Unfortunately, this ideal cannot be realized due to a variety of unfavorable circumstances: [...] lower intellectual receptivity of girls compared to boys [...] It is necessary to design a girls education program that trains good wives and good mothers and that does not neglect practical branches such as gardening, cooking, washing, ironing, sewing, child care, hygiene, and housekeeping.(Service de l'Enseignement (1952), p.18)".

The difference with the Protestant viewpoint is striking as illustrated by the following excerpt of the minutes of the Congo Missionary Conference of 1924 (this conference gathered annually Protestant missions active in the Congo):

> "In this land, where the womans social position is so much lower than that of the mans we believe that a mixed school is a powerful factor in lifting the woman to be mans equal. The girls being in the same class as boys have in the daily competition an opportunity of showing an equal aptitude for learning as the boys. Thus we expect the old idea of woman's inferiority to man will be done away with.(Congo Missionary Conference (1924), p.33)".

Differences across Protestant and Catholic schools.- The analysis of the 1950 Report to the Houses of Parliament shows that, in addition to a greater number of schools and pupils, the number of pupils per school was also higher among Catholics than Protestants (41 vs. 30)(Administration coloniale, 1950). While these figures should be treated with caution (and may mask considerable heterogeneity), they suggest that Catholic schools were larger than Protestant ones. Unfortunately it is difficult to draw conclusions regarding school quality from these reports (there is no consistent information on the number of teachers per school for example). While the education promoted by Protestants may have been more content- and quality- driven than that of Catholics (see Section 4.1), the latter were much better connected to colonial administration and large companies. Vinck et al. (2006) indicates that Catholic education (especially at secondary level) provided access to a better occupational opportunities, especially for boys.

C. Long-term Effects

C.1. Econometric specification

Let i denote individuals, c denote DHS clusters, and p denote provinces.[57] Our regression equation is:

$$y_{icp} = \beta_0 + \beta_C M_c^C + \theta_C M_c^C F_i + \beta_P M_c^P + \theta_P M_c^P F_i + \lambda F_i + \mathbf{fl} W_i + \mathbf{\Phi} X_c + \alpha_p + \sigma_r + \varepsilon_{icp} \quad (2)$$

where $y_{i,c,p}$ is the educational outcome of interest. M_c^C (M_c^P) measures the exposure of cluster c to Catholic (Protestant) missions and it is expressed as the negative (log) distance of cluster c to the nearest Catholic (Protestant) mission in 1948. α_p denotes province fixed effects, which are included in all regressions and capture time-invariant province-level characteristics, such as colonial policies or unobserved heterogeneity, that may be correlated with average education and missionary activity. Finally, σ_r represents the DHS survey round (2007 or 2013). The vector $W_i^{'}$ contains individual-level control variables such as age, age squared, and gender. Finally, we also include the vector $X_c^{'}$, which contains geographic and historical controls computed at the DHS cluster level. Their selection is based on the main determinants of mission location listed by Jedwab et al. (2022) (see Section 3.1 for details about these variables).[58] Since our main variables of interest are defined at the DHS cluster level, we cluster the standard errors at this level.

C.2. Threats to identification and Robustness Checks

In this Appendix we provide further discussion and analyses in order to rule out a spurious correlation between mission location and education outcomes today. First, we show that our results are not sensitive to the inclusion of any particular variable related to the determinants of mission location. Second, we discuss the potential endogeneity of historical post location. Third, we investigate whether missions left lasting impact on

[57] Today, the national territory of the DRC is divided into 26 provinces, the province into territories/cities, the territory/city into *collectivités*, the *collectivité* into groups/neighborhoods, and the group/neighborhood into villages/streets.

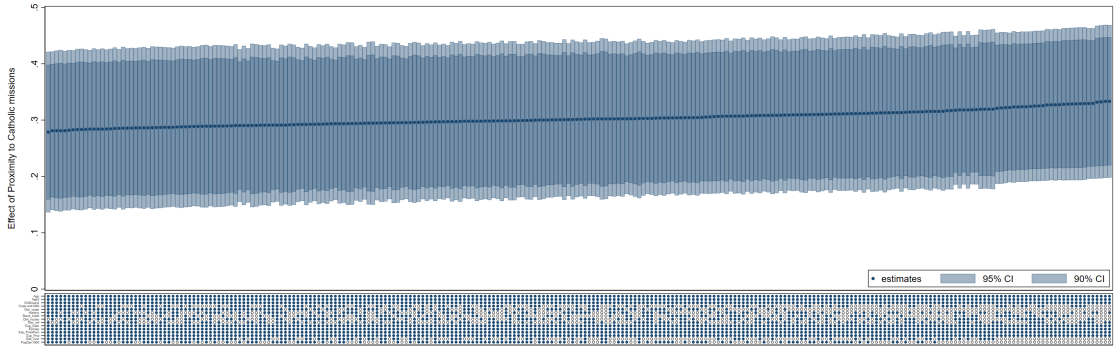
[58] One standard control used in the literature (which mainly includes multi-country analyses) is distance to the capital city. However, we do not include it here, since in the DRC context, distance to the coast is highly correlated with distance to the capital city, Kinshasa ($\rho = 0.95$).

the main dimension they targeted: religion. Finally, we perform a series of robustness checks.

Specification curves.– Figures C1 and C2 show the coefficient stability of our main results to different combinations of relevant control variables. Under random assignment, point estimates should not dramatically change after the inclusion of control variables. We examine whether the coefficients associated to missionary presence are sensitive to some of the variables usually identified as determining the initial position of missions. As we observe below, the coefficients are barely sensitive to the inclusion of different combinations of control variables, reducing the concerns about the endogeneity bias potentially introduced by missionary location. This effects are consistent with the accounts of Wantchekon et al. (2015), who argue that colonial schools' location was made with very little information, and therefore were quasi-random in practice. In the context of the DRC, early reports from missionary provide qualitative evidence in this direction:

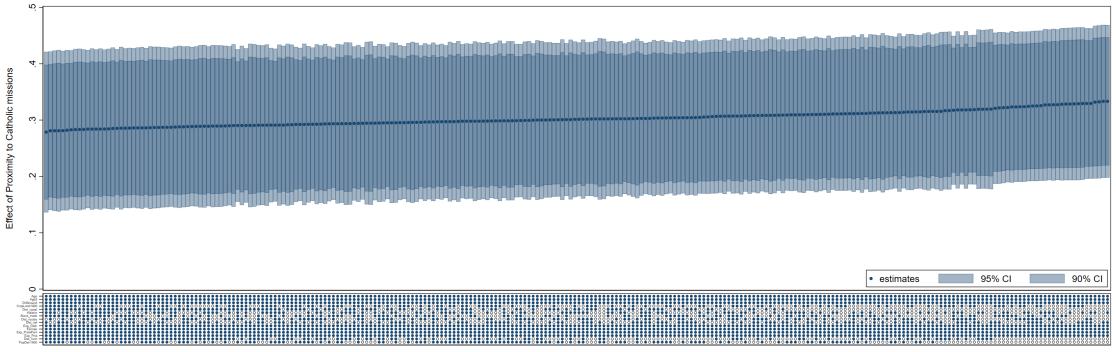
> "Sometimes the question is asked, why was a station established at such and such place, or why was a house built on such and such a site, because today, to a newcomer, it appears that the place or site is not the best that could have been chosen. But if one could realize the circumstances that existed at the time of selection, one would probably find that there was no other course open to our brethren who had to act. The people were not always so willing to welcome white men as now, and there was much jealousy on the part of neighboring villages, which often operated against the most desirable site being occupied (Congo Missionary Conference (1911), p.23)".

Figure C1: Specification curve: Exposure to Catholics x Female



Note: The figure shows the specification curve for the effect of exposure to Catholic missions interacted with the female dummy. Each dot is a coefficient from Eq. 2 with a different combination of control variables. The vertical bars, from darkest to lightest, denote the 95% and 90% confidence intervals. All specifications include province fixed effects.

Figure C2: Specification curve: Exposure to Protestants x Female



Note: The figure shows the specification curve for the effect of exposure to Protestant missions interacted with the female dummy. Each dot is a coefficient from Eq. 2 with a different combination of control variables. The vertical bars, from darkest to lightest, denote the 95% and 90% confidence intervals. All specifications include province fixed effects.

Endogenous location of historical posts.–Our main specification includes an extensive state-of-art set of historical and geographic controls, which the literature have identified as important determinants of mission location (Jedwab et al., 2022). Nevertheless, we cannot rule out the presence of unobserved confounding factors. In particular, an important concern is that our treatment (exposure to missions) mainly captures differential degrees of urbanization. Indeed, most colonial posts in the DRC (but not all), became cities or towns after independence. In that case, our treatment would be a proxy for distance to the city, and our empirical strategy would consist in a comparison between more and less urbanized areas. To account for this, we have always included population density in 1900 as a proxy for current urbanization (the correlation between our measure of population density in 1900 and an index of urbanization in 2000 is 0.81).[59] We do not directly control for current urbanization in our main specifications as it can be understood as a *bad control* in the sense of Angrist and Pischke (2009). Table C1 shows the re-estimation of Equation 2 including urbanization in 2000 as a control and shows that our results barely move. Even if this variable can be considered as a *bad control*, the fact that our results change little when we control for current urbanization suggests that it is not the main driver of the correlation (Table C1). Similarly, Table C2 estimates Eq. 2 using the sample of rural DHS cluster only. Although this sample restriction may

[59]We acknowledge that these historical measures of population density have strong limitations (Guinnane, 2023). However, in our case, historical population density strongly correlates with current population density, which is derived from remote sensing image collections (e.g. Landsat imagery). Moreover, our results are not sensitive to the inclusion of historical population density.

introduce bias due to the bad control problem, we obtain similar results.[60]

Table C1: Exposure to missions, education and urbanization

	(1) Years of education	(2) Literacy	(3) Secondary or more	(4) $>$ 12 years of educ
Exposure to Catholic	0.00298	-0.0184*	-0.0110	0.0157**
	(0.103)	(0.0104)	(0.0118)	(0.00787)
Exp to Catholic x Female	0.262***	0.0566***	0.0463***	-0.0134**
	(0.0730)	(0.0104)	(0.0106)	(0.00642)
Exposure to Protestant	-0.0341	-0.0172	-0.0139	-0.00240
	(0.110)	(0.0112)	(0.0131)	(0.00839)
Exp to Protestant x Female	0.348***	0.0563***	0.0524***	0.0110
	(0.0699)	(0.00967)	(0.0108)	(0.00713)
Urbanization	4.828***	0.459***	0.491***	0.280***
	(0.471)	(0.0402)	(0.0437)	(0.0362)
FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Mean Y	6.933	0.637	0.566	0.165
R-squared	0.334	0.223	0.255	0.179
N	34654	34654	34654	34654

Note. Data: Pooled 2007 and 2013 waves of the DHS. The sample includes women aged 15-49 and men aged 15-59. The sample is restricted to people living in DHS clusters located 50km or closer to a mission. The table reports OLS estimates. We use $-log(distance)$ to measure exposure. "Urbanization" is an index ranging from 0.00 (extremely rural) to 1.00 (extremely urban) for the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location. Outcome variables are defined as follows: single years of education in column (1), ability to read a whole sentence in column (2), probability of having post-primary education in column (3) and probability to have 12 or more years of education in column (4). Standard errors () are clustered at DHS cluster level. Geographical, historical and individual controls included. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

[60]Now, the long-term effects of Protestant missions seem to be somewhat stronger, driven by their influence on post-primary education. This is not very surprising. As we briefly discuss in Section 3.1, 35% of the 35.000 secondary schools that operate in the DRC in 2020 are Protestant, as compared with 15% of Catholic secondary schools.

Table C2: Exposure to missions, education and urbanization

	(1) Years of education	(2) Literacy	(3) Secondary or more	(4) $>$ 12 years of educ
Exposure to Catholics	-0.133	-0.0143	-0.00826	-0.00428
	(0.117)	(0.0129)	(0.0152)	(0.00761)
Exposure to Catholics x Female	0.215**	0.0293**	0.0187	0.00734
	(0.0959)	(0.0120)	(0.0128)	(0.00741)
Exposure to Protestants	0.122	0.0155	0.0180	-0.0117
	(0.138)	(0.0142)	(0.0183)	(0.0103)
Exposure to Protestants x Female	0.336***	0.0382***	0.0322**	0.0250**
	(0.112)	(0.0145)	(0.0164)	(0.0106)
FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Mean Y	5.209	0.480	0.392	0.0716
R-squared	0.231	0.170	0.189	0.0877
N	20031	20031	20031	20031

Note. Data: Pooled 2007 and 2013 waves of the DHS. The sample includes women aged 15-49 and men aged 15-59. The sample is restricted to people living in rural DHS clusters. The table reports OLS estimates. We use $-log(distance)$ to measure exposure. Outcome variables are defined as follows: single years of education in column (1), ability to read a whole sentence in column (2), probability of having post-primary education in column (3) and probability to have 12 or more years of education in column (4). Standard errors () are clustered at DHS cluster level. Geographical, historical and individual controls included (see Section 4 for details). *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

Mission location and religion. – Conversion was the primary goal of missionary and historians accounts suggest that providing education was often seen as a strategy to attract new converts. If missionary activities are to have left long-lasting traces, we expect them first on the religious dimension. To investigate whether this is the case, we replace education with religion in the estimated model (Equation 2). Results presented in Table C3 confirm that missions triggered changes that have persisted until today: proximity to historical Catholic missions increases the probability of being Catholic, while proximity to historical Protestant missions increases the probability of being Protestant.

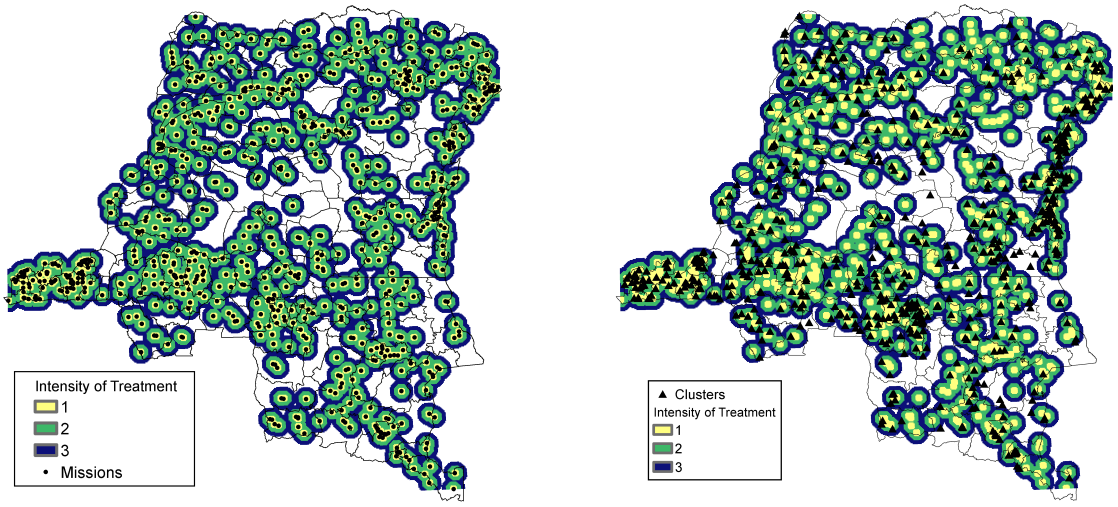
Table C3: Historical missionary presence and religion today

	(1) Religion today: Catholic	(2) Religion today: Protestant
Exposure to Catholic missions	0.0395***	-0.0172
	(0.0108)	(0.0107)
Exposure to Protestant missions	0.00913	0.0270**
	(0.0134)	(0.0134)
FEs	Yes	Yes
Controls	Yes	Yes
Mean Y	0.293	0.267
R-squared	0.0877	0.0957
N	34654	34654

NOTE. Data: Pooled 2007 and 2013 waves of the DHS. The sample includes women aged 15-49 and men aged 15-59. The sample is restricted to people living in DHS clusters located 50km or closer to a mission. We use $-log(distance)$ to measure exposure. The table reports OLS estimates. Outcome variables are dummies that equal one if the respondent reports that he/she is Catholic or Protestant, respectively. Standard errors () are clustered at DHS cluster level. Geographical, historical and individual controls included. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

Alternative measure of exposure to Christian missions – First, we consider an alternative measure of exposure to past missionary presence. Instead of measuring exposure as the negative (logarithmic) distance from each DHS cluster to the nearest mission, which in the context of the DHS data introduces classical measurement error due to random cluster displacements, we measure exposure using different sets of Euclidean distance buffers. Therefore, we classify each DHS cluster into three categories: highly exposed (0-16 km away from a mission), moderately exposed (16-33 km away), and weakly exposed (33-50 km away) (Figure C3). As in our main specification, we find in Table C4 that women in highly exposed clusters are more educated today. Second, if missions are located in better areas and educated people sort themselves into more favorable areas, migration would be a concern.

Figure C3: Intensity of Treatment



(a) Missions and area of influence (b) DHS clusters and missions' intensity

Notes: Figure C3a shows the different intensity buffers. Number "1" corresponds to the area within a 16km radius from a mission, while "2" represents the area between 16km-33km and "3" represents the area between 33km-50km away from a mission. On the other hand, Figure C3b shows the correspondence between DHS clusters and the intensity measures. If a DHS cluster is 20km away from its closest mission, then it is treated with "medium" intensity. Only the closest mission is considered when assigning exposure.

Table C4: Degree of exposure to missions and education

	(1) Years of education	(2) Literacy	(3) Secondary or more	(4) $>$ 12 years of educ
High Intensity Treatment	0.128	-0.0206	-0.0125	0.0112
	(0.288)	(0.0273)	(0.0321)	(0.0206)
HIT x Female	0.502**	0.106***	0.0987***	-0.00624
	(0.206)	(0.0280)	(0.0249)	(0.0184)
Low intensity treatment	0.355	0.0555	0.0456	-0.0319
	(0.311)	(0.0364)	(0.0382)	(0.0223)
LIT x Female	-0.561*	-0.0930**	-0.0654*	0.0201
	(0.294)	(0.0383)	(0.0335)	(0.0239)
FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Mean Y	6.933	0.637	0.566	0.165
R-squared	0.301	0.197	0.229	0.165
N	34654	34654	34654	34654

Note. Data: Pooled 2007 and 2013 waves of the DHS. The sample includes women aged 15-49 and men aged 15-59. The sample is restricted to people living in DHS clusters located 50km or closer to a mission. The table reports OLS estimates. Exposure measures at described in Section C.2. Outcome variables are defined as follows: single years of education in column (1), ability to read a whole sentence in column (2), probability of having post-primary education in column (3) and probability to have 12 or more years of education in column (4). Standard errors () are clustered at DHS cluster level. Geographical, historical and individual controls included. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

Spatial autocorrelation. – A final concern is related to the existence of spatial autocorrelation, which could reduce our estimated standard errors. Following recent sugges-

tions (i.e., Kelly (2019) or Voth (2020)), we always include province fixed effects, latitude, longitude and malaria as controls, and several variables measuring distances. Moreover, we always restrict our sample to DHS clusters located less than 50km away from the closest mission, reducing the concern that DHS clusters are similar to distant places, and omitting regions with extreme values of the explanatory variable. Finally, we show in Table C5 that the statistical significance of our results is unaffected when adjusting for spatial autocorrelation using Conley standard errors(Conley, 1999).[61]

Table C5: Education in 2000 and missions, 100km Conley standard errors

	(1)	(2)	(3)	(4)
	Years of education	Literacy	Secondary or more	$>$ 12 years of educ
Exposure to Catholic	0.0720	-0.0118	-0.00398	0.0197**
	(0.119)	(0.0108)	(0.0117)	(0.00928)
	[0.137]	[0.010]	[0.013]	[0.009]
Exposure to Catholic x Female	0.273***	0.0576***	0.0475***	-0.0127*
	(0.0712)	(0.0102)	(0.0103)	(0.00649)
	[0.077]	[0.008]	[0.010]	[0.007]
Exposure to Protestant	-0.00320	-0.0143	-0.0107	-0.000604
	(0.121)	(0.0119)	(0.0136)	(0.00923)
	[0.127]	[0.013]	[0.015]	[0.006]
Exposure to Protestant x Female	0.348***	0.0564***	0.0525***	0.0110
	(0.0689)	(0.00946)	(0.0106)	(0.00722)
	[0.055]	[0.011]	[0.012]	[0.008]
FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Mean Y	6.933	0.637	0.566	0.165
R-squared	0.304	0.202	0.232	0.165
N	34654	34654	34654	34654

NOTE. Data: Pooled 2007 and 2013 waves of the DHS. The sample includes women aged 15-49 and men aged 15-59. The sample is restricted to people living in DHS clusters located 50km or closer to a mission. The table reports OLS estimates. We use $-log(distance)$ to measure exposure. Outcome variables are defined as follows: single years of education in column (1), ability to read a whole sentence in column (2), probability of having post-primary education in column (3) and probability to have 12 or more years of education in column (4). Standard errors are either clustered at DHS cluster level () or account for spatial autocorrelation (100km Conley standard errors) []. Geographical, historical and individual controls included (see Section 4 for details). *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

[61] These results are consistent to the election of different thresholds such as 25km or 50km (not shown).

C.3. *Long term effects by birth decade*

Table C6: Education in 2000 and missions, by birth decade

	Dependent variable: Years of education			
	(1)	(2)	(3)	(4)
Exposure to Missions	0.315**	0.468***		
	(0.142)	(0.159)		
Exposure to Mission x (1980-1989)	0.102	-0.00476		
	(0.115)	(0.141)		
Exposure to Mission x (1970-1979)	0.0186	-0.108		
	(0.130)	(0.151)		
Exposure to Mission x (1950-1970))	0.120	0.0270		
	(0.160)	(0.173)		
Exposure to Catholic			0.271*	0.320**
			(0.139)	(0.157)
Exposure to Catholic x (1980-1989)			0.00921	-0.0124
			(0.113)	(0.132)
Exposure to Catholic x (1970-1979)			-0.0464	-0.0756
			(0.138)	(0.154)
Exposure to Catholic x (1950-1970)			-0.000437	0.0693
			(0.148)	(0.171)
Exposure to Protestant			0.144	0.259*
			(0.141)	(0.147)
Exposure to Protestant x (1980-1989)			0.155	0.136
			(0.107)	(0.120)
Exposure to Protestant x (1970-1979)			0.166	0.0873
			(0.130)	(0.141)
Exposure to Protestant x (1950-1970)			0.194	0.0212
			(0.135)	(0.159)
FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Mean Y	7.005	6.101	7.005	6.101
R-squared	0.337	0.317	0.338	0.319
N	25561	17266	25561	17266

NOTE. Data: Pooled 2007 and 2013 waves of the DHS. The sample includes women aged 15-49 and men aged 15-59. The reference category is individuals born after 1989. Columns 2 and 4 restrict the sample to women. The sample is restricted to people living in DHS clusters located 50km or closer to a mission. The sample is restricted to respondents over the age of 20. The table reports OLS estimates. We use $-log(distance)$ to measure exposure. The outcome variable is single years of education. Standard errors are either clustered at DHS cluster level (). Geographical, historical and individual controls included (see Section 4 for details). *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

C.4. *Exposure to missions and children's education*

Table C7: Children's education and exposure to former missionary presence

	Years of education	
	Children aged 6-11 (1)	Children aged 12-17 (2)
Exposure to Catholics	0.00385	0.0536
	(0.0251)	(0.0668)
Exposure to Catholics x Female	0.0213	0.114*
	(0.0223)	(0.0636)
Exposure to Protestants	0.0146	-0.0251
	(0.0285)	(0.0601)
Exposure to Protestants x Female	0.0363	0.126**
	(0.0298)	(0.0562)
FEs	Yes	Yes
Controls	Yes	Yes
Mean Y	1.311	4.933
R-squared	0.500	0.348
N	16707	9361

NOTE. Data: Pooled 2007 and 2013 waves of the DHS. The sample includes boys and girls aged 6-17. The sample is restricted to people living in DHS clusters located 50km or closer to a mission. The table reports OLS estimates. We use $-log(distance)$ to measure exposure. The outcome variable is single years of education. Standard errors () are clustered at DHS cluster level. Geographical, historical and individual controls included (see Section 4 for details). *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

D. Short-term Effects

D.1. *Econometric specification, 1970s data*

Let i index individuals, t index territory of birth (since most of the sample consists of urban migrants), a index year of birth and m the type of mission ($m = 1$ denotes Catholic missions, $m = 2$ denotes Catholic missions with at least one nun and $m = 3$ refers to Protestant missions). Using these notations, Equation 3 describes our preferred empirical specification:

$$y_{i,t,l} = \beta_0 + \beta_C D^C_{t,l} + \beta_P D^P_{t,l} + \beta_N D^N_{t,l} + \alpha_C F_i D^C_{t,l} + \alpha_P F_i D^P_{t,l} + \alpha_N F_i D^N_{t,l} + X_i'\Phi + \theta_l + \sigma_{coh} + \varepsilon_{i,t,l} \quad (3)$$

where $y_{i,t,l}$ is an educational outcome. $D^C_{t,l}$, $D^P_{t,l}$ and $D^N_{t,l}$ are our variables of interest and measure the average exposure of territory l in year t as described in Section 3.2. The coefficient associated to $D^N_{t,l}$ (exposure to Catholic missions with nuns) captures the effect of posts with at least one Catholic nun, conditional on the exposure to general Catholic

missions and Protestant missions. The variable F_i is an indicator $= 1$ if individual i is female, and its interaction with our measures of missionary presence captures the differential impact of missions by gender. X_i denotes a vector of individual-level control variables, which includes the F_i gender indicator, year of installation in the current city and whether the respondent was born in a rural or urban area. Finally, θ_l and σ_{coh} denote territory of birth and age-cohort fixed effects respectively. Since our treatment variables are defined at the territory of birth level, we cluster the standard errors at this level.[62] The inclusion of territory fixed effects allows for a within-territory analysis. Therefore, the coefficients of interest are identified by time variations in the exposure to missions in a given territory.

D.2. Threats to identification

D.2.1. Selective migration: discussion

As a first step, we verify that the patterns between the 1950s and 1970s databases are qualitatively comparable. If the effects of mission exposure are similar, then we can assume that the 1970s sample does not indicate any particular differential selection of migrants along the mission exposure dimension. Table D1 in Appendix D.2.2 below shows the results. Overall, the results using the 1950s data go in the same direction as our main results with the 1970s survey.

In order to rule out that selective migration is driving the results on the individual sample from the 1970s, we explicitly examine the extent of selection into migration that exists in the 1970s dataset. First, we quantify the degree of selection into migration by territory and age cohort by calculating the share of educated people (people with some education) in a given territory and age cohort in the representative 1950s sample and by comparing it with the share of educated people in the same age cohort and territory of birth in the 1970s urban sample.[63] Figure D1 in Appendix D.2.3 shows that there is generalized positive selection into migration: the share of educated people in the

[62]Note, however, that clustering at this level may be highly conservative for two reasons (Abadie et al., 2023). First, we observe all clusters in the population and second, treatment assignment is imperfectly correlated within clusters. Still, since conventional robust standard errors may be underestimated, we prefer being overly cautious and report clustered standard errors.

[63]We focus on people over the age of 15, as they should have started school by this age. We assume that the territory of birth is the same as the territory of residence in the 1950s sample, as migration was very limited in the early colonial period.

1970s sample is usually higher than the share of educated people in the 1950s sample (comparing the same territories and age cohorts).

Second, we examine whether the degree of selection into migration is differentially affected by exposure to missionary presence (i.e., whether educated people in highly exposed territories are more likely to migrate to cities than educated people in moderately exposed territories). To do this, we regress the degree of selection into migration (observed at the territory x age cohort level, shown in figure D1) on our measures of exposure to missionaries. Table D2 shows that the magnitude of selection into migration does not depend on the type of exposure. Putting together both analyses suggest that educated people are more likely to migrate, yet this increased propensity is *not* a function of their missionary exposure. This suggests that our estimated effects of missionary exposure on (migrants) education are not the results of a composition effect (whereby the effect of education on the propensity to migrate would be a function of exposure), and can thus be interpreted as a direct influence of mission on the local population's levels of education.

D.2.2. Selective migration: Short-term Effects in the 1950s

In this section, we rely on the demographic survey that was carried out in the 1950s (see Section A.3 for details).

When using the 1950s demographic survey to estimate equation 3, we need to modify our specification slightly. First, instead of using the territory of birth, as in the 1970s survey, to match individuals to our measures of exposure to missionary presence, we use the territory of residence, since the data are aggregated by territory of residence. In addition, since the data are also aggregated by age cohort, we compute the average of age-specific exposure measures (i.e., for the age group 5-9, we first compute the exposure at birth of those aged 5, 6, 7, 8, and 9, and then take the average).

Let t index territory of residence, a index age-group and d index districts.[64] Then, we run the following regression equation for men, women and both sex separately:

$$y_{t,a} = \alpha + D'_{t,a}\beta + X'_t\Phi + \theta_d + \sigma_a + \varepsilon_{t,a} \tag{4}$$

[64]Districts are one level higher administrative units than territories.

Where $y_{t,a}$ is the educational outcome of age group a in territory t. Instead of looking at individual years of education or whether individual i has at least one year of education, the survey provides information on the share of people with (completed) primary and secondary education, the share of people who can read and write, or the share of people who have not studied. Since the variation is higher now, we also include people born before 1930. The vector $D_{t,a}$ contains the (negative) logarithm of our distance measures (averages of age-specific distances) described in section 3.2, but only for Catholics and Protestants due to the high correlation between distance to Catholic missions and distance to Catholic missions with nuns in this dataset ($\rho = 0.94$).[65] Finally, θ_d and σ_a are district and age group fixed effects.

Since we use district fixed effects (θ_d), we need to account for potential endogeneity in the location of missions within the territories of a given district.[66] To do so, we introduce the vector X_t, which contains geographic and historical controls computed at the territory level. Their choice is based on the main determinants of mission location listed by Jedwab et al. (2022) and include: distance to Catholic and Protestant missions in 1885, area, altitude, ruggedness, latitude, longitude and their product, distance to the nearest colonial route, population density and arable land in 1900, a malaria index, a tsetse susceptibility index, number of slaves exported in the Indian and Atlantic trade, distance to the nearest navigable river, and distance to the coast.[67] Standard errors are clustered at the territory of residence level.

Table D1 presents the results. Proximity to Catholic mission increases education for both men and women, while Protestant missions has no detectable effect. Unfortunately, here we are not able to identify the differential effect by gender of Catholic missions with nuns, so we only report the average effect of exposure to general Catholic missions. In terms of magnitude, halving the distance to a Catholic mission increases the share of people with primary school completed by almost 7%, or the share of people able to read and write by about 15%.

[65]This high correlation may introduce a problem of (almost) perfect multicollinearity due to the fact that our exposure measure is now an average of age-specific exposures, which reduces the variation in our explanatory variables.

[66]Including territory fixed effects leaves us with very little variation. When using data from the 1950s, we move our geographic fixed effects up one level.

[67]We do not include distance to the capital because it is highly correlated with distance to the coast (($\rho = 0.95$).

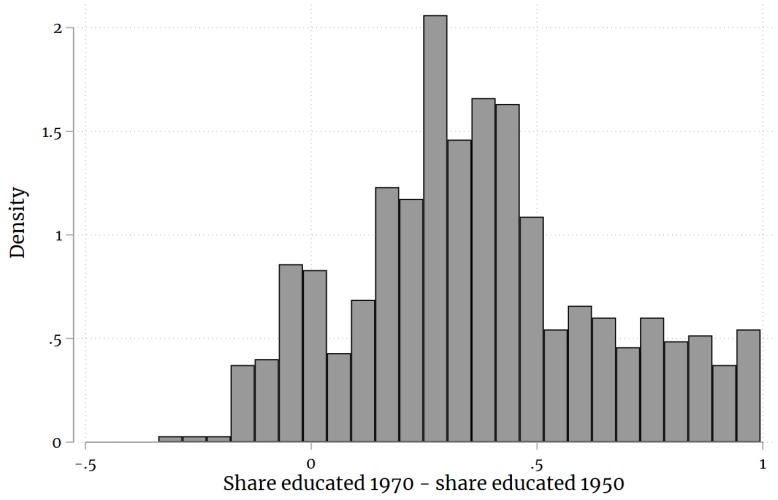
Table D1: Education in 1950s and exposure to missionary presence

	(1) Share Primary	(2) Share Secondary	(3) Share read/write	(4) Share Studies
Panel A: Both sex				
Exposure to Catholic	0.0213***	0.00445**	0.0335***	0.0268***
	(0.00729)	(0.00171)	(0.00860)	(0.00810)
Exposure to Protestant	-0.00481	-0.00249	-0.00878	-0.00529
	(0.00996)	(0.00189)	(0.0103)	(0.0108)
Panel B: Men				
Exposure to Catholic	0.0195**	0.00743**	0.0336**	0.0334***
	(0.00919)	(0.00302)	(0.0132)	(0.0103)
Exposure to Protestant	0.0120	-0.00446	0.00932	0.00137
	(0.0120)	(0.00323)	(0.0141)	(0.0136)
Panel C: Women				
Exposure to Catholic	0.0159**	0.0000411	0.0159*	0.0237***
	(0.00738)	(0.000561)	(0.00880)	(0.00736)
Exposure to Protestant	-0.00783	-0.00138**	-0.00805	-0.00711
	(0.00945)	(0.000666)	(0.00915)	(0.00997)
FEs	Yes	Yes	Yes	Yes
Mean Y both	0.215	0.0107	0.160	0.221
Mean Y men	0.368	0.0222	0.300	0.376
Mean Y women	0.0834	0.00228	0.0517	0.0808
Controls	Yes	Yes	Yes	Yes
R-squared (both)	0.844	0.582	0.775	0.841
N (both)	1082	836	825	1064

NOTE. Data: Demographic Survey of the 1950s. The unit of observation is a "*Territory x Age-Cohort*". In column (2), the sample is restricted to cohorts older than 15 years old. The table reports OLS estimates. Exposure to missionary presence is measured as $-log(distance)$. Outcome variables are defined as follows: share of people in a given cohort and territory who attended primary (column 1) or secondary (column 2) school, share of people in a given cohort and territory able to read or write (column 3), and share of people with studies in a given cohort and territory (column 4). Standard errors () are clustered at the territory level. Total population, geographical and historical controls included. Cohort and district fixed-effects included in all specifications. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

D.2.3. Estimating the selection into migration

Figure D1: Degree of Selection into Migration



NOTE.: Data: Demographic Survey of 1950s and Demographic Survey of 1970s. The graph displays the distribution of the difference between the share of educated people in age-cohort a and territory t in the 1970s sample and the share of educated people in age-cohort a and territory t in the 1950s sample. For example, if this difference is 0.2, it means that the share of educated people in age-cohort a and territory t in the 1970s sample is 20 percentage points larger than the share of educated people in the 1950s sample from that same territory and age-cohort.

Table D2: Selection into migration and exposure to missionary presence

	Degree of Selection into Migration		
	(1)	(2)	(3)
Exposure to Catholic missions	0.00499	0.00621	-0.00215
	(0.0426)	(0.0422)	(0.0447)
Exposure to Protestant missions	0.0523**	0.0213	0.0399
	(0.0260)	(0.0268)	(0.0305)
Exposure to Catholic missions Nuns	0.0197	0.0220	0.0176
	(0.0308)	(0.0351)	(0.0355)
Cohort FE	No	Yes	Yes
District FE	No	No	Yes
Territory-level Controls	Yes	Yes	Yes
Mean Y	0.373	0.373	0.373
R-squared	0.147	0.215	0.249
N	645	645	645
p-value Catholic=Protestant	0.4083	0.7848	0.4968
p-value Catholic=Nuns	0.8337	0.8271	0.7882
p-value Nuns=Protestant	0.4159	0.9869	0.6169

NOTE. Data: Demographic Survey of the 1950s and Demographic Survey of the 1970s. The unit of observation is a "*Territory x Age-Cohort*". The table reports OLS estimates. Exposure to missionary presence is measured as the $-log(distance)$. It is measured as the average of age-specific exposure measures (at birth) of people from age-cohort a born in territory t. The outcome variable is the difference between the share of educated people in cohort a and territory t in the 1970s sample and the share of educated people in cohort a and territory t in the 1950s sample. Territory-level controls are described in Section D.2.2. Standard errors () are clustered at the territory level. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

D.3. Endogenous opening of missions

D.3.1. Determinants of Mission location

Table D3: Differences between Catholic and Protestant posts, by opening decade

	1900-1910			1920-1930			1940-1948		
	Catholic	Protestant	p-value(1-2)	Catholic	Protestant	p-value(1-2)	Catholic	Protestant	p-value(1-2)
	(1)	(2)		(1)	(2)		(1)	(2)	
Longitude	21.963	23.228	0.396	23.132	24.449	0.123	22.561	23.498	0.284
Latitude	-2.599	-1.689	0.482	-3.923	-3.622	0.743	-2.810	-2.776	0.963
Elevation	568.664	664.965	0.393	726.201	735.825	0.899	685.208	719.918	0.659
Ruggedness Index	5.227	4.719	0.754	4.694	5.450	0.422	6.056	5.333	0.515
Malaria Suitability Index	14.588	12.528	0.240	11.790	12.239	0.719	14.236	12.739	0.147
Distance to navigable river (km)	58.271	90.860	0.210	140.256	98.444	0.077	82.527	112.076	0.128
Distance to coast (km)	1194.633	1275.922	0.650	1331.350	1264.488	0.489	1243.386	1509.821	0.010
Distance to Kinshasa (km)	868.453	926.350	0.716	1011.975	941.810	0.438	918.114	1166.291	0.005
Disantance to colonial routes (km-	79.256	108.589	0.189	99.683	107.704	0.581	101.217	91.132	0.524
Distance to colonial railroad (km)	655.757	511.468	0.213	560.390	644.113	0.163	574.660	694.486	0.022
Population density in 1900	18.146	74.364	0.088	10.300	13.945	0.499	14.790	8.341	0.186
Area suitable for agriculture in 1900	5.352	5.578	0.886	4.125	2.739	0.044	3.975	3.024	0.107
TseTse Fly Suitability Index	0.735	0.515	0.340	0.339	0.535	0.217	0.586	0.663	0.554
Exposure to the Atlantic Slave Trade	344,711.844	6,007.636	0.317	149,143.385	86,064.467	0.624	80,851.923	2,497.541	0.207
N	45	11		78	45		52	74	

Note: This table shows the differences in means between Catholic and Protestant posts along selected variables, depending on the decade in which the posts were opened. A description of the variables and their sources can be found in Section 3.1.

D.3.2. Endogenous opening of missions: time-varying controls

Table D4: Education in 1970s and exposure to missionary presence

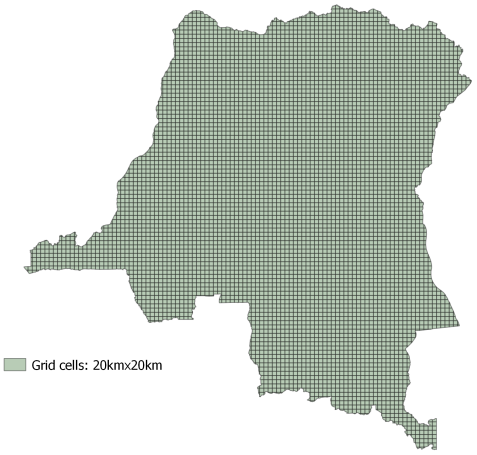
	(1)	(2)	(3)	(4)	(5)	(6)
	Years educ	Any educ	Primary of more	Years educ	Any educ	Primary of more
Catholic missions	0.520***	0.0296	0.0500**	0.700***	0.0862*	0.0804***
	(0.197)	(0.0277)	(0.0228)	(0.241)	(0.0467)	(0.0301)
Catholic missions x female	0.0361	-0.0162	-0.00384	-0.419	-0.147	-0.0698
	(0.255)	(0.0379)	(0.0312)	(0.391)	(0.0908)	(0.0537)
Protestant missions	-0.124	0.0310	0.00935	-0.147	0.0250	0.00658
	(0.294)	(0.0294)	(0.0329)	(0.297)	(0.0296)	(0.0332)
Protestant missions x female	-0.154	0.0188	-0.00959	-0.110	0.0313	-0.00333
	(0.240)	(0.0286)	(0.0310)	(0.245)	(0.0288)	(0.0321)
Catholic with nuns				-0.0176	-0.0314	-0.0259
				(0.232)	(0.0321)	(0.0285)
Catholic with nuns x female				0.420	0.121*	0.0609
				(0.319)	(0.0634)	(0.0396)
FEs	Yes	Yes	Yes	Yes	Yes	Yes
Territory and cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-Varying Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	4.995	0.734	0.435	4.995	0.734	0.435
R-squared	0.443	0.293	0.339	0.444	0.294	0.339
N	41655	41655	41655	41655	41655	41655

Note. Data: Demographic Survey of the 1970s. The sample is restricted to people born after 1930. The table reports OLS estimates. Exposure to missionary presence is measured as $-log(distance)$. Outcome variables are defined as follows: single years of education in columns (1) and (2), any year of education in columns (2) and (5) and the probability of completing primary education (>5 years of education) in columns (3) and (6). Standard errors () are clustered at the territory of birth level. Time-varying district-level information includes the size of the white Belgian and white non-Belgium population, the size of the indigenous population and the amount of per capita tax levied from the indigenous population. Fixed-effects include city of residence, territory of birth, 5yr age cohorts and whether the individual was born in an urban area. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

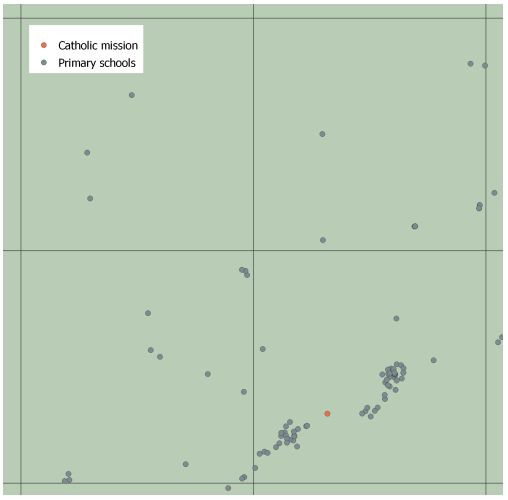
E. Persistence in supply of education

E.1. Historical missions and contemporary schools in the DRC

Figure E1: Historical missions and contemporary schools in the DRC



(a) 20x20km cells



(b) Example

E.2. Missions and contemporary schools: robustness

The strong positive correlation between historical mission schools and contemporary schools is not surprising if the missions were located in urban areas. While this may be true, it is less likely that urbanization explains all of our results. First, we have looked at *schools per 1000 inhabitants* instead of at the absolute number of schools in a grid-cell. Second, we always control for factors that are highly correlated with urbanization in the context of the DRC, such as historical population density, or the presence of a navigable river in the grid-cell. Here we show that the strong positive correlation between historical and contemporary schools holds even when we control for contemporary population density (Tables E1 and E2), and when we restrict the sample to rural schools only (Table E3).[68] Population density in 2020 comes from The Gridded Population of the World (GPW) collection (fourth version), which has an output resolution of about $1km^2$ at the equator.

[68]Note that the introduction of contemporary population density and the sample restrictions are bad controls.

Table E1: Former missionary presence and supply of primary education

	(1) Schools per 1000 people	(2) Proportion of Girls	(3) Schools per 1000 people	(4) Proportion of Girls
Catholic mission	0.831***	0.0613***	0.831***	0.0617***
	(0.155)	(0.0101)	(0.153)	(0.0104)
Protestant mission	0.399***	0.0401***	0.362***	0.0381***
	(0.138)	(0.0129)	(0.135)	(0.0131)
Mean Y	0.701	0.288	0.701	0.286
Collect FE	Yes	Yes	Yes	Yes
Pop den 2020	No	No	Yes	Yes
Grid-cell controls	Yes	Yes	Yes	Yes
R-squared	0.224	0.390	0.231	0.400
N	5786	6010	5786	5814

Note. Data: Universe of primary schools in the DRC in 2020, from Ministry of education. The table reports OLS estimates. The unit of observation is a 20kmx20km grid cell. Contemporary population density is described in Appendix E.2. Outcome variables are defined as follows: number of schools per 1000 inhabitants in 2020 in columns (1) and (3) and average share of girls in school by grid cell in columns (2) and (4). Controls at the grid-cell level are always included: presence of navigable river, presence of colonial railroad, presence of precolonial explorer route, population density in 1900, cropland area in 1900, mean altitude, ruggedness, and distance from the centroid of the grid-cell to Kinshasa. Sector fixed-effects (collectivités) included in all specifications. Robust standard errors in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

Table E2: Former missionary presence and supply of secondary education

	(1) Schools per 1000 people	(2) Proportion of Girls	(3) Schools per 1000 people	(4) Proportion of Girls
Catholic mission	0.495***	0.0691***	0.495***	0.0692***
	(0.0994)	(0.00986)	(0.0988)	(0.0101)
Protestant mission	0.254***	0.0492***	0.233***	0.0470***
	(0.0911)	(0.0112)	(0.0902)	(0.0113)
Mean Y	0.371	0.189	0.371	0.186
Collect FE	Yes	Yes	Yes	Yes
Pop den 2020	No	No	Yes	Yes
Grid-cell controls	Yes	Yes	Yes	Yes
R-squared	0.220	0.433	0.226	0.443
N	5786	5984	5786	5788

Note. Data: Universe of secondary schools in the DRC in 2020, from Ministry of education. The table reports OLS estimates. The unit of observation is a 20kmx20km grid cell. Contemporary population density is described in Appendix E.2. Outcome variables are defined as follows: number of schools per 1000 inhabitants in 2020 in columns (1) and (3), and average share of girls in school by grid cell in columns (2) and (4). Controls at the grid-cell level are always included: presence of navigable river, presence of colonial railroad, presence of precolonial explorer route, population density in 1900, cropland area in 1900, mean altitude, ruggedness, and distance from the centroid of the grid-cell to Kinshasa. Sector fixed-effects (collectivités) included in all specifications. Robust standard errors in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

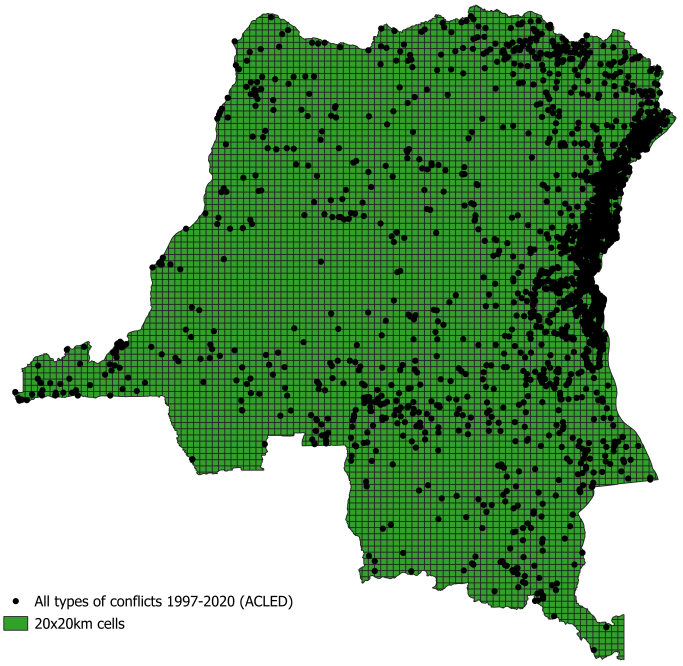
Table E3: Historical Missions and Contemporary Schools

	(1) Schools per 1000 people	(2) Proportion of Girls
Panel A: Primary Education		
Catholic mission	0.437***	0.0571***
	(0.0955)	(0.0106)
Protestant mission	0.179*	0.0336**
	(0.0920)	(0.0132)
Mean Y	0.592	0.283
Collectivite FE	Yes	Yes
Grid-cell controls	Yes	Yes
R-squared	0.183	0.388
N	5786	6009
Panel B: Secondary Education		
Catholic mission	0.208***	0.0653***
	(0.0523)	(0.00990)
Protestant mission	0.120**	0.0437***
	(0.0593)	(0.0110)
Mean Y	0.300	0.184
Collectivite FE	Yes	Yes
Grid-cell controls	Yes	Yes
R-squared	0.177	0.424
N	5786	5985

NOTE. Data: Universe of primary and secondary schools in the DRC in 2020, from Ministry of education. The sample is restricted to rural schools as defined by the Ministry of Education. The table reports OLS estimates. The unit of observation is a 20kmx20km grid cell. Outcome variables are defined as follows: schools per 1000 population in 2020 in column (1) and average share of girls in school by grid cell in column (2). Controls at the grid-cell level are always included: presence of navigable river, presence of colonial railroad, presence of precolonial explorer route, population density in 1900, cropland area in 1900, mean altitude, ruggedness, and distance from the centroid of the grid-cell to Kinshasa. Sector fixed-effects (collectivités) included in all specifications. The number of observations is lower is column 1 because grid-cells with very low population are excluded to obtain meaningful values. Robust standard errors in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

E.3. Post-colonial schools and conflict

Figure E2: Distribution of conflict events in the DRC, 1997-2022



NOTE.: Data comes from The Armed Conflict Location Event Data Project (ACLED). The figure shows the distribution of all type of conflicts in the DRC between 1997-2020. Events include battles, explosions/remote violence, protests, riots, strategic developments, and violence against civilians.

Table E4: Historical Missions and Contemporary Schools

	(1) Schools per 1000 people	(2) Proportion of Girls
Panel A: Primary Education		
Catholic mission	0.505***	0.0325***
	(0.132)	(0.0114)
Protestant mission	0.282**	0.0433***
	(0.119)	(0.0131)
Mean Y	0.640	0.282
Collectivite FE	Yes	Yes
Grid-cell controls	Yes	Yes
R-squared	0.229	0.429
N	5515	5739
Panel B: Secondary Education		
Catholic mission	0.289***	0.0408***
	(0.0846)	(0.0107)
Protestant mission	0.207**	0.0480***
	(0.0866)	(0.0109)
Mean Y	0.330	0.181
Collectivite FE	Yes	Yes
Grid-cell controls	Yes	Yes
R-squared	0.210	0.467
N	5515	5713

NOTE. Data: Universe of primary and secondary schools in the DRC in 2020, from Ministry of education. The sample is restricted to grid-cells with less than 10 conflicts. The table reports OLS estimates. The unit of observation is a 20kmx20km grid cell. Outcome variables are defined as follows: schools per 1000 population in 2015 in column (1) and average share of girls in school by grid cell in column (2). Controls at the grid-cell level are always included: presence of navigable river, presence of colonial railroad, presence of precolonial explorer route, population density in 1900, cropland area in 1900, mean altitude, ruggedness, and distance from the centroid of the grid-cell to Kinshasa. Sector fixed-effects (collectivités) included in all specifications. The number of observations is lower is column 1 because grid-cells with very low population are excluded to obtain meaningful values. Robust standard errors in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

Table E5: Competition between religious affiliations

	(1) Catholic school	(2) Protestant school	(3) Catholic school	(4) Protestant school
Catholic between t and t-3	-0.104***	0.0556***		
	(0.00987)	(0.0108)		
Protestant between t and t-3	0.0307***	-0.0518***		
	(0.00802)	(0.00948)		
Catholic between t and t-5			-0.106***	0.0598***
			(0.00923)	(0.0107)
Protestant between t and t-5			0.0331***	-0.0691***
			(0.00828)	(0.00960)
Time FEs	Yes	Yes	Yes	Yes
Grid-cell FEs	Yes	Yes	Yes	Yes
Mean Y	0.221	0.469	0.221	0.469
R-squared	0.199	0.243	0.200	0.245
N	17426	17426	17426	17426

NOTE. The dependent variable is a dummy that equals one if a new school of type x opens in cell c at year. The sample is restricted to grid-cells with less than 100 conflicts between 1997 and 2022. The independent variables are dummies that equal one if a new school of the opposite type opened in cell c in year t-1, t-2 or t-3, or in cell c in year t-1, t-2, t-3, t-4 or t-5, respectively. Standard errors () clustered at the 20x20km grid cell level. Robust standard errors in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

E.4. Competition between schools and religious affiliations

When the lagged dependent variable is included as a regressor, OLS estimates are inconsistent since they are correlated with the error term. Moreover, when T is small, the introduction of individual fixed effects does not solve the problem (Nickell, 1981). Here, we show that our results (two-way fixed effects model with large T) do not change when we implement the estimator developed by Arellano and Bond (1991), which is designed to avoid the problem raised by the inclusion of lagged dependent variables. To this end, we estimate the following first difference model:

$$\Delta y_{it} = \gamma_1 \Delta y_{i,t-1} + \gamma_2 \Delta y_{i,t-2} + \gamma_3 \Delta y_{i,t-3} + \beta_0 \Delta x_{i,t} + \beta_1 \Delta x_{i,t-1} + \beta_2 \Delta x_{i,t-2} + \beta_3 \Delta x_{i,t-3} + \Delta u_{i,t}$$

Where y_{it} is the opening of a school of type x in year t, and x_{it} is the opening of a school of the opposite type in year t. Because of practical reasons, we now separate the different lags in the right hand side of the equation. Table E6 shows evidence of strong substitution effects between schools of the same religion, but strong clustering dynamics between schools of different religion, suggesting competition between religious affiliations.

Table E6: Competition between religious affiliations

	2 to 10 lags used		2 to 5 lags used	
	Catholic school	Protestant school	Catholic school	Protestant school
	(1)	(2)	(3)	(4)
Catholic		0.0843***		0.121***
		(0.0292)		(0.0410)
Lag 1 Catholic	-0.149***	0.0139**	-0.237***	0.0216**
	(0.0221)	(0.00702)	(0.0297)	(0.00903)
Lag 2 Catholic	-0.00570	-0.00138	-0.00747	0.00259
	(0.00486)	(0.00655)	(0.00525)	(0.00777)
Lag 3 Catholic	-0.0155***	0.000587	-0.0185***	0.00323
	(0.00409)	(0.00554)	(0.00435)	(0.00636)
Protestant	0.0701***		0.0564**	
	(0.0180)		(0.0236)	
Lag 1 Protestant	0.00745	-0.169***	0.0148**	-0.237***
	(0.00453)	(0.0250)	(0.00590)	(0.0331)
Lag 2 Protestant	0.00194	-0.000436	0.00628	0.00362
	(0.00414)	(0.00558)	(0.00501)	(0.00677)
Lag 3 Protestant	0.00203	-0.0136***	0.00454	-0.0108**
	(0.00350)	(0.00468)	(0.00400)	(0.00529)
Time FEs	Yes	Yes	Yes	Yes
Mean Y	0.0526	0.118	0.0526	0.118
N	62218	62218	62218	62218

NOTE. The dependent variable is a dummy that equals one if a new school of type x opens in cell c at year t. The independent variables are dummies that equal one if a new school opens in year t (Catholic or Protestant), in year $t-1$ (lag 1), $t-2$ (lag 2), or $t-3$ (lag 3). Standard errors () clustered at the 20x20km grid cell level are reported in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

E.5. Competition across religious denominations: robustness

Table E7: Competition between religious affiliations, controlling for education trends

	(1) Catholic school	(2) Protestant school	(3) Catholic school	(4) Protestant school
Catholic between t and t-3	-0.0324***	0.0197***		
	(0.00319)	(0.00408)		
Protestant between t and t-3	0.00981***	-0.0110***		
	(0.00257)	(0.00370)		
Catholic between t and t-5			-0.0370***	0.0162***
			(0.00278)	(0.00354)
Protestant between t and t-5			0.00667***	-0.0224***
			(0.00238)	(0.00328)
Time FEs	Yes	Yes	Yes	Yes
Grid-cell FEs	Yes	Yes	Yes	Yes
Mean Y	0.0581	0.120	0.0581	0.120
R-squared	0.126	0.271	0.127	0.271
N	72132	72132	72132	72132

NOTE. The dependent variable is a dummy that equals one if a school of type x opens in cell c at year t. The independent variables are dummies that equal one if a new school of the opposite type opened in cell c in year t-1, t-2 or t-3, or in cell c in year t-1, t-2, t-3, t-4 or t-5, respectively. All specifications include the number of school openings in the province of cell c at year t. Standard errors () clustered at the 20x20km grid cell level and are reported in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

Table E8: Competition between religious affiliations, Conley standard errors

	(1) Catholic school	(2) Protestant school	(3) Catholic school	(4) Protestant school
Catholic between t and t-3	-0.0317***	0.0215***		
	(0.00492)	(0.00522)		
Protestant between t and t-3	0.0108***	-0.00843		
	(0.00370)	(0.00568)		
Catholic between t and t-5			-0.0366***	0.0172***
			(0.00440)	(0.00534)
Protestant between t and t-5			0.00690*	-0.0218***
			(0.00360)	(0.00532)
Time FEs	Yes	Yes	Yes	Yes
Grid-cell FEs	Yes	Yes	Yes	Yes
Mean Y	0.0580	0.118	0.0580	0.118
R-squared	0.00222	0.000602	0.00345	0.00116
N	72727	72727	72727	72727

NOTE. The dependent variable is a dummy that equals one if a new school of type x opens in cell c at year t. The independent variables are dummies that equal one if a new school of the opposite type opened in cell c in year t-1, t-2 or t-3, or in cell c in year t-1, t-2, t-3, t-4 or t-5, respectively. Standard errors () clustered at the 20x20km grid cell level. 100km Conley standard errors in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

E.6. Competition between religious schools: rural

To reduce further the concerns associated with confounding factors associated with urbanization dynamics (and therefore related to changes in the returns to education and structural transformation), we show here that our results do not change when we exclusively focus on rural areas. First, Table E9 restricts the sample to rural schools as defined by the Congolese Ministry of Education. Second, Table E10 restricts the sample to grid-cells with a population density in 2010 lower than 1300 (removing the top 10% of population density). In both cases, our results barely change, suggesting that economic factors alone are unlikely to explain the the positive coefficient of opening a school of a different denomination.

Table E9: Competition between religious affiliations: rural sample

	(1) Catholic school	(2) Protestant school	(3) Catholic school	(4) Protestant school
Catholic between t and t-3	-0.0372***	0.0119***		
	(0.00320)	(0.00434)		
Protestant between t and t-3	0.00655**	-0.0191***		
	(0.00268)	(0.00404)		
Catholic between t and t-5			-0.0410***	0.00958**
			(0.00288)	(0.00384)
Protestant between t and t-5			0.00367	-0.0298***
			(0.00246)	(0.00362)
Time FEs	Yes	Yes	Yes	Yes
Grid-cell FEs	Yes	Yes	Yes	Yes
Mean Y	0.0522	0.114	0.0522	0.114
R-squared	0.0744	0.155	0.0758	0.156
N	69554	69554	69554	69554

NOTE. The sample is restricted to rural or semi-rural schools. The dependent variable is a dummy that equals one if a new school of type x opens in cell c at year t. The independent variables are dummies that equal one if a new school of the opposite type opened in cell c in year t-1, t-2 or t-3, or in cell c in year t-1, t-2, t-3, t-4 or t-5, respectively. Standard errors () clustered at the 20x20km grid cell level and are reported in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

Table E10: Competition between religious affiliations: rural sample

	(1) Catholic school	(2) Protestant school	(3) Catholic school	(4) Protestant school
Catholic between t and t-3	-0.0314***	0.0204***		
	(0.00335)	(0.00450)		
Protestant between t and t-3	0.0106***	-0.00964**		
	(0.00272)	(0.00416)		
Catholic between t and t-5			-0.0355***	0.0167***
			(0.00294)	(0.00396)
Protestant between t and t-5			0.00650***	-0.0227***
			(0.00250)	(0.00368)
Time FEs	Yes	Yes	Yes	Yes
Grid-cell FEs	Yes	Yes	Yes	Yes
Mean Y	0.0573	0.121	0.0573	0.121
R-squared	0.0831	0.164	0.0841	0.165
N	68771	68771	68771	68771

NOTE. The sample is restricted to grid-cells with a population density in 2010 lower than 1300 (removing top 10%). The dependent variable is a dummy that equals one if a new school of type x opens in cell c at year t. The independent variables are dummies that equal one if a new school of the opposite type opened in cell c in year t-1, t-2 or t-3, or in cell c in year t-1, t-2, t-3, t-4 or t-5, respectively. Standard errors () clustered at the 20x20km grid cell level and are reported in parenthesis. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

E.7. Structural transformation and returns to education

Using the same methodology as the demographic survey of the 1970s, a household budget survey was conducted in parallel, in which 1/50 of the identified households were interviewed. This survey contains information on expenditures and economic transactions made by the household during the month of the interview, including information on education expenditures, which will allow us to explore additional questions related to the demand for education.

Table E11: Exposure to missions and occupational specialization

	(1) Work	(2) Work	(3) Agriculture	(4) Services	(5) Manufacture	(6) Log(nightlights)
Exposure to Catholic	-0.0181**	-0.0150**	-0.0563***	0.0193***	0.0121***	0.0260
	(0.00850)	(0.00761)	(0.0137)	(0.00660)	(0.00357)	(0.0257)
Exposure to Protestant	-0.0243***	-0.0196**	-0.00832	-0.00314	-0.00779**	0.0299
	(0.00817)	(0.00774)	(0.0139)	(0.00776)	(0.00340)	(0.0261)
FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	0.719	0.821	0.393	0.178	0.0781	0.652
R-squared	0.220	0.101	0.290	0.100	0.102	0.823
N	34208	25221	34208	34208	34208	675

NOTE. Data: Pooled 2007 and 2013 waves of the DHS. The sample is restricted to people living in DHS clusters located 50km or closer to a mission. The sample in column 2 is restricted to people over 20. The unit of observation in column 7 is the DHS cluster. The table reports OLS estimates. We use $-log(distance)$ to measure exposure. Outcome variables are defined as follows: "Employed" is a dummy variable equal to one if the respondent works (column 1) or if the respondent works and is older than 20 years old (column 2). Dummy equal to one if the respondents works in agriculture (column 3), services (column 4), or manufactures (column 5). The outcome in column 6 is a dummy variable that equals one if the respondent belongs to the top 40% of the wealth distribution. Column 7 reports the log(1+nightlights), where nightlights are defined as the average nighttime luminosity of the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location (removing outliers such as flares from petroleum extraction and other short-duration lights. The background data was also shifted to 0 to account for moonlight). Standard errors () are clustered at DHS cluster level. Geographical, historical and individual controls included. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

E.8. Distance to school and girls' education

Table E12: Distance to School and Enrollment rates

	Dependent variable: Out-of-school Child					
	Primary			Secondary		
	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
Primary school 1-2km	0.0400***	0.0424**	0.0369*			
	(0.0152)	(0.0170)	(0.0203)			
Primary school 3-5km	0.0901***	0.0822***	0.0944***			
	(0.0220)	(0.0245)	(0.0298)			
Primary school 5-10km	0.185***	0.242***	0.125***			
	(0.0348)	(0.0468)	(0.0402)			
Primary school >10km	0.477***	0.461***	0.490***			
	(0.0905)	(0.113)	(0.0699)			
Secondary school 1-2km				0.00731	-0.00210	0.0145
				(0.0128)	(0.0154)	(0.0178)
Secondary school 3-5km				0.0307	-0.0116	0.0740**
				(0.0211)	(0.0233)	(0.0296)
Secondary school 5-10km				0.0610***	0.0357	0.0842***
				(0.0226)	(0.0220)	(0.0301)
Secondary school >10km				0.290***	0.239***	0.345***
				(0.0440)	(0.0553)	(0.0470)
Mean Y	0.267	0.258	0.276	0.185	0.140	0.232
R-squared	0.145	0.147	0.145	0.0923	0.0588	0.119
N	12086	6170	5916	9030	4577	4453

NOTE. Data: Out-of-School Children and Adolescents Survey. The sample in Columns (1)-(3) is restricted to children in primary school age (6-11) while it is restricted to children in secondary school age (12-17) in Columns (4)-(6). The dependent variable is a dummy indicating whether the child attends school at the time of the survey. Standard errors () are clustered at the "grappe" level. Province fixed-effects are always included. Controls include age, whether the respondent lives in a rural/urban area and sex.

F. Persistence in demand for education

F.1. Correlation between parents' and children education

Table F1: Intergenerational correlation in education in the 1970s

	(1) Years educ	(2) Primary or more	(3) Post-primary	(4) Years educ	(5) Primary or more	(6) Post-primary
Father's y.o.e	0.0998***	0.0175***	0.0165***	0.0778***	0.0155***	0.0157***
	(0.0110)	(0.00161)	(0.00131)	(0.00590)	(0.00131)	(0.00165)
Mother's y.o.e	0.0452***	0.00928***	0.00917***	0.0327***	0.00654***	0.00677***
	(0.0103)	(0.00214)	(0.00188)	(0.0115)	(0.00185)	(0.00242)
Father's y.o.e x F				0.0457**	0.00422	0.00179
				(0.0210)	(0.00300)	(0.00285)
Mother's y.o.e x F				0.0268***	0.00580***	0.00506**
				(0.00738)	(0.00181)	(0.00206)
FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	6.080	0.636	0.439	6.080	0.636	0.439
R-squared	0.354	0.261	0.298	0.356	0.262	0.298
N	22458	22458	22458	22458	22458	22458

NOTE. Data: Demographic Survey conducted during the 1970s. The sample is restricted to children aged 14-19 during the time of the survey. In the explanatory variables, "*y.o.e*" account for "Years of education" to shorten their names. The table reports OLS estimates. Outcome variables are defined as follows: single years of education in columns (1) and (2), a binary variable that equals one if the respondent has at least one year of education in columns (2) and (5) and the probability of completing primary education (>5 years of education) in columns (3) and (6). Standard errors () are clustered at the territory of birth level. Fixed-effects include city of residence, territory of birth, age and the zone of residence of the individual within the city. Controls also include age of parents. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

Table F1 presents the results of regressing children's education on their parents' education, controlling for an extensive set of demographic and geographic characteristics. Overall, the relationship between parents' and children's education is strongly significant but small in size: an additional year of father's (mother's) education is associated with 0.1 (0.05) years of education for a child (column 1). These coefficients are small compared to estimates in the literature, even accounting for the fact that we absorb the "regional exposure effect" of parents. For instance, Hertz et al. (2008) report coefficients from simple regressions of children number of years of education on parents years of education for 42 countries and find coefficients of 0.80 for their African sample.[69]

[69]Other regional averages are: 0.79 for Latin America, 0.47 for Eastern European countries and 0.54 for West-European countries and North America.

F.2. *Exposure to missions and educational expenditures*

Table F2: Exposure to missions and education expenditures

	(1) Education expenditures	(2) Ln(1+Educ Expenditures)	(3) Expenditure per capita	(4) Clothes expenditure
Exposure to Catholic	-0.0382	-0.211	-54.95	0.00542
	(0.0399)	(0.162)	(98.16)	(0.0403)
Exposure to Nuns	-0.0140	-0.0508	75.52	0.00148
	(0.0205)	(0.104)	(75.75)	(0.0213)
Exposure to Protestant	0.0665**	0.278*	60.68	0.00627
	(0.0307)	(0.146)	(87.33)	(0.0265)
FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Mean Y	0.348	1.511	1311.9	0.264
R-squared	0.162	0.179	0.183	0.160
N	4731	4731	4731	4731

NOTE. Data: Budgetary Survey of the 1970s. The table reports OLS estimates. (Log) Exposure to missions is described in Section 3.2. Outcome variables are defined as follows: dummy variable that equals one if the household spends in education in column (1), 1 + the logarithm of total expenditure in education in column (2) , total expenditure per capita in column (3), and whether the household spends in clothes for children in column (4). Standard errors () are clustered at the territory of birth level. Fixed-effects include city of residence, territory of birth, and 5-year age cohorts. Controls include total number of household members, sex of household head, whether respondents were born in urban/rural area, and total expenditure. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

F.3. *Missions and the demand for education of uneducated parents.*

We examine whether the demand for education of uneducated parents is different when they grew up closer to a Protestant mission. Formally we estimate the following equation, for both sons and daughters separately:

$$y_{iatz} = \alpha + ParEduc_i^{'}\beta + ParExp_i^{'}\gamma + ParExp_i * ParEduc_i^{'}\pi + X_i^{'}\Phi + \theta_t + \sigma_a + \tau_z + \varepsilon_{iatz} \quad (5)$$

Where y_{iatz} is the educational outcome of child i of age a born in territory t and living in the zone z within her city of residence. $ParEduc$ is a vector containing two binary variables indicating whether the mother and the father have at least one year of education. Finally, $ParExp$ captures whether the father was born in a territory highly exposed to Catholic or to Protestant missions.[70] In our main specification, high exposure to Catholic (Protestant) mission is defined as a binary variable that equals one if the father was born in a territory where the distance to Catholic (Protestant) mission is above the median.[71] X_i is a vector of individual level controls (gender, year of installation in

[70]Results are similar (although of smaller size) if we use the exposure of the mother instead of the father. We prefer using father's exposure to missions because most mothers are uneducated.

[71]Below, we show that these results are robust to alternative thresholds (i.e., top 40% or top 30% of the

the current city of residence and the age of parents). θ_t, σ_a and τ_z are territory of birth, age and neighborhood (within the city) fixed effects, respectively.

The main coefficient of interest is that on exposure, γ indicating whether uneducated parents who were (highly) exposed to missions take different education decisions (for them $ParEduc = 0$ and thus $ParEduc * ParExp = 0$). Table F3 reports the results and confirms that uneducated parents' exposure to Protestant missions is associated with higher investment in the education of their children, even if, again sizes are modest. Children of uneducated parents have 0.55 more years of education (9.5% of the sample mean) when parents' exposure to Protestant missions is above the median.[72] Exposure to Catholic missions has no significant effect. Furthermore the effect of exposure to Protestant missions is stronger for daughters than for sons.

Table F3: Father's exposure to missions and children's education: heterogeneity between educated and uneducated fathers

	Daughters			Sons		
	Years educ	$\geq$ Primary	Post-primary	Years educ	$\geq$ Primary	Post-primary
	(1)	(2)	(3)	(4)	(5)	(6)
Father educated	0.833***	0.110***	0.0703***	0.479***	0.0843***	0.0678***
	(0.125)	(0.0272)	(0.0195)	(0.0938)	(0.0188)	(0.0182)
Mother educated	0.420***	0.0797***	0.0674***	0.208***	0.0437***	0.0403***
	(0.0466)	(0.0102)	(0.0100)	(0.0362)	(0.00881)	(0.00937)
High Exposure to Catholic	0.0135	-0.0153	-0.0210	-0.229	-0.0537	-0.0250
	(0.203)	(0.0293)	(0.0237)	(0.159)	(0.0330)	(0.0245)
High Exposure to Protestant	0.547***	0.0701**	0.0654***	0.369**	0.0788***	0.0441*
	(0.179)	(0.0301)	(0.0195)	(0.150)	(0.0302)	(0.0259)
HEC x Father Educ	0.0866	0.0152	0.0542**	0.219	0.0593*	0.0220
	(0.223)	(0.0338)	(0.0233)	(0.159)	(0.0328)	(0.0249)
HEP x Father Educ	-0.455**	-0.0465	-0.0517**	-0.288*	-0.0784**	-0.0294
	(0.190)	(0.0309)	(0.0200)	(0.150)	(0.0342)	(0.0276)
FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	5.769	0.590	0.391	6.347	0.673	0.480
R-squared	0.350	0.263	0.282	0.371	0.267	0.311
N	9853	9853	9853	10562	10562	10562

NOTE. Data: Demographic Survey of the 1970s. The sample is restricted to daughters aged 14-19 at the time of the survey in columns (1)-(3) and to sons aged 14-19 in columns (4)-(6). High exposure to mission of type X is a dummy variable that equals one if father's exposure to mission of type x is in the top 50% of the distribution. Outcome variables are: single years of education in column (1), the probability of completing primary education (>5 years of education) in column (2), and the probability of having post-primary education (>6 years of education) in column (3). Standard errors () are clustered at the father's territory of birth level. Fixed-effects include territory of birth, age and zone of residence. Controls include sex, year of installation in the current city and age of parents. *** for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$.

distribution) or to a continuous measure of exposure.

[72]Interestingly the coefficient on the interaction is of a similar magnitude and an opposite sign, suggesting that educated and uneducated parents had similar demand for education, when highly exposed to Protestant missions.

These results are robust to alternative specifications. First, Table F4 shows the same specification but uses instead mother's exposure to missionary presence. Second, Tables F5 and F6 measure high exposure to missions by using alternative thresholds: high exposure is a dummy variable that takes value one if the father was born in a territory belonging to the top 40% or to the top 30% of the distribution of exposure to missionary presence, respectively. Finally, Table F7 measures exposure to missionary presence by using the logarithmic distance, as reported in Section 4.4.

Table F4: Mother's exposure to missions and children's education: heterogeneity between educated and uneducated mothers

	Daughters			Sons		
	Years educ	$\geq$ Primary	Post-primary	Years educ	$\geq$ Primary	Post-primary
	(1)	(2)	(3)	(4)	(5)	(6)
Father educated	0.651***	0.0925***	0.0696***	0.454***	0.0771***	0.0652***
	(0.103)	(0.0161)	(0.0137)	(0.0621)	(0.0142)	(0.0132)
Mother educated	0.522***	0.0935***	0.0929***	0.271***	0.0511***	0.0455***
	(0.0637)	(0.0132)	(0.0166)	(0.0502)	(0.0132)	(0.0125)
High Exposure to Catholic	0.103	-0.00292	0.0274*	0.0251	0.00170	0.00425
	(0.0830)	(0.0156)	(0.0155)	(0.0656)	(0.0132)	(0.0159)
High Exposure to Protestant	0.190***	0.0318**	0.0158	0.182**	0.0202	0.0286*
	(0.0685)	(0.0142)	(0.0147)	(0.0740)	(0.0167)	(0.0173)
HEC x Father Educ	-0.140	-0.0324*	-0.0484**	-0.0353	-0.00762	0.00377
	(0.101)	(0.0182)	(0.0231)	(0.0840)	(0.0229)	(0.0221)
HEP x Father Educ	-0.0541	0.0151	0.0152	-0.102	-0.00665	-0.0158
	(0.0895)	(0.0198)	(0.0189)	(0.102)	(0.0266)	(0.0251)
FEs and controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	5.770	0.590	0.393	6.358	0.676	0.481
R-squared	0.348	0.262	0.284	0.368	0.265	0.311
N	9561	9561	9561	10138	10138	10138

Note. Data: Demographic Survey of the 1970s. The sample is restricted to daughters aged 14-19 at the time of the survey in columns (1)-(3) and to sons aged 14-19 in columns (4)-(6). High exposure to mission of type X is a dummy variable that equals one if mother's exposure to mission of type x is in the top 50% of the distribution. Outcome variables are: single years of education in column (1), the probability of completing primary education (>5 years of education) in column (2), and the probability of having post-primary education (>6 years of education) in column (3). Standard errors () are clustered at the mother's territory of birth level. Fixed-effects include territory of birth, age and zone of residence. Controls include sex, year of installation in the current city and age of parents.

Table F5: Father's exposure to missions and children's education: heterogeneity between educated and uneducated fathers (threshold at 40 %)

	Daughters			Sons		
	Years educ	≥ Primary	Post-primary	Years educ	≥ Primary	Post-primary
	(1)	(2)	(3)	(4)	(5)	(6)
Father educated	0.801***	0.108***	0.0680***	0.468***	0.0783***	0.0622***
	(0.116)	(0.0251)	(0.0187)	(0.0921)	(0.0191)	(0.0182)
Mother educated	0.419***	0.0794***	0.0672***	0.208***	0.0441***	0.0405***
	(0.0467)	(0.0102)	(0.00997)	(0.0360)	(0.00878)	(0.00933)
High Exposure to Catholic	0.0142	-0.0157	-0.0199	-0.228	-0.0506	-0.0250
	(0.204)	(0.0307)	(0.0247)	(0.145)	(0.0308)	(0.0227)
High Exposure to Protestant	0.551***	0.0737**	0.0613***	0.360**	0.0843***	0.0354
	(0.140)	(0.0290)	(0.0209)	(0.146)	(0.0288)	(0.0294)
HEC x Father Educated	0.0854	0.0197	0.0485**	0.238	0.0650**	0.0273
	(0.221)	(0.0336)	(0.0239)	(0.147)	(0.0316)	(0.0244)
HEP x Father Educated	-0.461***	-0.0568*	-0.0458**	-0.318**	-0.0812**	-0.0238
	(0.154)	(0.0295)	(0.0222)	(0.137)	(0.0330)	(0.0290)
FEs and controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	5.769	0.590	0.391	6.347	0.673	0.480
R-squared	0.350	0.263	0.282	0.371	0.267	0.311
N	9853	9853	9853	10562	10562	10562

Note. Data: Demographic Survey of the 1970s. The sample is restricted to daughters aged 14-19 at the time of the survey in columns (1)-(3) and to sons aged 14-19 in columns (4)-(6). High exposure to mission of type X is a dummy variable that equals one if father's exposure to mission of type x is in the top 40% of the distribution. Outcome variables are: single years of education in column (1), the probability of completing primary education (>5 years of education) in column (2), and the probability of having post-primary education (>6 years of education) in column (3). Standard errors () are clustered at the father's territory of birth level. Fixed-effects include territory of birth, age and zone of residence. Controls include sex, year of installation in the current city and age of parents.

Table F6: Father's exposure to missions and children's education: heterogeneity between educated and uneducated fathers (threshold at 30 %)

	Daughters			Sons		
	Years educ	≥ Primary	Post-primary	Years educ	≥ Primary	Post-primary
	(1)	(2)	(3)	(4)	(5)	(6)
Father educated	0.784***	0.110***	0.0737***	0.482***	0.0834***	0.0653***
	(0.105)	(0.0217)	(0.0180)	(0.0818)	(0.0180)	(0.0162)
Mother educated	0.415***	0.0791***	0.0666***	0.206***	0.0436***	0.0401***
	(0.0468)	(0.0102)	(0.00986)	(0.0361)	(0.00869)	(0.00930)
High Exposure to Catholic	-0.293	-0.0315	-0.0246	-0.166	-0.0241	-0.0214
	(0.321)	(0.0459)	(0.0356)	(0.128)	(0.0271)	(0.0287)
High Exposure to Protestant	0.838***	0.0982***	0.0642**	0.270**	0.0580**	0.0326
	(0.189)	(0.0279)	(0.0250)	(0.116)	(0.0239)	(0.0276)
HEC x Father Educated	0.402	0.0393	0.0506	0.180	0.0431	0.0308
	(0.339)	(0.0481)	(0.0353)	(0.141)	(0.0306)	(0.0323)
HEP x Father Educated	-0.768***	-0.0860***	-0.0528**	-0.251**	-0.0613**	-0.0276
	(0.203)	(0.0286)	(0.0253)	(0.111)	(0.0267)	(0.0305)
FEs and controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	5.769	0.590	0.391	6.347	0.673	0.480
R-squared	0.351	0.263	0.282	0.371	0.267	0.311
N	9853	9853	9853	10562	10562	10562

Note. Data: Demographic Survey of the 1970s. The sample is restricted to daughters aged 14-19 at the time of the survey in columns (1)-(3) and to sons aged 14-19 in columns (4)-(6). High exposure to mission of type X is a dummy variable that equals one if father's exposure to mission of type x is in the top 30% of the distribution. Outcome variables are: single years of education in column (1), the probability of completing primary education (>5 years of education) in column (2), and the probability of having post-primary education (>6 years of education) in column (3). Standard errors () are clustered at the father's territory of birth level. Fixed-effects include territory of birth, age and zone of residence. Controls include sex, year of installation in the current city and age of parents.

Table F7: Father's exposure to mission and children's education: heterogeneity between educated and uneducated fathers (logarithmic distance)

	Daughters			Sons		
	Years educ	≥ Primary	Post-primary	Years educ	≥ Primary	Post-primary
	(1)	(2)	(3)	(4)	(5)	(6)
Father educated	-0.0483	-0.00134	0.102*	0.253	0.0378	0.0377
	(0.494)	(0.0749)	(0.0608)	(0.369)	(0.0823)	(0.0929)
Mother educated	0.416***	0.0793***	0.0667***	0.208***	0.0437***	0.0402***
	(0.0463)	(0.0101)	(0.00992)	(0.0359)	(0.00874)	(0.00929)
Father's Exposure to Cath	-0.196	-0.0343*	-0.0194	-0.0983	-0.0237	-0.00740
	(0.159)	(0.0190)	(0.0176)	(0.103)	(0.0171)	(0.0182)
Father's Exposure to Protestant	0.507***	0.0734***	0.0505***	0.200*	0.0436***	0.0236
	(0.129)	(0.0193)	(0.0164)	(0.107)	(0.0165)	(0.0225)
Exposure to Cath x Educated	0.295	0.0364	0.0488***	0.112	0.0352*	0.0116
	(0.191)	(0.0224)	(0.0181)	(0.0956)	(0.0182)	(0.0190)
Exposure to Prot x Educcated	-0.456***	-0.0584***	-0.0365**	-0.154	-0.0423**	-0.0176
	(0.145)	(0.0198)	(0.0163)	(0.102)	(0.0178)	(0.0228)
FEs and controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	5.769	0.590	0.391	6.347	0.673	0.480
R-squared	0.351	0.263	0.283	0.371	0.267	0.311
N	9853	9853	9853	10562	10562	10562

Note. Data: Demographic Survey of the 1970s. The sample is restricted to daughters aged 14-19 at the time of the survey in columns (1)-(3) and to sons aged 14-19 in columns (4)-(6). Exposure to missionary presence is measured as $-log(distance)$. Outcome variables are: single years of education in column (1), the probability of completing primary education (>5 years of education) in column (2), and the probability of having post-primary education (>6 years of education) in column (3). Standard errors () are clustered at the father's territory of birth level. Fixed-effects include territory of birth, age and zone of residence. Controls include sex, year of installation in the current city and age of parents.

F.4. Intergenerational correlation and father's occupation

Table F8: Intergenerational correlation in education in the 1970s

	(1)	(2)	(3)	(4)	(5)	(6)
	Years educ	Primary or more	Post-primary	Years educ	Primary or more	Post-primary
Father's y.o.e	0.0976***	0.0169***	0.0163***	0.0754***	0.0149***	0.0154***
	(0.0105)	(0.00149)	(0.00120)	(0.00553)	(0.00128)	(0.00167)
Mother's y.o.e	0.0452***	0.00924***	0.00911***	0.0325***	0.00646***	0.00662***
	(0.0103)	(0.00215)	(0.00188)	(0.0115)	(0.00189)	(0.00244)
Father's y.o.e x F				0.0459**	0.00425	0.00177
				(0.0210)	(0.00304)	(0.00286)
Mother's y.o.e x F				0.0271***	0.00589***	0.00523**
				(0.00746)	(0.00179)	(0.00207)
FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y	6.080	0.636	0.439	6.080	0.636	0.439
R-squared	0.355	0.262	0.299	0.357	0.263	0.299
N	22458	22458	22458	22458	22458	22458

Note. Data: Demographic Survey conducted during the 1970s. The sample is restricted to children aged 14-19. "*y.o.e*" account for "Years of education". Outcome variables are: single years of education in columns (1) and (2), a binary variable that equals one if the respondent has at least one year of education in columns (2) and (5) and the probability of completing primary education (>5 years of education) in columns (3) and (6). Standard errors () are clustered at the territory of birth level. Fixed-effects include city of residence, territory of birth, age, zone of residence of the individual within the city, and father's sector of activity. These sectors are 10: 1) Agriculture, silviculture, hunting and fishing; 2) Extractive industries; 3) Agricultural processing industries; 4) Other processing industries and rubber; 5) Building and civil engineering; 6) Electricity, gaz, water and health services; 7) Financial institutions, insurance, real state; 8) Transport, warehouses and communications; 9) Services; 10) Trade. Controls also include age of parents.