PROJECTING THE IMPACT OF COVID-19 ON EDUCATION AND INTERGENERATIONAL MOBILITY IN SUB-SAHARAN AFRICA

Guido Neidhöfer, Nora Lustig and Patricio Larroulet

COMMITMENT TO EQUITY

CEQ INSTITUTE
COMMITMENT TO EQUITY
Tulane University

Working Paper 132
August 2023

The CEQ Working Paper Series

PROJECTING THE IMPACT OF COVID-19 ON EDUCATION AND INTERGENERATIONAL MOBILITY IN SUB-SAHARAN AFRICA

Guido Neidhöfer[*], *Nora Lustig, and Patricio Larroulet*

CEQ Working Paper 132

AUGUST 2023

ABSTRACT

Using microsimulations, we project the effect of instructional losses caused by COVID-19 on secondary school completion rates and intergenerational mobility of education in eight countries in Sub-Saharan Africa. On average, secondary school completion rates could decrease by 12 percentage points overall and by 16 points for children with low-educated parents. Interestingly, in most countries the gender gap diminishes because, for men, the projected decrease in secondary school completion is higher. A small additional impact on girls' education due to the potential rise in teenage pregnancy is observed in some countries. Intergenerational mobility of education in the eight Sub-Saharan countries in our sample is expected to decrease, on average, by 10%.

JEL Codes: I24, I38, J62

Keywords: COVID-19, lockdowns, human capital, school closures, intergenerational persistence, education, inequality, Africa

---

[*] Corresponding author: Guido Neidhöfer, Senior Researcher at ZEW – Leibniz Centre for European Economic Research, Mannheim (guido.neidhoefer@zew.de).

Acknowledgements: Paper prepared for the project "Measuring Fiscal Equity in the post-COVID- 19 World." The authors are grateful to the Bill & Melinda Gates Foundation for its generous support. Guido Neidhöfer acknowledges funding by the Volkswagen Foundation within the grant call "Corona Crisis and Beyond – Perspectives for Science, Scholarship and Society".
This paper was prepared as part of the Commitment to Equity Institute's country-cases research program and benefitted from the generous support of the Bill & Melinda Gates Foundation. For more details, click here www.ceqinstitute.org.

Projecting the Impact of COVID-19 on Education and Intergenerational Mobility in Sub-Saharan Africa

Guido Neidhöfer* *ZEW Mannheim*

Nora Lustig *Tulane University*

Patricio Larroulet *Central Bank of Argentina*

This version: August 2023

Abstract

Using microsimulations, we project the effect of instructional losses caused by COVID-19 on secondary school completion rates and intergenerational mobility of education in eight countries in Sub-Saharan Africa. On average, secondary school completion rates could decrease by 12 percentage points overall and by 16 points for children with low-educated parents. Interestingly, in most countries the gender gap diminishes because, for men, the projected decrease in secondary school completion is higher. A small additional impact on girls' education due to the potential rise in teenage pregnancy is observed in some countries. Intergenerational mobility of education in the eight Sub-Saharan countries in our sample is expected to decrease, on average, by 10%.

Keywords: COVID-19, lockdowns, human capital, school closures, intergenerational persistence, education, inequality, Africa. *JEL Codes:* I24, I38, J62.

----

* Corresponding author: Guido Neidhöfer, Senior Researcher at ZEW – Leibniz Centre for European Economic Research, Mannheim (guido.neidhoefer@zew.de).

Acknowledgements: Paper prepared for the project "Measuring Fiscal Equity in the post-COVID-19 World." The authors are grateful to the Bill & Melinda Gates Foundation for its generous support. Guido Neidhöfer acknowledges funding by the Volkswagen Foundation within the grant call "Corona Crisis and Beyond – Perspectives for Science, Scholarship and Society".

1. Introduction

In Sub-Saharan Africa (SSA), the COVID-19 pandemic resulted in a significant contraction of economic activity. After an unprecedented decline of GDP per capita of, on average, 4.5% in 2020, the region recovered in 2021 and is expected to expand again in 2022 and 2023 (IMF, 2022; World Bank, 2022). However, these figures do not consider the persistence of the impact of the pandemic on some dimensions of human development and their potential long-run consequences. One of these dimensions is the effect of the pandemic on education.

Like almost everywhere in the world, SSA schools closed their doors to mitigate the spread of the COVID-19 virus. In most countries, schools remained closed over long periods in 2020 and 2021, adding up to a loss of more than 50% of instruction time (UNESCO, 2022). Available evidence for other parts of the world suggests large negative effects of school closures on schooling achievements, particularly among children from disadvantaged background (see e.g. the reviews by Betthäuser et al., 2023; Hammerstein et al., 2021; Werner and Woessmann, 2021; Zierer, 2021). However, as highlighted by Betthäuser et al. (2023), among others, the lack of research on this is particularly significant in low-income countries.

The scant existing evidence for SSA countries confirms the negative effect on education, as revealed by the simulation of associated learning losses (Angrist et al., 2021; Azevedo et al., 2021), comparing data on reading pre- and post-school closures (Ardington et al., 2021), or reports on learning activity (or lack thereof) of children during school closures collected through household-level data from telephone-surveys (Dang et al, 2021). The evidence suggests that the effect on education has been different for children depending on their socioeconomic background. The long-term effects of the pandemic on education inequality could be strong enough to reduce intergenerational mobility and produce higher labor income inequality in the future.

To the best of our knowledge, this is the first study analyzing the long-run impact of the pandemic on educational achievements and intergenerational mobility for SSA. Adopting the microsimulation framework proposed by Neidhöfer et al. (2021), we nowcast the potential effect of instructional losses on educational achievements and project their future implications for secondary school completion rates and intergenerational educational mobility.[1] Our analysis

[1] Nowcasting is a term that combines the words "now" and "forecasting". Essentially, it is the prediction of current outcomes using information that is published early and often at a higher frequency than the target variable of interest. This term is primarily used in macroeconomics, where the information on current macroeconomic outcomes is available with a certain delay. Economists attempt to estimate the current state of these outcomes using data that is available earlier. The main goal is to provide an early estimate of the current status of the outcomes. In our paper, we aim to assess the long-term effects of the recent COVID-19 pandemic on education. Through a model of educational learning losses, using individual level data, we simulate the educational losses of cohorts affected by the COVID-19

focuses on eight Sub-Saharan African countries: Ethiopia, Ghana, Kenya, Liberia, Malawi, Nigeria, South Africa, and Tanzania. We perform the simulations for the country as a whole, by parental educational background, by gender, and for rural and urban areas. We explain our approach in detail in the nontechnical summary and the presentation of the formal model in the main body of the article.

To estimate the education supply shock, we collect data on the duration of school closures and mitigation policies implemented by governments to support learning from home during periods of school closures. In addition, we estimate the distribution of access to the internet in each country. We begin by nowcasting the potential consequences of instructional losses due to school closures on educational attainments by using pre-pandemic nationally representative household surveys that include information on educational attainment of individuals and their parents' educational background. Then, we project the effect on secondary school completion rates and intergenerational mobility. When estimating the impact by gender, we also consider the potential increase in teenage pregnancy rates caused by the pandemic and its effect on school attendance. Finally, we simulate the capacity of policies to mitigate the impact of school closures on education through the provision of offline and online remote learning.

Our main results suggest that the pandemic will have a significant and persistent negative effect on educational inequality and its intergenerational persistence in SSA. On average, among the eight countries in our sample, secondary school completion rates decrease by 12 percentage points overall and by 16 points for children with low-educated parents. Interestingly, in most countries, the gender gap diminishes because the projected decrease in secondary school completion is higher for men. However, a small additional effect on girls' education due to the COVID-19 induced rise in teenage pregnancy is estimated in some countries. Intergenerational mobility of education in the eight Sub-Saharan countries in our sample is expected to decrease, on average, by 10%. These findings indicate the importance of remedial actions that should be taken to prevent the likely and unprecedented decrease in human capital that could potentially offset years of favorable development in creating opportunities. Our microsimulations must inevitably rely on a number of assumptions which are described in detail in Section 2.1.2. To assess their validity, we either test how changing them would affect the results in a series of robustness checks or explain how altering them is likely to impact the effects. Although order of magnitudes vary in some cases, the results are consistent, even if assumptions are relaxed or changed.

---

shock. We started this project in early 2021, so our estimates are based on the most recent (pre-pandemic) data available at that time. That is why we refer to our estimates as a "nowcast" of the learning losses occurred during the pandemic.

The remainder of the paper is structured as follows: Section 2 presents the methodology adopted to estimate the impact of the pandemic on intergenerational mobility and describes the data. Section 3 reports our results on the impact of COVID-19 on education in SSA while highlighting heterogeneities by parental background and gender, presenting estimates on the additional effect of increasing teenage pregnancy due to the pandemic on education. Furthermore, the mitigation effect of policies is evaluated. Section 7 concludes.

2. Methodology and Data

2.1 Simulation of the impact of COVID-19 on educational achievements

2.1.1 Non-technical summary

As indicated, we follow the simulation methodology proposed by Neidhöfer et al. (2021) to nowcast the impact of instructional losses on educational achievements and estimate the resulting impact on secondary school completion rates and intergenerational mobility of education through a counterfactual exercise.[2] As done elsewhere in the literature, we assume that there is a human capital production function where the production factors are schooling, through in-person classes and remotely, and the family (e.g. Hanushek and Woessmann, 2012).[3] The approach builds on measuring the amount of instructional time lost due to the closure of educational facilities (see e.g. Abadzi, 2009; Adda, 2016) and considers asymmetries in the response of countries and the families' capabilities to cushion instructional disruptions. The main component driving the shock to human capital is the loss in instructional time suffered by children due to school lockdowns during the pandemic. Families may hereby partly or perfectly substitute formal schooling depending on the highest educational degree attained by parents. For example, parents with tertiary education are assumed to be able to perfectly substitute in-person schooling. An additional channel causing instructional losses is the likelihood of infection and death within the family, approximated by the relative number of COVID-19 cases and deaths per inhabitant.

Figure 1 shows a schematic representation of the functioning of our model. In a regular school year, parental inputs and school inputs contribute to children's education. The pandemic shock brings in a number of additional factors. Due to the closure of educational facilities, in-person school inputs are only available when schools are open. However, during school closures education

[2] Strictly speaking, we are estimating the intergenerational persistence of human capital.

[3] A further production factor are innate abilities. While, we would not have the necessary data to incorporate them in the analysis, we expect innate abilities not to be affected by the pandemic. Hence, we focus on the role played by the two other factors.

systems provide remote online and offline learning tools. While we assume offline resources are available for all children, access to online resources depends on the existing internet infrastructure and its distribution. If governments (or other education providers) can replace in-person with remote learning in full, there would not be any instructional losses. However, if nothing else, internet access is not universal so it is quite unlikely that such a scenario could materialize. Another source of compensation is parental involvement in remote learning. We assume that the extent of compensation through this channel depends on the parents' socioeconomic status, measured by their level of education. This is how children's background affects instructional losses differentially. Finally, other shocks due to COVID-19 (health, pregnancy, income losses) can affect education negatively because of its impact on demand for schooling.

Given the limited time that has elapsed since the pandemic, we cannot use actual data to econometrically identify the impact of the pandemic on intergenerational mobility. As an alternative, to estimate the potential future impact we construct a counterfactual scenario. This involves simulating the shock using groups of individuals who had completed their education shortly before the pandemic (see the explanation of the data in Section 2.3.1). The underlying assumption is that these individuals are similar enough to those affected by the pandemic for the simulation to predict effects fairly accurately (see more details under item (3) right below).

2.1.2 Assumptions

The procedure we use is necessarily based on a series of assumptions. We will either explain how altering these assumptions would impact the estimates or test them in a robustness check. To ensure that we do not overestimate the instructional loss when choosing between different assumptions, we always choose the one that is more likely to result in downward bias rather than upward bias in our estimates. Here, we will list our procedure's assumptions and potential limitations, discuss each of these, and explain why the results are consistent when assumptions are relaxed. In later sections, we will discuss them in more detail with reference to the performed robustness checks.

(1) The human capital production function adopted in our model includes schooling and parental inputs. Parental inputs are closely linked to parental resources (e.g. Becker and Tomes, 1979, 1986), which in our analysis are proxied by parental education (see also Black and Devereux, 2011). In line with theories of intergenerational mobility (e.g. Becker et al., 2018), we assume that any factors that enhance human capital investments, such as the quantity and quality of monetary and non-monetary investments made in children's education (including school quality, tutoring, extra-curricular activities etc.), are positively correlated with parental resources.

One limitation of using parental education as a measure of a child's learning during school closures is that it does not consider other factors that may influence a parent's ability to teach their child effectively. These factors can include the parent's teaching skills and personality, as well as their profession. For example, parents who are teachers themselves or who have patient personalities may be more effective at substituting for school closures. Our analysis does not consider these potential differences within parental education groups. Considering them should not change the results obtained for the impact of the pandemic across socioeconomic groups because, if anything, the ability to minimize instructional losses for children with higher socioeconomic status would be even greater since parents with better teaching abilities are likely to have more formal education.

(2) The main measure of school inputs used in the model is days of instruction (i.e., days in which schools are open), with a measure of remote learning calculated to estimate the reduction in instructional losses during school closures. In our main specification, the combination of remote learning through offline and online tools can potentially substitute a day of in-person classes fully. Giving lower weights to remote learning days with respect to in-class instruction would increase the instructional losses and yield estimates that show an even stronger impact on educational attainments and mobility. So, our results are closer to a lower bound.

(3) In order to simulate a counterfactual, the simulations of instructional loss and the resulting changes in intergenerational mobility are conducted on a sample of individuals who are older than 18 and have completed their education. Hereby, we restrict the sample to those born between 1987 and 1994, who are the youngest possible respondents fulfilling the chosen age restriction. Because they are rather close in time to those in school during the COVID-19 pandemic, we expect that the composition of our sampled individuals, in terms of factors such as gender and parental education, does not differ substantially and that potential differences do not affect the estimates. In other words, we treat our sampled individuals as a laboratory by assuming that they are the best available proxy to simulate the impact of the pandemic shock on individuals that are currently in school. In the Appendix, we relax this assumption by reweighting the observations in our sample such that the composition by parental background resembles the composition of the cohorts in school during 2020. The results of this additional exercise are consistent with our main estimates.

(4) The instructional loss is not cumulative, meaning that students do not forget what they have learned before the school closures and that their education is only affected during the pandemic years and not thereafter. As shown in one of the robustness checks included in the Appendix, considering cumulative learning losses increases the negative impact of COVID-19 on education.

(5) A day of instructional loss always counts the same regardless of the age and grade. This is in line with the findings of the meta-analysis by Betthäuser et al. (2023), which found that the effects of school closures due to COVID-19 do not vary significantly across different ages and grades.

(6) The days schools remained closed and the extensiveness of mitigation measures vary across countries but not within countries (no differences across regions). While we do not have information on school closures disaggregated by subnational regions, we can at least observe the number of weeks that schools were fully closed in the whole country as opposed to being partially closed, i.e. closed in only in some parts of the country or with shorter days of instruction. In a robustness check, we only consider days of countrywide school closures as days of lost instruction while we treat the days of partial school closures as if regular schooling would have taken place. The estimates—albeit lower—does not significantly differ from the main results.

(7) We consider the quantity of (in-class and remote) education but not its quality. If one assumes that the quantity and quality of human capital investments are positively correlated across households (see e.g. Gruijters and Behrman, 2020), estimates considering the differences in the quality of remote instruction would yield an even stronger inequality in instructional losses across advantaged and disadvantaged children.[4]

(8) We incorporate the effect of health shocks on instructional losses based on existing research on COVID-19 and the impact of health-related events on education. However, in our model the contribution of health shocks to the overall impact on instructional losses is relatively low compared to other factors, such as school closures and remote learning. Therefore, modifying the numbers associated with health shocks would have minimal impact on the overall estimates.

(9) The threshold for completing secondary education is set at 12 full years of education, which is the typical length of study in the countries analyzed and is associated with a completed secondary degree. Individuals whose counterfactual education falls below this threshold after the simulation are classified as individuals with incomplete secondary education. In this case, one could argue that our estimates are more an upper than a lower bound given that we treat a loss of 1 day identically to the loss of a full school year. In robustness checks, we relax this assumption by allowing children to be unaffected by small instructional losses: e.g., if less than

[4] The quality of education can also vary considerably across countries. As shown by Filmer et al. (2020) Learning-adjusted years of schooling, which consider the quality of education, rather than just its quantity, are substantially lower than aggregate measures of years of schooling. This is particularly true in developing countries. While this is surely a very interesting aspect, since in our analysis the measure of interest is the within-country difference between the education in a regular year and in a counterfactual year characterized by the pandemic, considering a factor that varies at the cross country level and would affect similarly both sides of the equation (such as a factor converting years of education into LAYS) would not change our measures of interest.

25% and 50% percent of one regular school year, we assume that the instructional loss was zero. However, because in most cases instruction days lost over the years 2020 and 2021 were above these thresholds, the impact on high-school completion rates in all but one country is substantially the same.

(10) The baseline exercise does not consider additional effects of the pandemic on the externalities associated with schooling, such as nutrition, mental health, and non-cognitive skills. If these effects were taken into account, the impact of the pandemic on intergenerational mobility and inequality in educational achievements would likely be stronger.

(11) Due to the lack of data to directly measure the impact of the pandemic on intergenerational mobility in the African countries in our sample it is not possible to directly validate the predictions of our model. Despite these limitations, we attempt to validate our results by, firstly, contextualizing our estimates within the current literature on the impact of the pandemic on education and, secondly, using alternative data sources. Appendix C in the Supplemental Material shows a comparative analysis of the parameters used in our model with these additional sources of information. Comparisons with indicators such as the percentage of households involved in learning recovery activities, duration of school closures, and types of recovery actions suggest qualitative consistency of real-time data with our model and the estimates.

2.1.3 Model formalization

The core of the simulation exercise by Neidhöfer et al. (2021) is to estimate the instructional loss of cohorts who are currently in school by constructing a counterfactual human capital equivalent for each individual in the sample, which encompasses individuals belonging to cohorts that completed their schooling prior to the onset of the pandemic.[5] This counterfactual assessment aims to determine the level of education an individual would have attained if they had experienced the pandemic while still enrolled in their country of residence. The counterfactual education $\hat{e}$ is defined as:

[5] Monroy-Gomez-Franco et al. (2021) extend the methodology proposed by Neidhöfer et al. (2021) to, first, include parental income as further component of the model and, second, consider the additional impact of instructional losses in the short-run on cumulative learning losses. For the first extension, information on parental income is required, while for the second extension data on test scores (or a comparable measure of learning) is needed. Since we do not have this information available, we consider the original version of the methodology. Due to the usually high correlation between parental education and earnings, we do not expect that applying the first extension would change our results significantly. Applying the second extension could indeed lead to estimates of even higher learning losses. Hence, also for this reason our results should be considered as lower bound estimates of the persistent learning losses caused by the COVID-19 pandemic in Sub-Saharan Africa. In the Supplemental Material, we provide simple estimates of cumulative educational losses deriving from our estimates.

$$\widehat{e_{ijc}} = e_{ijc} - \kappa_{ijc}. \tag{1}$$

Here, e is the actual reported years of education of individual i, with parental educational background j, living in country c.[6] κ is the human capital loss, measured as the share of instructional time lost during the pandemic:

$$\kappa_{ijc} = \frac{(t_c - M_{ijc} + \tau_{ic}) \cdot (1 - p_j)}{T_c}. \tag{2}$$

This human capital loss is defined by several components: i) the instructional loss due solely to school closures (t_c); ii) the days of instructional loss compensated by government policies geared to support not-in-person schooling (M_{ijc}); iii) the additional instructional loss associated with health shocks suffered by households (τ_{ic}); iv) the parental factor of substitution (p_j); v) the total days of school in a regular school year in the country (T_c). Since we consider the years 2020 and 2021, T_c is the sum of regular days of schooling in two years.

The single components of equation (2) are defined in the following way:

The instructional loss due solely to school closures (t_c) captures the number of days that schools were closed, or partly closed, due to the pandemic. Hereby, days in which schools were partly closed count as half a day of instructional loss and days of full closures as entire days of instructional loss.[7] We obtain the data on the total duration of school closures from UNESCO (see Section 2.4.2 and Appendix A in the Supplemental Material).

The days of instructional loss compensated by government policies supporting not-in-person schooling (M_{ijc}) can be calculated using the following formula:

$$M_{ijc} = \delta_1 t_c f_c + \delta_2 t_c n_c \cdot P(A_{ijc} = 1). \tag{3}$$

This term specifies that the days schools were closed (t_c) can be compensated by remote learning through offline resources (such as TV, radio, cellphone, or printed copies)—in the first term—and online resources—in the second term. f_c and n_c are indices that range from zero to one, constructed to measure the alternative supply of education with offline and online resources in each country, respectively. We derive these two measures from the UNESCO-Survey of National

[6] To avoid measurement error and harmonize the measure of education across countries, following most of the literature we do not use the actually reported years of schooling, but impute the regular years of education necessary to obtain the attained educational level. Hence, incomplete primary education is equivalent to three years of schooling, complete primary to five, incomplete secondary to eight, complete secondary to 12, and more than secondary education to 15. We assign years of parental education following the same procedure.

[7] We relax this assumption and estimate a lower bound by considering only days of nationwide school closures (i.e. full days of school closures). This robustness check provides estimates which are not significantly different from the main estimates (see Appendix B7 in the Supplemental Material).

Education Responses to COVID-19 School Closures (see Section 2.4.2 and Appendix A in the Supplemental Material). δ_1 and δ_2 are weights that define the relative efficiency of these tools. In our baseline specification, in order to obtain a lower-bound estimate, both weights have the value of 0.5 (i.e., the combination of offline and online resources can potentially substitute a day of in-person classes fully). Relaxing this assumption would yield even stronger estimates of the instructional loss.[8]

Online learning is, however, only available if the household is connected to the internet. Hence, we interact the term capturing the mitigation through online learning with the likelihood that the individual lived in a household with access to the internet in his or her childhood. We approximate this likelihood by the probability that a household with the education of the household head j in country c has access to the internet $(A_{ijc} = 1)$, which we estimate using household survey data for each country in the sample (see Section 2.4.2 and Appendix A in the Supplemental Material).

The additional instructional loss associated with health shocks suffered by households (τ_{ic}) is obtained by computing:

$$\tau_{ic} = \tau^q \cdot P(q_{ic} = 1) + \tau^d \cdot P(d_{ic} = 1). \tag{4}$$

This calculation considers the human capital loss due to infection ($P(q_{ic} = 1)$) of one of the household members with COVID-19 and death ($P(d_{ic} = 1)$) due to the latter. To estimate the likelihood of this happening, the number of infections and deaths per inhabitant in the country is multiplied by the average country-level household size. τ^q and τ^d are the number of days of schooling lost due to the time the child cannot dedicate to learning in case he/she or someone in the household is infected, sick, or in the case of the death of a family member.[9] Following Neidhöfer et al. (2021), we set τ^q to the average days of symptom duration—around one week (i.e. five days of school)—and τ^d to a three-week loss of instructional time (i.e. 15 days).

These three components contributing to the loss and recovery of instructional time during the pandemic can be compensated to a certain degree by parental inputs, captured by the term $1 - p_j$ in equation (2). Here, p_j is the parental factor of substitution. It is defined as:

$$p_j = \frac{e_j^p}{\max\left(e_j^p\right)}, \tag{5}$$

[8] Appendix B8 in the Supplemental Material shows the main estimates and all alternative estimates obtained from the robustness checks.

[9] This assumption follows the literature on the impact of household health shocks on education (e.g. Gertler et al., 2004).

where e_j^p are the completed years of schooling of parents with education level j and $\max(e_j^p)$ the maximum level that can be attained. Consequently, $1 - p_j$ is zero for children with high-educated parents that may fully substitute the instructional losses (i.e. $\kappa = 0\ \forall t, M, \tau$), and one for children of illiterate parents who depend entirely on the provision of remote learning by the education system. For other levels of parental education, the value of $1 - p_j$ lies within this interval. Generally, $1 - p_j$ represents various factors that can influence a child's education—such as the ability of the parents to help with learning material, the value placed on education, and the resources available to support education, like parental time and technology—and is aimed to capture the higher capabilities of parents with higher levels of education to support their children's education when schools are closed, as evidenced in several empirical studies (e.g. Engzell et al., 2021; Jaume and Willen, 2019; Maldonando and De Witte, 2020). These capabilities are typically more prevalent in parents with higher levels of education and socio-economic status. Hence, κ_{ijc} determines the impact of lost instructional time due to school closures on a child's education, taking into account the different risks to households of different socio-economic statuses during the COVID-19 pandemic.

For the simulation of changes in intergenerational mobility, the instructional losses resulting from the equations above are attributed randomly to the share of the sample with parental education j that mirrors the likelihood that the household in which the individual grew up has the characteristics displayed above (likelihood to have access to the internet, likelihood of death and infection). The parameter $1 - p_j$ instead can be understood as a proportional instructional loss experienced by all children of parents with a j level of education or as a certain share of children of parents with a j level of education who suffer the entire instructional loss, while the rest are unaffected. In the first scenario, $1 - p_j$ is the degree to which parents with educational level j can substitute schooling; in this case, for the simulation, the shock is distributed to the degree $(t_c - M_{ijc} + \tau_{ic}) \cdot (1 - p_j)$ evenly to all individuals with the respective parental educational background. In the second scenario, $1 - p_j$ is the likelihood of parents with educational level j to substitute schooling perfectly; in this scenario, a shock of the amount of $(t_c - M_{ijc} + \tau_{ic})$ is attributed to a randomly selected share $1 - p_j$ of the population within the group of individuals with parental education j. We report the results of the simulations for the second scenario in the main body of the text and the first scenario in Appendix B of the Supplemental Material.

2.2 Intergenerational mobility estimates

Once obtained the counterfactual post-pandemic education of individuals following the procedure explained above, we estimate two standard measures for the intergenerational mobility of education: the intergenerational slope coefficient (a measure of relative intergenerational persistence) and the probability of (absolute) upward mobility (see e.g. Neidhöfer et al., 2018). The first indicator is obtained by regressing children's education on parents' education. [10] The second indicator is obtained by estimating the likelihood of individuals with low-educated parents to complete secondary schooling.

In the post-pandemic counterfactual, secondary school completion changes for individuals with a completed secondary degree (and not more) whose counterfactual post-pandemic education lies under the threshold of 12 years of education.[11] We also estimate the same likelihood for children of high-educated parents to provide a benchmark for comparison. Following the literature on educational mobility in African countries, we define low-educated parents as those who did not complete primary schooling and those that completed at least primary schooling as "high-educated" parents (Alesina et al., 2021). [12] In order to simulate and quantify the potential impact of the pandemic on intergenerational mobility, we estimate the two indicators using the actual years of education of individuals. Then, we re-estimate them with the counterfactual education. Finally, we measure the difference between each indicator's resulting measures.

[10] Other measures used to estimate intergenerational mobility, besides of the slope coefficient, are the intergenerational correlation and the slope of a rank-rank regression (see e.g. Asher at al., *forthcoming*). Ahsan et al. (2022) discuss the various measures used to estimate intergenerational educational mobility and how they can lead to conflicting conclusions. It suggests that the intergenerational rank-rank slope (IRRS) estimates can provide dramatically different results than the intergenerational regression coefficient (IGRC) and intergenerational correlation (IGC) estimates. The study presents evidence that the idiosyncratic component of children's schooling variance unrelated to family background plays an important role in IGC, and that IGC estimates are less responsive to changes in economic forces compared to IGRC estimates. The study concludes by suggesting that policy advice should focus on the effects of policies on the influence of inherited circumstances on children's education, as captured by IGRC, even if policies fail to affect IRRS significantly. Although it goes beyond the aim of our study to establish causal effects, we believe that these findings are relevant for our analysis and focus therefore our main results on the IGRC (slope coefficient). Simulation results of the predicted impact of the pandemic on the IGC and IRRS are included in Appendix B10 in the Supplemental Material.

[11] In robustness checks, we relax this assumption by allowing children to be unaffected by small instructional losses (e.g., if less than 25% and 50% percent of one regular school year). However, due to the strong instructional losses, the impact on high-school completion rates in all but one country is substantially the same. Results are included in Appendix B12 of the Supplemental Material.

[12] Results using a different threshold to define low-educated and high-educated parents (completed secondary schooling) are included in Appendix B of the Supplemental Material.

2.3 Data

2.3.1 Household surveys

To simulate the impact of the pandemic on education, our primary source of individual data for each country is nationally representative household surveys that include information on respondents' education level. To avoid co-habitation bias in our intergenerational mobility estimates, which arises from incomplete information about children who have left their household of origin, we exclusively utilize surveys that include retrospective questions on parental education (see Emran et al., 2015). We use each country's most updated household survey that includes these characteristics. In order to prevent bias due to truncation, we limit the sample to individuals aged 19 years or older. This is because we want to make sure that all individuals in the sample have completed at least secondary education, which typically takes 12 years of schooling. By restricting the sample to respondents aged 19 or older, we can ensure that all individuals in the sample have had sufficient time to complete their secondary education. Furthermore, in order to simulate the impact for a cohort of individuals whose average level of education and educational mobility is as close as possible to those who are in the education system in 2020 and 2021, we restrict the sample to the youngest available cohort of adult respondents (i.e. people born between 1987 and 1994).[13] Our final sample comprises 26,147 individuals, ranging from 1,565 individual observations for Kenya to 5,209 for Ghana. Table 1 shows descriptive statistics for the samples used for each country.

To make our results comparable across countries, we adjust the variables we use to account for differences in the number of years corresponding to conventional levels of education (primary, secondary, etc.). Each country has a unique system with different lengths of primary and secondary education and distinct definitions of vocational training and post-secondary education. To account for this diversity, we use the ISCED scale provided by UNESCO to standardize education levels across countries and estimate the number of years of schooling for each level. This allows us to consistently define the following education categories for all countries: incomplete primary or no schooling, complete primary, incomplete secondary, complete secondary, at least some tertiary education. The imputed number of years of schooling for each category are 0, 5, 8, 12, and 15, respectively. We use the same categorization for children and parents. Our measure for parental education is always the education of the parent with the highest degree in the household. If one of the two parents has missing information, we use the information that is available. For more

[13] As mentioned, the results are consistent when reweighting the observations to resemble the composition by parental background of cohorts in school during 2020. See Appendix B13.

information on the household survey data and the education classification used in each country, see Appendix A in the Supplemental Material.

2.3.2 Country-level data on school closures and other characteristics

To compute the single measures for t, M, and τ described in Section 2, we complement the household survey data with country-level data on school closures in 2020 and 2021, information on educational mitigation strategies, and epidemiological parameters on the number of COVID-19 cases and deaths per inhabitant. Data on school closures is retrieved from UNESCO. Information on educational mitigation strategies is retrieved from the UNESCO-Survey of National Education Responses to COVID-19 School Closures. Data on COVID-19 cases and deaths is retrieved from the website *Our World in Data*. As mentioned, we multiply this last two indicators by the average number of people living in the same household in each country, retrieved from data by the United Nations, Department of Economic and Social Affairs, Population Division. Furthermore, we estimate the distribution of internet access by the level of education of the household head – our measure for A in equation (3) – using household survey data for each country in the sample. Table 2 displays all variables used in the simulations. Appendix A in the Supplemental Material describes the country-level data and its sources in more detail.

3. Results

3.1 The impact of COVID-19 on education

3.1.1 Instructional losses, educational inequality and intergenerational persistence

Applying the methodology explained in Section 2, we obtain estimates for the impact of the pandemic on educational achievements and intergenerational mobility. Table 1 provides a first estimate of the inequality in instructional time losses during the pandemic. Considering the days of school closures in 2020 and 2021 and the educational mitigation measures, the application of our methodology suggests that disadvantaged children (children whose parents did not complete primary education) could have lost between 31 and 118 days of instruction, while their peers with a more favorable parental educational background (children whose parents completed at least primary education) only between 10 and 33. The country with the strongest inequality in instructional loss is South Africa, where we observe a difference of 85 days of instruction between the two groups.

The estimated instructional time losses imply detrimental repercussions on the predicted change in the likelihood of children to complete secondary schooling. Figure 2 shows the predicted change in secondary completion. As can be seen, the share of individuals with complete secondary education decreases substantially due to the pandemic. On average, across the eight Sub-Saharan countries in our sample, secondary school completion rates decrease by 12 percentage points.

Figure 3 shows the potential heterogeneous impact of the pandemic on the likelihood of completing secondary education of children with low and high-educated parents. Children with high-educated parents are affected, but the shock hits strongly the already relatively low likelihood to complete secondary education of children with low-educated parents in Sub-Saharan countries. In all countries, this likelihood decreases substantially. On average, across all countries, the decline equals 16 percentage points. In most countries, the chances of secondary school completion of disadvantaged children reach a level even lower than ten percent, i.e. less than one of ten children with low-educated parents leaves education with at least a secondary school degree. In Malawi, the likelihood even decreases to less than one percent. As mentioned before, this effect is driven by children at the margin, namely those who completed secondary education in the pre-COVID but did not continue with tertiary education afterwards. Since this condition mainly applies to children with low-educated parents, our estimates for impact of the pandemic are particularly strong at the lower bottom of the distribution and suggest that inequality in educational achievements should increase.

Based on our projections, the intergenerational mobility of education is expected to fall in all countries under analysis. Conventionally, intergenerational mobility is measured by the "persistence" coefficient, which measures the partial correlation between the years of schooling of parents and children. The larger (smaller) the coefficient, the lower (higher) is mobility. Figure 4 shows the slope coefficient of intergenerational persistence measured pre-COVID and the post-pandemic counterfactual. Since this measure considers the entire distribution of years of education, and not just a marginal threshold as captured by the likelihood of secondary school completion discussed above, the magnitude of the effect is small. However, relative to the size of the slope coefficient, the effect ranges between a one percent increase in persistence and an increase by almost 50 percent.

3.1.2 Impact by gender

We estimate the effect separately for men and women and evaluate the impact of the pandemic on gender inequality in educational attainments. We here mainly focus on the overall results. The results for children of low and high-educated parents separately are available in Appendix B in the Supplemental Material.[14]

Figure 5 shows that in all countries except Kenya and South Africa, the likelihood of male children belonging to the 1987-1994 cohort completing secondary education is substantially higher than their female peers. In Kenya, completion rates are similar among males and females, while in South Africa, females have a slightly higher completion rate than males. Our estimates suggest that secondary completion rates of male children may decrease more in absolute terms than those of female children.

Figure 6 shows the resulting impact of the gender gap on the likelihood of completing secondary schooling. While in most countries, the gender gap diminishes because of the higher decrease in secondary school completion rates among men, in Kenya, the gender gap stays at a constant level but changes the pattern: before the pandemic, completion rates in Kenya were higher for male than for female, while after the pandemic this trend is reversed. In Ghana, the gender gap stays virtually the same as before.

The explanation for these differential effects by gender depends on the different quantity of boys and girls who completed secondary education in the pre-COVID scenario but did not continue with tertiary education afterwards. In most countries, our estimates suggest that boys are more likely to complete secondary education. However, among those who complete it, the likelihood of continuing to tertiary education after completing secondary is higher for females. This explains why our estimation of the potential impact of the pandemic on the fundamental threshold of secondary school completion is lower for female than for male students.[15]

3.2 Additional effect of increases in teenage pregnancy on girls' education

In this part of the analysis, we estimate the additional impact on secondary school completion of a further potential consequence of the pandemic, which affects the well-being of young girls and their educational achievements: the rise in teenage pregnancy. Teenage pregnancy reduces the probability of receiving a high school diploma and enrolling in tertiary education while increasing the likelihood of leaving school without a qualification (e.g. Fergusson and Woodward, 2000; Fletcher and Wolfe, 2009). Confinement and deprivation during lockdowns are believed to

[14] We also include separate results for rural and urban areas in Appendix B of the Supplemental Material. Generally, the projection of the decrease in completion rates and upward mobility is similar across areas.

[15] While this analysis highlights the potential impact of the pandemic on children's education, it is important to note that parental financial and non-financial investments may differ by gender, leading to potential gender bias against girls in developing countries (e.g. Pasqua, 2005). The discussion of different results by children's gender emphasizes the need to consider these differences in future research. If parents invest less in girls, the impact of school closures on girls' education may be even stronger.

dramatically worsen the situation regarding child abuse and teenage pregnancy, especially among vulnerable families. Indeed, media reports, statements by local NGOs, and reports of international organizations in several African countries provide anecdotal evidence that anticipates an increase in adolescent birth rates during the COVID-19 pandemic (UNICEF, 2020). Recent reports confirm this picture. In Malawi, for instance, an 11% increase in teenage pregnancies was recorded from January to August 2020 compared to the same period in 2019 (UNFPA, 2021). Scientific studies document a dramatic rise in the risk of young girls becoming pregnant, for instance, in Kenya (Zulaika et al., 2022), Nigeria (Musa et al., 2021) and other African countries (see the review by Willie, 2021).

To take into account the effect of teenage pregnancy on education, we extend the exercise described in Section 2 to account for this additional shock affecting the human capital formation of girls during the pandemic. We do so by including an additional component in the model that simulates the impact of an increase in the likelihood of young girls becoming pregnant on secondary school completion rates in each country.

Formally, the counterfactual post-pandemic education becomes:

$$\widetilde{e_{ijc}} = \widehat{e_{ijc}} - P(\zeta_{ijc} = 1) \cdot Z. \tag{6}$$

$\widehat{e_{ijc}}$ are the counterfactual years of schooling defined in equation (1). $P(\zeta_{ijc} = 1)$ is the likelihood of a girl with parental background *j* in country *c* to drop out from school due to pregnancy during the COVID-19 pandemic. To account for socioeconomic differences in the probability of this event to occur and its consequences for the educational career of girls, $P(\zeta_{ijc} = 1)$ is obtained by multiplying the percentage point increase in teenage pregnancy in the country due to the pandemic, with one minus the parental factor of substitution, defined in equation (5).[16] Z quantifies the consequences of pregnancy on education, which we set to a loss equal to the entire amount of two years of education.

To estimate the parameters of the model, we collect data on the increase in adolescent birth rates for each country. However, due to underreporting during the pandemic, birth registry data from low-income countries remains incomplete. Therefore, no clear conclusions can be drawn about how COVID-19 affected births, in general, and teenage pregnancy rates (UNFPA, 2021). Hence,

[16] Hence, in this simulation the increase in the risk of getting pregnant is set to be equal for girls of all parental backgrounds, namely equal to the average increase in the adolescent birth rate in the country of residence. What varies by parental background is the risk that this pregnancy leads to drop out from education, which is obtained by multiplying with one minus the parental factor of substitution. This means that the likelihood to drop out from school in case of pregnancy of girls whose parents are highly educated is lower than the likelihood of girls whose parents have low education.

we rely on estimates of the increase in teenage pregnancy from different reports (see a more detailed description in Appendix A in the Supplemental Material). Among these, we choose the worst-case scenarios, which indicate an increase in teenage pregnancy due to the pandemic of about 75%.[17] Based on this figure for the potential increase and on data on adolescent birth rates before the pandemic, we project the percentage point increase in adolescent birth rates for each country. Then, we simulate the impact of this additional factor, besides school closures, on the education of girls. Finally, we estimate the resulting impact on secondary school completion rates, similar to the estimations performed before.

Figure 7 shows the results. The first bar shows, for each country, the estimated likelihood of secondary school completion of girls born between 1987 and 1994, while the second bar shows the post-pandemic counterfactual; both are already shown in Figure 5. The third bar shows the post-pandemic counterfactual, considering the estimated increase in teenage pregnancy rates. It turns out that the potential increase in teenage pregnancy contributes marginally to a further increase in educational drop-out. In three countries, the impact on the decline in secondary school completion rates is around one percentage point, while in the other countries, it is lower than one percentage point. The main effect of the pandemic on education is, hence, confirmed to be driven by school closures. Although this may come as a surprise to some, our result is in line with studies that found that for the majority of young women, pregnancy occurred after dropping out from school rather than the opposite (Fergusson and Woodward, 2000).

In conclusion, concentrating resources on keeping girls in school could also counteract teenage pregnancy as part of the post-COVID remedial actions. A caveat is in order, however. The pandemic created unique circumstances. First, school closures could be seen as equivalent to "forced dropouts," especially for girls of low socioeconomic backgrounds. In that case, the findings by Ferguson and Woodward (2000) would apply. Furthermore, lockdowns exacerbated the circumstances within the household, which can lead to teenage pregnancy, given that members were forced to stay at home for lengthy periods.[18]

---

[17] Estimates based on this worst case scenario can be considered an upper bound. Indeed, in this part of the analysis we are interested in understanding the maximum contribution of teenage pregnancy on the top of the effect of school closures. Estimates based on the other scenarios show an even lower impact of teenage pregnancy and are available upon request.

[18] A further factor affecting children's education during the pandemic are family income losses and child labor. In Appendix B9 we provide estimates for two countries, Ethiopia and Ghana, for which we could find estimates from other studies on the share of population expected to suffer income losses (Geda, 2021; Issahaku and Abu, 2020). The results suggest that, while the main detrimental impact on education is driven by school closures, income losses can have an additional negative effect on intergenerational mobility. If child labour is a direct consequence of household income losses, these results are indicative of the detrimental effect of child labour on education.

3.3 Decomposition of effects and simulation of policy alternatives

The projections shown in the previous sections refer to the bundled effect of school closures (adding the potential impact of COVID-related health issues) and the mitigating impact of remote learning policies. The results suggest significant differences in how the pandemic affected intergenerational mobility across countries. These differences can be partially explained by the length of school closures, the quality of remote learning, and the distribution of access to the internet across countries (see Table 2). However, the patterns in the association with these factors across countries and the projected impact on mobility are not completely clear. For instance, while South Africa has the strongest increase in intergenerational persistence, the number of days of school closures is not among the highest, remote learning efforts were around average, and it is among the countries with the highest average access to the internet. South Africa is also the country where the 1987-1994 cohort has the highest pre-COVID level of intergenerational mobility. This shows that the distribution of education in the children's and parents' generation in each country also plays an important role in the projected effect of the pandemic on intergenerational mobility. With this in mind, we now evaluate the potential of educational mitigation policies to reduce instructional losses and temper the decrease in intergenerational mobility. First, we decompose the impact of the enacted policies vis-à-vis the effect that school closures would have had without any mitigation measure. Then, we simulate different scenarios, either improving the policies or the infrastructures that interact with the effectiveness of these policies.

Figure 8 ranks the countries in our sample by the estimated decrease in intergenerational mobility, measured by the slope coefficient, taking into account the closure of schools and the enacted offline and online educational mitigation strategies. The figure also shows the projected decrease in mobility due to school closures that would have occurred without mitigation measures. As can be seen clearly in the graph, although the effect of mitigation measures is sizable in most countries, it is not sufficient to close the gap caused by the closure of schools.

As a next step, we simulate which combination of measures and infrastructural improvements would allow to cushion the negative effects of the pandemic on intergenerational mobility. We successively change the parameters for online learning and internet coverage, keeping the offline level at the current level, and measure the impact it would have on the difference between the pre-pandemic level of intergenerational mobility and the post-pandemic counterfactual. Figure 9 shows the results of this policy exercise. The first bar in the graph shows the baseline situation, namely the decrease in intergenerational mobility, given the current distribution of internet coverage and online learning tools provided by the education systems. The change is always displayed in points of the partial correlation between parents' and children's education (i.e. the slope coefficient). The second bar shows the estimates obtained with improved online learning – i.e. setting the index for online learning tools, i.e. n in equation (3), to one – while keeping constant the distribution of internet access in the population. In the third and fourth bar, instead, results are obtained by improving internet coverage while keeping the current value of online learning constant: first, internet coverage is improved in such a way that each individual in the population has twice the likelihood of having access to the internet; then, universal internet coverage is granted, i.e. the likelihood of internet access is set to 100% for all individuals in the sample. The last bar shows the results of granting both full internet access and improved online learning.

The analysis shows that in all countries, given the current distribution of internet access, an improvement of the provided online learning resources would have no sizeable effect on reducing the negative effect of the pandemic on intergenerational mobility. At the same time, improving internet access alone, even granting universal internet access, would not be enough to close the gap. An unrealistically strong and costly effort by states would have been necessary, both improving online learning tools and the current infrastructure, to fully mitigate the impact of the closure of schools on educational disruptions by offering online remote learning.[19]

4. Conclusions

Using microsimulation and the framework proposed by Neidhöfer et al. (2021), we estimated the potential impact of the pandemic-related instructional losses on educational achievements and intergenerational mobility in eight Sub-Saharan African countries. We focused on the asymmetric effects of school closures on the education of children with different socioeconomic backgrounds and analyzed the potential mitigating impact of policies. Our findings show that the pandemic is likely to significantly negatively affect schooling and secondary school completion rates. Intergenerational mobility of education is bound to be lower as well. Educational inequality is expected to rise because the effect of school closures is stronger for disadvantaged children, who are at a higher risk of dropping out of the education system without completing a secondary school degree. On average, our simulations show that intergenerational mobility of education in the eight Sub-Saharan countries in our sample is expected to decrease by 10%, while the likelihood of children from low-educated families completing secondary education could decrease, on average, even by 16 percentage points. For several SSA countries, the likelihood of children from

[19] The negative sign of some coefficients, i.e. a higher level of intergenerational mobility in the counterfactual with respect to the pre-pandemic scenario, derives from the fact that with universal internet coverage children from low-educated families benefit most from the mitigation measures, while the likelihood of infection and death in the family may affect individuals over the entire distribution. However, the size of the difference shows that this equalizing effect would be negligible.

disadvantaged families completing secondary education may even become lower than ten percent. This means that in these countries, less than one of every ten children with low-educated parents affected by the COVID-19 crisis may leave education with a secondary schooling degree.

This alarming picture mirrors the projections of Neidhöfer et al. (2021) for Latin America, which found some confirmation in current analyses of education drop-outs using household survey data (Bracco et al., 2022) and administrative data (Lichand et al., 2021). Studies on the effects of the COVID-19 related school closures on student achievements in other parts of the world also validate our findings. As shown by the reviews by Hammerstein et al. (2021) and Blanden et al. (2022), the vast majority of studies—carried out, for example, in Australia, Belgium, China, Germany, Netherlands, Switzerland, and the US—found negative effects in mathematics, reading and other subjects of students in both elementary and secondary school. The average negative effect size found in these studies ranges between 0.09 and 0.29 standard deviations associated with 7-8 weeks of school closures, which is comparable in size with the effects found in past studies on the detrimental impact of summer vacations on learning (e.g. Alexander et al., 2007; Kuhfeld et al., 2020; Kuhfeld et al., 2022). A recent meta-analysis by Betthäuser et al. (2023) reports an overall effect size equivalent to a loss of approximately 35% of a full school year of learning. The estimates of the learning loss deriving from our simulations are somewhat lower—our estimations show that in the country with the highest impact, an instructional loss due to school closures of around 40 weeks is equivalent to around 20% of a school year of learning—but generally consistent with these patterns, and in line with our intention to provide lower bounds rather than overestimating the potential effect. Furthermore, the studies described above mostly analyzed learning losses during spring 2020, at the beginning of the pandemic, when institutions and families were unprepared for remote learning, while our simulations pertain to the entire years 2020 and 2021. The reported studies also confirm that learning losses differed strongly among children depending on the socioeconomic status of their family: while the impact on students from disadvantaged families was significant, students from families with high socioeconomic status were hardly affected.

The few existing studies measuring, rather than simulating, the short-term impact of the COVID-19 pandemic on learning in Africa with current data are also in line with our estimates. Dessy et al. (2021) found that in Nigeria, 7% of students did not return to school when they reopened. As shown by Kidman et al. (2022) in Malawi this share is even 14%. Ardington et al. (2021) found that in three South African provinces, due to the COVID-19 related disruptions in 2020, students in grade 2 had reading losses between 57% and 70% of a year of learning, and students in grade 4 had losses between 62% and 81%. Dang et al. (2022) shows by analysing household survey data for Burkina Faso, Ethiopia, Malawi, Mali, Nigeria, Tanzania, and Uganda that the pandemic reduced children's learning activities in all countries and that the amount of learning activities is positively associated with parents' level of education and household income. Our analysis adds a longer-term perspective to this picture of the effects of the COVID-19 pandemic on education in South Saharan Africa by showing the potential future impact on intergenerational mobility. Altogether, these findings highlight the severity of the resulting educational crisis and the importance of remedial actions that should be taken as quickly as possible to temper the impact on learning and educational losses, especially for the more disadvantaged children.

Declarations

*Ethical Approval: not applicable

*Competing interests: not applicable

*Authors' contributions: All authors contributed equally.

*Funding: Paper prepared for the project "Measuring Fiscal Equity in the post-COVID-19 World." The authors are grateful to the Bill & Melinda Gates Foundation for its generous support. Guido Neidhöfer acknowledges funding by the Volkswagen Foundation within the grant call "Corona Crisis and Beyond – Perspectives for Science, Scholarship and Society".

*Availability of data and materials: Datasets and codes are available upon request.

References

Abadzi, H. (2009). Instructional time loss in developing countries: Concepts, measurement, and implications. The World Bank Research Observer, 24(2), 267-290.

Adda, J. (2016). Economic activity and the spread of viral diseases: Evidence from high frequency data. The Quarterly Journal of Economics, 131(2), 891-941.

Ahsan, M. N., Emran, M. S., Jiang, H., Murphy, O., & Shilpi, F. (2022). When Measures Conflict: Towards a Better Understanding of Intergenerational Educational Mobility. Available at SSRN 4231514.

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2021). Intergenerational mobility in Africa. Econometrica, 89(1), 1-35.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. American sociological review, 72(2), 167-180.

Angrist, N., de Barros, A., Bhula, R., Chakera, S., Cummiskey, C., DeStefano, J., ... & Stern, J. (2021). Building back better to avert a learning catastrophe: Estimating learning loss from COVID-19 school shutdowns in Africa and facilitating short-term and long-term learning recovery. International Journal of Educational Development, 84, 102397.

Ardington, C., Wills, G., & Kotze, J. (2021). COVID-19 learning losses: Early grade reading in South Africa. International Journal of Educational Development, 86, 102480.

Asher, S., Novosad, P., & Rafkin, C. (forthcoming). Intergenerational mobility in India: Estimates from new methods and administrative data. American Economic Journal: Applied Economics.

Azevedo, J. P., Hasan, A., Goldemberg, D., Geven, K., & Iqbal, S. A. (2021). Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates. The World Bank Research Observer, 36(1), 1-40.

Becker, G. S., & Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. Journal of political Economy, 87(6), 1153-1189.

Becker, G. S., & Tomes, N. (1986). Human capital and the rise and fall of families. Journal of labor economics, 4(3, Part 2), S1-S39.

Becker, G. S., Kominers, S. D., Murphy, K. M., & Spenkuch, J. L. (2018). A theory of intergenerational mobility. Journal of Political Economy, 126(S1), S7-S25.

Betthäuser, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. Nature Human Behaviour, 1-11.

Black, S. E., & Devereux, P. J. (2011). Chapter 16 - Recent Developments in Intergenerational Mobility (D. Card & O. Ashenfelter, Eds.). In Handbook of Labor Economics (pp. 1487–1541).

Blanden, J., Doepke, M., & Stuhler, J. (2023). Chapter 6 - Educational inequality (E. A. Hanushek, S. Machin, & L. Woessmann, Eds.). In (pp. 405–497).

Bracco, J., Ciaschi, M., Gasparini, L., Marchionni, M., & Neidhöfer, G. (2022). The Impact of COVID-19 on Education in Latin America: Long-Run Implications for Poverty and Inequality. World Bank Policy Research Working Paper 10259.

Dang, Hai-Anh; Oseni, Gbemisola; Zezza, Alberto; Abanokova, Kseniya. (2021). Learning Inequalities during COVID-19: Evidence from Longitudinal Surveys from Sub-Saharan Africa. IZA DP no. 15684.

Dessy, S., Gninafon, H., Tiberti, L., & Tiberti, M. (2021). COVID-19 and children's school resilience: evidence from Nigeria (No. 952). GLO Discussion Paper.

Emran, M. S., Greene, W., & Shilpi, F. (2018). When measure matters coresidency, truncation bias, and intergenerational mobility in developing countries. Journal of Human Resources, 53(3), 589-607.

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. Proceedings of the National Academy of Sciences, 118(17).

Fergusson, D. M., & Woodward, L. J. (2000). Teenage pregnancy and female educational underachievement: A prospective study of a New Zealand birth cohort. Journal of Marriage and Family, 62(1), 147-161.

Filmer, D., Rogers, H., Angrist, N., & Sabarwal, S. (2020). Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education. Economics of Education Review, 77, 101971.

Fletcher, J. M., & Wolfe, B. L. (2009). Education and labor market consequences of teenage childbearing evidence using the timing of pregnancy outcomes and community fixed effects. Journal of Human Resources, 44(2), 303-325.

Geda, A. (2021). The macroeconomic and social impact of COVID-19 in Ethiopia in the global context. UNCTAD, Addis Ababa, 78.

Gertler, P., Levine, D. I., & Ames, M. (2004). Schooling and parental death. Review of Economics and Statistics, 86(1), 211-225.

Gruijters, R.J., & Behrman, J.A. (2020). Learning Inequality in Francophone Africa: School Quality and the Educational Achievement of Rich and Poor Children. Sociology of Education, 93, 256 - 276.

Hammerstein, S., König, C., Dreisörner, T., & Frey, A. (2021). Effects of COVID-19-Related School Closures on Student Achievement-A Systematic Review. Frontiers in Psychology, 4020.

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. Journal of economic growth, 17, 267-321.

IMF (2022). Regional economic outlook. Sub-Saharan Africa. Washington DC: International Monetary Fund (IMF).

Issahaku, H., & Abu, B. M. (2020). COVID-19 in Ghana: Consequences for poverty, and fiscal implications. AERC Working Paper.

Jaume, D., & Willén, A. (2019). The long-run effects of teacher strikes: evidence from Argentina. Journal of Labor Economics, 37(4), 1097-1139.

Kidman, R., Breton, E., Behrman, J., & Kohler, H. P. (2022). Returning to school after COVID-19 closures: Who is missing in Malawi?. International Journal of Educational Development, 93, 102645.

Kubitschek, W. N., Hallinan, M. T., Arnett, S. M., & Galipeau, K. S. (2005). High school schedule changes and the effect of lost instructional time on achievement. The High School Journal, 89(1), 63-71.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impact of COVID-19 school closures on academic achievement. Educational Researcher, 49(8), 549-565.

Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test score patterns across three COVID-19-impacted school years. Educational Researcher, 51(7), 500-506.

Lichand, G., Dória, C. A., Neto, O. L., & Cossi, J. (2021). The impacts of remote learning in secondary education: Evidence from brazil during the pandemic. Inter-American Development Bank Technical Note N. IDB-TN-02214.

Maldonado, J. E., & De Witte, K. (2022). The effect of school closures on standardised student test outcomes. British Educational Research Journal, 48(1), 49-94.

Michael Mncedisi Willie. (2021). Teenage Pregnancy During a Pandemic. International Journal of Women's Health Care, 6(3), 218-219.

Monroy-Gómez-Franco, L., Vélez-Grajales, R., & López-Calva, L. F. (2022). The potential effects of the COVID-19 pandemic on learnings. International journal of educational development, 102581.

Musa, S. S., Odey, G. O., Musa, M. K., Alhaj, S. M., Sunday, B. A., Muhammad, S. M., & Lucero-Prisno, D. E. (2021). Early marriage and teenage pregnancy: The unspoken consequences of COVID-19 pandemic in Nigeria. Public Health in Practice, 2, 100152.

Neidhöfer, G., Lustig, N., & Tommasi, M. (2021). Intergenerational transmission of lockdown consequences: prognosis of the longer-run persistence of COVID-19 in Latin America. The Journal of Economic Inequality, 19(3), 571-598.

Neidhöfer, G., Serrano, J., & Gasparini, L. (2018). Educational inequality and intergenerational mobility in Latin America: A new database. Journal of Development Economics, 134, 329-349.

Pasqua, S. (2005). Gender bias in parental investments in children's education: A theoretical analysis. Review of Economics of the Household, 3, 291-314.

Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: a decennial review of the global literature. Education Economics, 26(5), 445-458.

The Economist (2022). Tracking covid-19 excess deaths across countries. https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker

UNESCO (2022). Global monitoring of school closures.

UNFPA (2021). How will the COVID-19 pandemic affect births? Technical Brief.

UNICEF (2020). COVID-19: A Catastrophe for Children in Sub-Saharan Africa. Technical Report.

Werner, K., & Woessmann, L. (2021). The legacy of covid-19 in education. CESifo Working Paper No. 9358.

World Bank (2022). Global Economic Prospects: Sub-Saharan Africa.

Zierer, K. (2021). Effects of pandemic-related school closures on pupils' performance and learning in selected countries: A rapid review. Education Sciences, 11(6), 252.

Zulaika, G., Bulbarelli, M., Nyothach, E., van Eijk, A., Mason, L., Fwaya, E., ... & Phillips-Howard, P. A. (2022). Impact of COVID-19 lockdowns on adolescent pregnancy and school dropout among secondary schoolgirls in Kenya. BMJ global health, 7(1), e007666.

Tables and Figures

Table 1 – Data and Samples

Country	Survey	Year	Sample			
			Average age	Average years of education	Average parental years of education	N
Ethiopia	Living standard measurement study	2018	27.29	4.22	1.31	3868
Ghana	Living Standards Survey	2017	26.59	9.15	6.22	5209
Kenya	STEP Skills Measurement Household Survey	2013	22.59	9.58	7.25	1565
Liberia	Household Income and Expenditure Survey	2016	25.66	5.98	5.40	2841
Malawi	integrated household panel survey	2019	28.11	3.70	2.13	4721
Nigeria	General Household Survey	2018	27.69	6.97	4.29	1984
South Africa	National Income Dynamics Study	2018	27.62	11.14	6.67	2117
Tanzania	Household Budgetary Survey	2017	26.57	5.64	4.00	3842

Table 2 – Country level data

	t	*T*	*f*	*n*	*Aj=1*	*Aj=2*	*Aj=3*	*Aj=4*	*Aj=5*	*P(d=1)*	*P(q=1)*	*HH-size*
Ethiopia	208	405	1.00	0.25	0.09	0.17	0.27	0.39	0.70	0.0001	0.0038	4.60
Ghana	123	380	0.33	0.67	0.08	0.10	0.35	0.84	1.00	0.0000	0.0047	3.50
Kenya	163	440	1.00	1.00	0.07	0.12	0.19	0.27	0.49	0.0001	0.0056	3.90
Liberia	130	445	1.00	1.00	0.09	0.17	0.21	0.43	0.80	0.0001	0.0013	5.00
Malawi	110	410	1.00	0.67	0.06	0.12	0.15	0.30	0.57	0.0001	0.0040	4.50
Nigeria	105	400	0.33	0.25	0.09	0.15	0.23	0.41	0.69	0.0000	0.0012	4.60
South Africa	195	430	0.67	0.75	0.18	0.21	0.56	1.00	1.00	0.0016	0.0591	3.20
Tanzania	65	410	1.00	0.50	0.08	0.16	0.19	0.39	0.73	0.0000	0.0005	4.90

Variable	Definition
t	Days of school closures in 2020/21
T	Days of school in two typical years
f	Remote learning index: Offline
n	Remote learning index: Online
Aj	Share of households with access to internet (education level of household head j)
P(d=1)	COVID-19 deaths per inhabitants 31/12/2021
P(q=1)	COVID-19 cases per inhabitants 31/12/2021
HH-size	Average country-level household size

Notes: Education of the household head j=1 incomplete primary or no schooling, j=2 complete primary, j=3 incomplete secondary, j=4 complete secondary, j=5 at least some tertiary education. For more information on single data sources, see Appendix A in the Supplemental Material.

Table 3 – Average days of instructional time lost considering mitigation strategies

	Days of instructional time lost in 2020 and 2021		… in percentage of the regular school years	
	disadvantaged children	advantaged children	disadvantaged children	advantaged children
Ethiopia	101	33	25%	8%
Ghana	99	30	26%	8%
Kenya	76	23	17%	5%
Liberia	59	17	13%	4%
Malawi	53	16	13%	4%
Nigeria	86	28	22%	7%
South Africa	118	33	27%	8%
Tanzania	31	10	8%	2%

Notes: Disadvantaged children are children whose parents did not complete primary education. Advantaged children are those with at least one parent who completed primary education.

Figure 1 – Schematic representation of the model's functioning

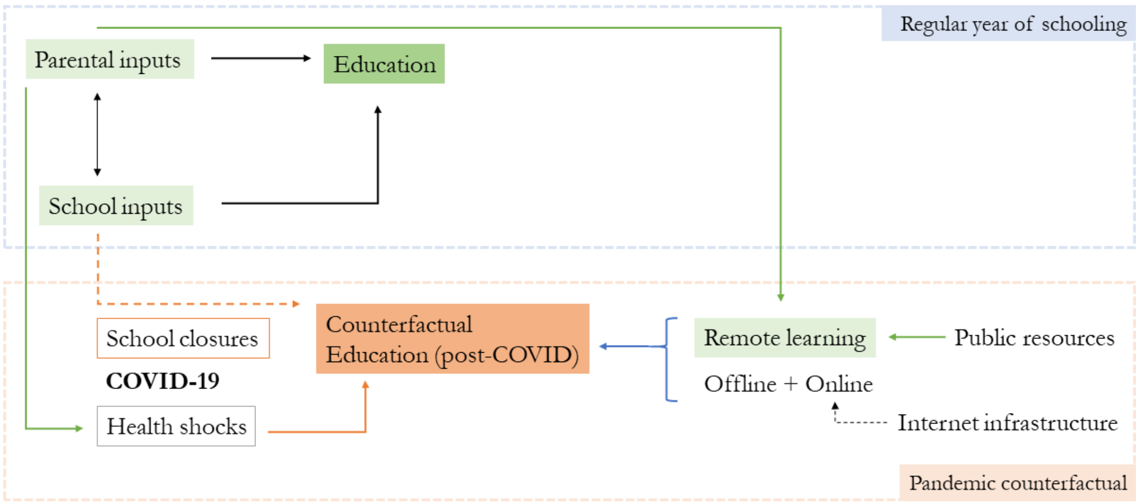
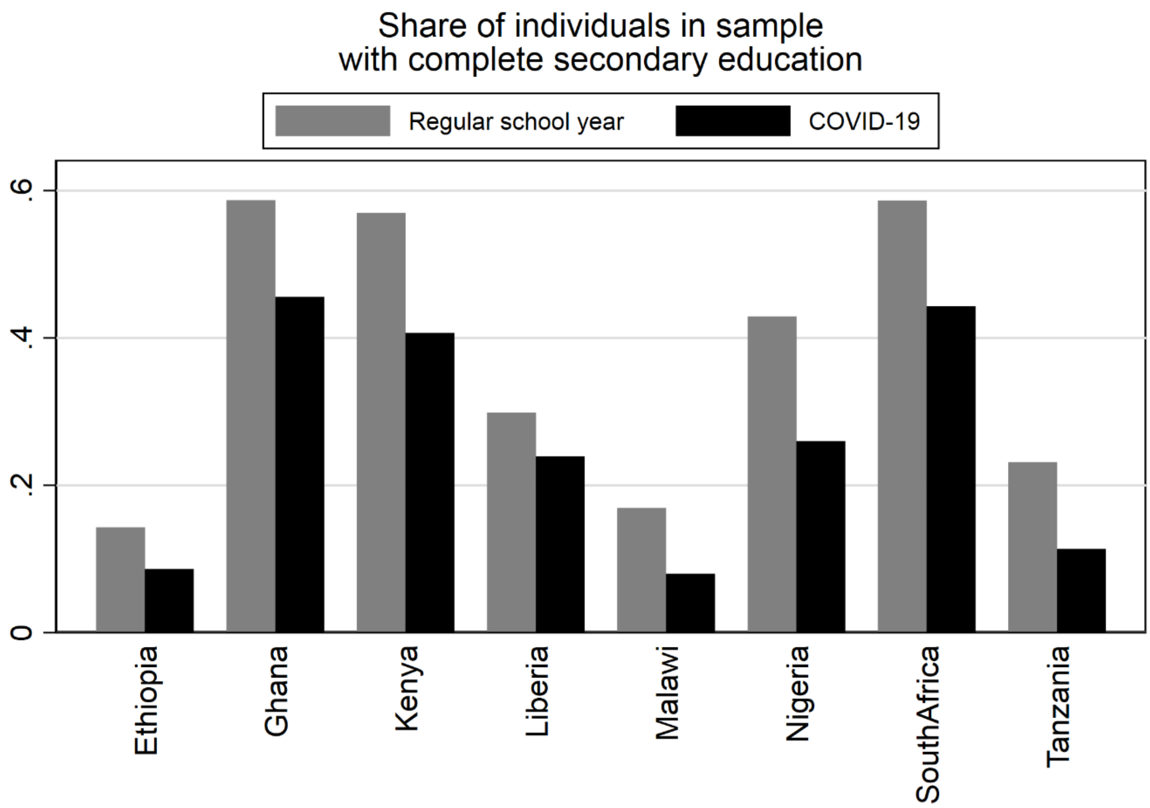
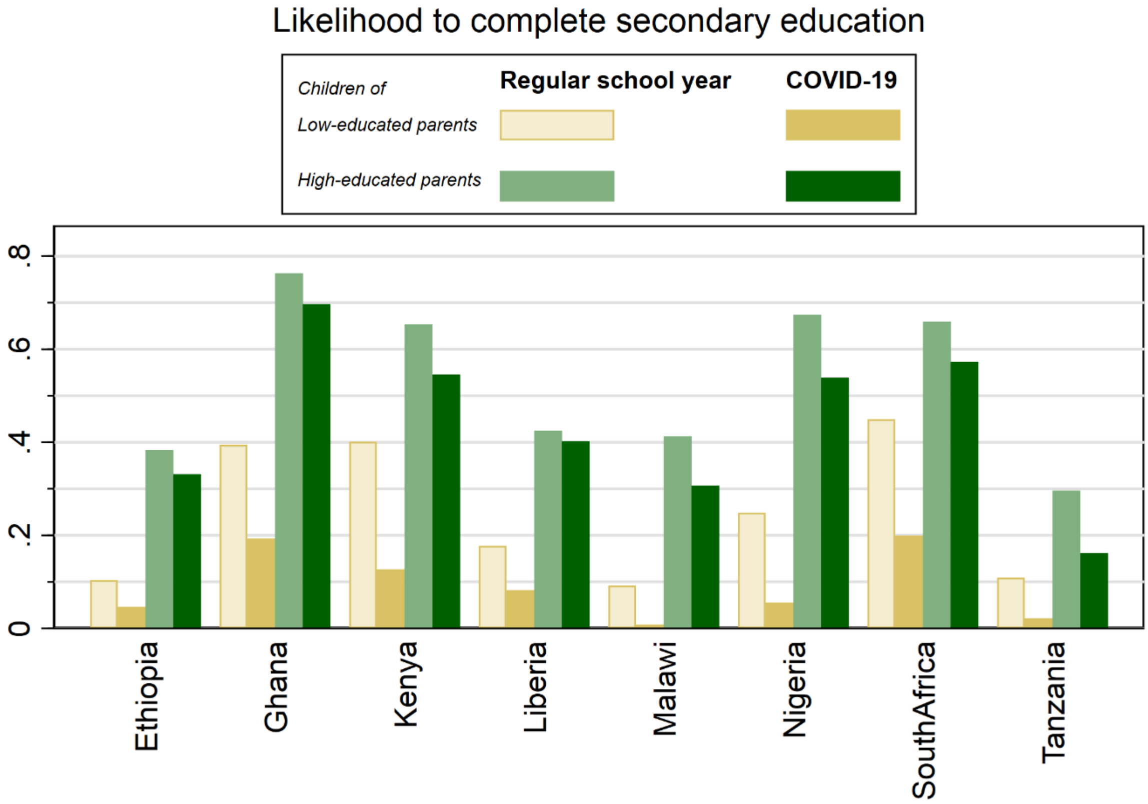


Figure 2 – Predicted impact of the pandemic on secondary school completion



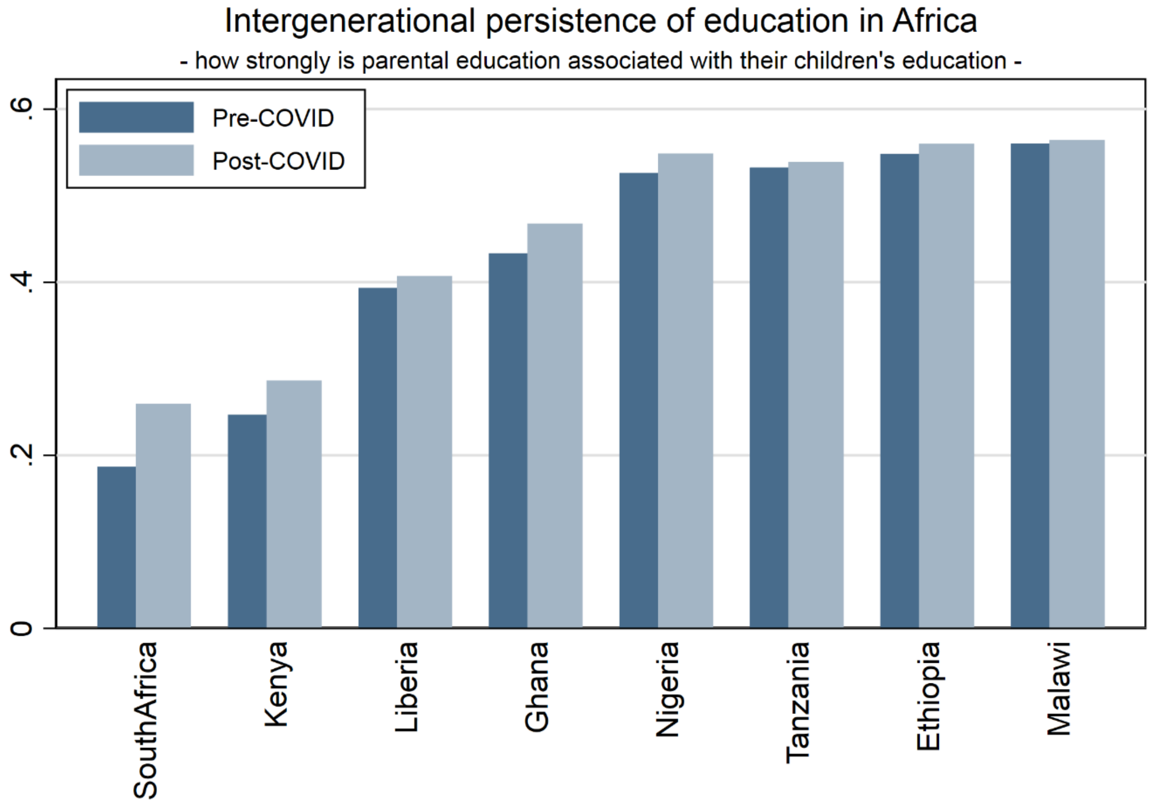
Notes: Completed secondary education is equivalent to 12 full years of schooling. First scenario shows actual share of individuals in sample with completed secondary schooling. Second scenario shows estimates of the same share after simulation of the COVID-19 shock. Source: National household surveys, own estimates.

Figure 3 – Predicted impact of the pandemic on secondary school completion by parental background



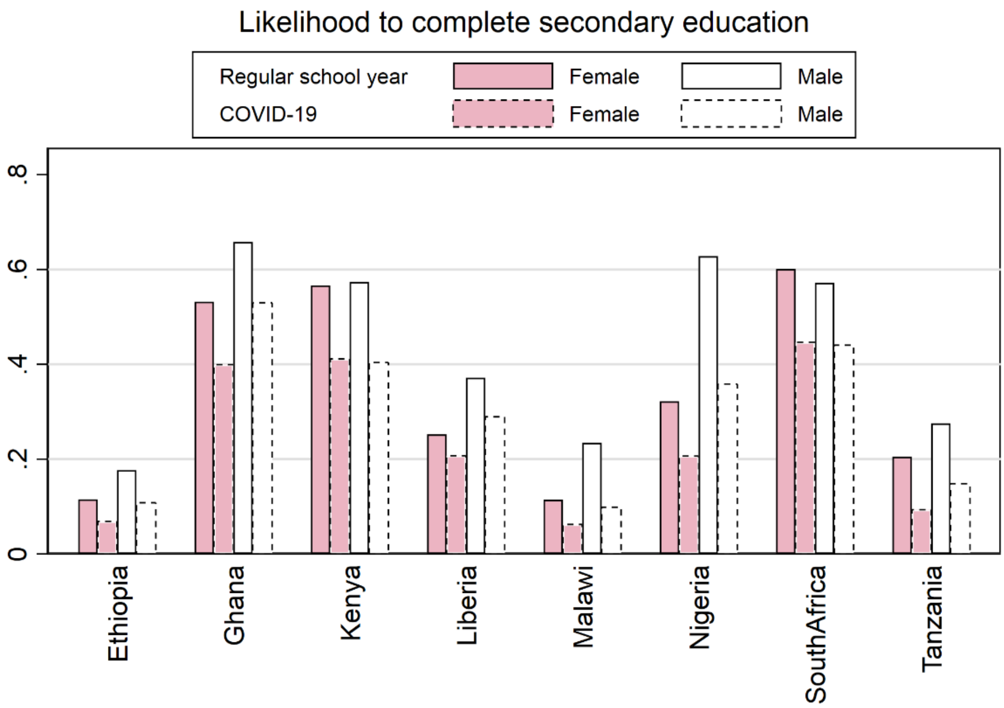
Notes: Bars show the likelihood to complete at least 12 years of schooling before and after simulation of the COVID-19 shock on education. High educated parents have at least completed primary education, low educated parents less than primary secondary education. Source: National household surveys, own estimates.

Figure 4 – Change in the intergenerational persistence of education due to the COVID-19 pandemic



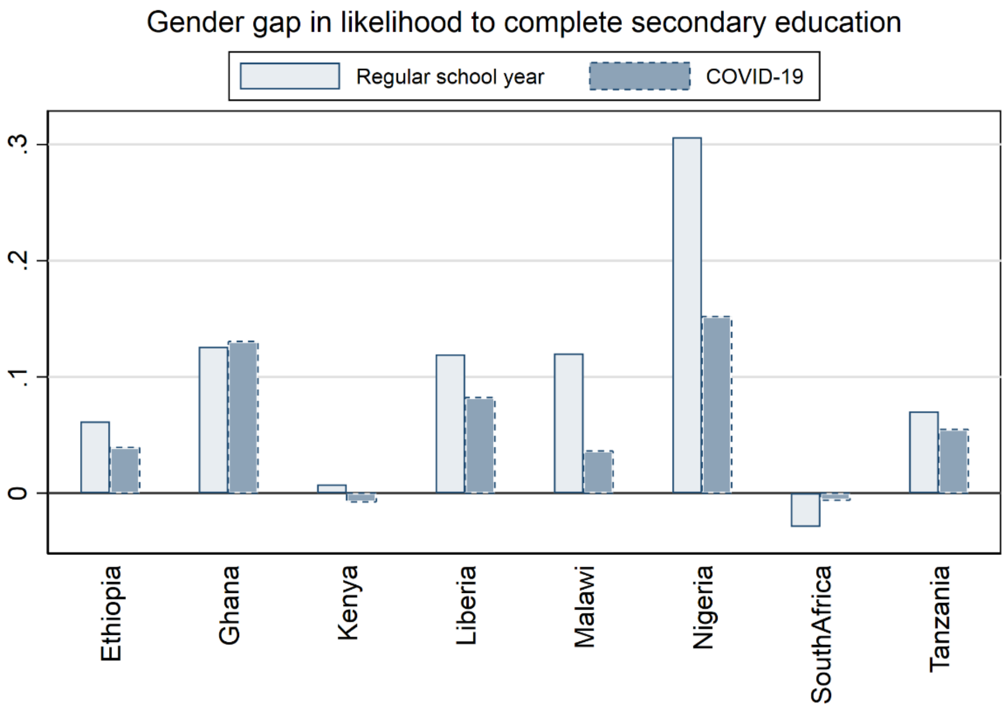
Notes: Bars show the pre-pandemic and counterfactual post-pandemic level of intergenerational persistence of education, measured by the slope coefficient. Source: National household surveys, own estimates.

Figure 5 – Heterogeneous impact of the pandemic on secondary school completion by gender



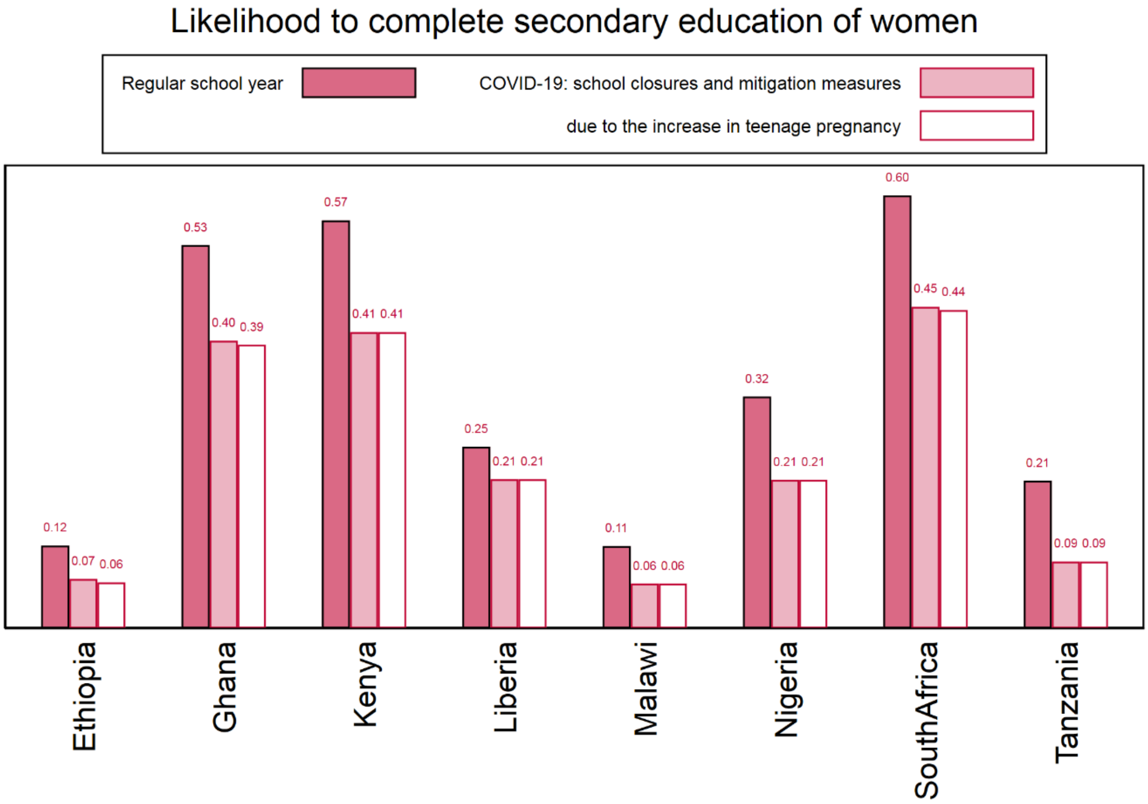
Notes: Bars show the pre-pandemic and counterfactual post-pandemic secondary school completion rate of men and women. Source: National household surveys, own estimates.

Figure 6 – Predicted change in the gender gap in secondary school completion due to the pandemic



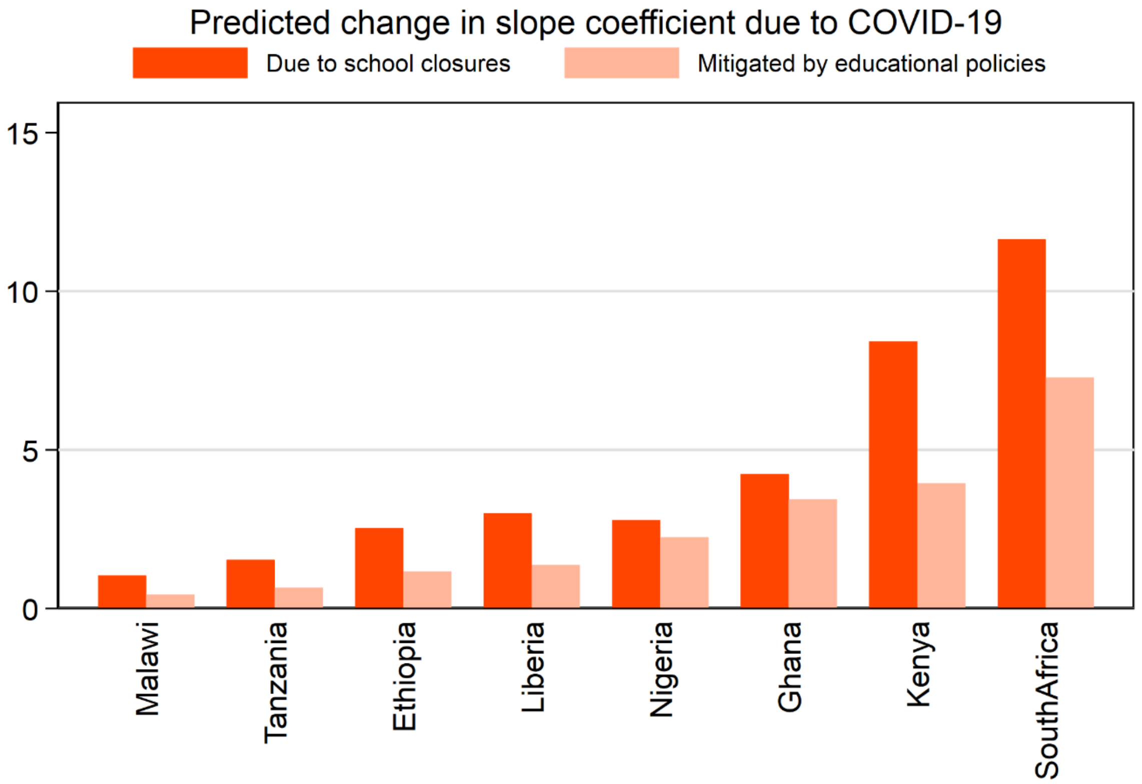
Notes: Bars show the pre-pandemic and counterfactual post-pandemic difference between secondary school completion rates of men and women. Source: National household surveys, own estimates.

Figure 7 – Predicted impact of increasing teenage pregnancy on secondary school completion rates



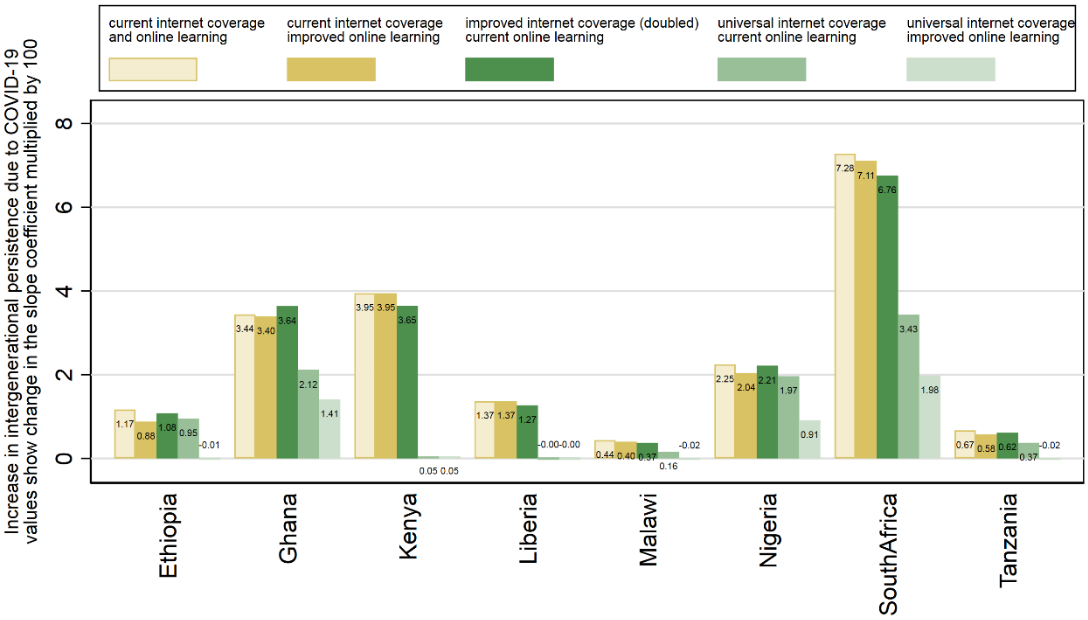
Notes: Bars show the likelihood to complete at least 12 years of schooling before and after simulation of the COVID-19 shock on education. The last bar shows the additional decrease in secondary school completion rates due to the predicted increase in teenage pregnancy. High educated parents have at least completed primary education, low educated parents less than primary secondary education. Source: National household surveys, own estimates.

Figure 8 – Evaluation of the impact of educational policies on reducing the effect of the pandemic on intergenerational mobility



Notes: Bars show the difference (multiplied by 100) between the pre-pandemic and counterfactual post-pandemic level of intergenerational persistence, measured by the slope coefficient, in two scenarios: i) only the effect of school closures on the instructional loss is taken into account; 2) including the mitigating effect of educational policies to provide offline and online remote learning. Source: National household surveys, own estimates.

Figure 9 – Policy simulation exercise to evaluate the combination of measures that would allow to cushion the effects of the pandemic on education



Notes: Bars show the difference (multiplied by 100) between the pre-pandemic and counterfactual post-pandemic level of intergenerational persistence, measured by the slope coefficient, in different scenarios of improved online remote learning and/or internet coverage. The first bar shows the difference between the pre-pandemic level and the baseline estimate for the post-pandemic counterfactual (i.e. given the current online remote learning efforts and distribution of internet access). Source: National household surveys, own estimates.

Projecting the Impact of COVID-19 on Education and Intergenerational Mobility in Sub-Saharan Africa

SUPPLEMENTAL MATERIAL

for online publication only

APPENDIX A: Data sources

Our analysis of instructional losses requires the following data: (i) a cross-sectional sample including information on educational attainment of adult respondents and their parents, as well as basic demographic characteristics of respondents (birth year, gender, and age) for each country; (ii) the share of households with internet access by educational level of the household head; (iii) the number of days of school closures due to COVID-19 pandemic; (iv) the mitigation policies implemented by governments to support learning from home during periods of school closures; (v) the number of COVID-19 cases and deaths. The evaluation of changes in long-run inequality requires, in addition, (vi) country data on average monthly earnings, and country data on wage returns to education. Furthermore, we retrieve (vii) data on teenage pregnancy before and during the pandemic.

A1 – Survey samples

We use the most updated household survey for each country that includes retrospective information on parental education.

- **Ethiopia**: Living standard measurement study. Socioeconomic Survey 2018-2019. Central Statistic Agency of Ethiopia. Link: https://microdata.worldbank.org/index.php/catalog/3823.
- **Ghana:** Ghana Living Standards Survey 2016-2017. Link: https://statsghana.gov.gh/gssdatadownloadspage.php.
- **Kenya**: STEP Skills Measurement Household Survey 2013 (Wave 2). Link: https://microdata.worldbank.org/index.php/catalog/2226.
- **Liberia:** Household Income and Expenditure Survey 2016-2017. Link: https://microdata.worldbank.org/index.php/catalog/2986.
- **Malawi:** Fifth integrated household survey 2019/2020 and the integrated household panel survey 2019-2020. Link: https://microdata.worldbank.org/index.php/catalog/3818.
- **Nigeria:** General Household Survey, Panel 2018-2019, Wave 4. Link: https://microdata.worldbank.org/index.php/catalog/3557.
- **Tanzania:** Household Budgetary Survey 2017-2018. Link: https://www.worldbank.org/en/programs/lsms/initiatives/lsms-ISA#acc40.
- **South Africa:** National Income Dynamics Study 2017, Wave 5. Link: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/712.

The surveys are collected from national statistical offices in different years and with different methodologies. To make the results comparable across countries, we harmonize the categories of all variables used. Specifically, the educational attainment variable varies between countries because each education system has specific features. Each country has a different quantity of years dedicated to primary and secondary education, as well as different definitions of vocational training, and post-secondary education. Considering this heterogeneity, we define the following educational categories: Illiterate, incomplete primary, complete primary, incomplete secondary, complete secondary, incomplete tertiary, and complete tertiary.[1]

---

[1] Since it is not possible in all countries to obtain such detailed information on the education of respondents and their parents to form this seven categories, for the final analysis they are reduced into five categories to

We convert the specific categories in each country into the ones listed above by using the harmonized educational categories provided in the 2011 standardized tables from ISCED. These tables are part of a collaborative process between the UNESCO Institute for Statistics (UIS) and the Member States to map national education systems according to the International Standard Classification of Education (ISCED). This classification combines expert information with surveys launched by the UIS to gather detailed information on national education systems in each country. The survey includes questions for each education program on entry requirements, entry age, duration and diplomas obtained, and their corresponding ISCED level for pre-primary to tertiary education. In addition, when it is necessary, countries are consulted individually to resolve potential classification problems which might compromise the comparability of their education data. Based on this information, the UIS produces individual ISCED mappings for each country in consultation with local authorities.[2] The ISCED tables contain the following categories: (i) Early childhood education; (ii) Primary education; (iii) Lower secondary education; (iv) Upper secondary education; (v) Post-secondary non-tertiary education; (vi) short-cycle tertiary education; (vii) Bachelor's or equivalent degree; (viii) Doctoral or equivalent level.[3] To match ISCED categories with the standard categories mentioned above, we consider as incomplete primary those observations in (i) and (ii) without completing the corresponding years of education for that level (whenever possible). Lower secondary education is considered as incomplete secondary education in our classification and upper secondary education to complete secondary education. Components (v) and (vi) correspond to incomplete tertiary, and we consider (vii) and (viii) as complete tertiary education.

This grouping differs slightly for Tanzania. In this case, we decided to define Form IV, which ISCED defines as lower secondary education, as complete secondary education. The reason is that most students stop learning after the fourth year of secondary education. This feature comes from the fact that all students must take an examination after Form IV to finish the remaining two years. In the past, this test had pass rates of around 50%. Approximately half of the students failed at that stage and had to decide whether going to a vocational training school or dropping out of the education system. However, even if students successfully approve the exam after Form IV, the remaining two years (Form V and VI) are taught in English. So, those students without good training in English during the previous cycle mostly fail in the last part of secondary school. Therefore, even though Form VI is the legal requirement to enter university, the turning point in secondary education in Tanzania's education system seems to be rather Form IV.

The exact classification of education that we use in our analysis, based on ISCED data and the sources mentioned above, is indicated in the following tables.

---

preserve the homogeneity of the information across countries. The classification used to obtain the estimates is: incomplete primary or no schooling, complete primary, incomplete secondary, complete secondary, at least some tertiary education.

[2] Further details about this classification can be found here: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

[3] All tables are available here http://uis.unesco.org/en/isced-mappings

Correspondence with ISCED Education tables

Ethiopia

Name of the education programme (National language)	ISCED 2011	Education level our category
Kindergarten	Early childhood education	Incomplete primary
Primary (grades 1 to 6)	Primary education	Complete primary
Primary (grades 7 to 8)	Lower secondary education	Incomplete secondary
Secondary First Cycle (grades 9-10)	Lower secondary education	Incomplete secondary
Secondary Second Cycle Preparatory Programme (grades 11-12)	Upper secondary education	Complete secondary
Pre-school Teaching certificate programme	Upper secondary education	Complete secondary
Technical / Vocational education training (TVET level 1)	Upper secondary education	Complete secondary
Technical / Vocational education training (TVET level 2)	Upper secondary education	Complete secondary
First cycle of primary teaching certificate (Grades 1-4) programme	Upper secondary education	Complete secondary
Second cycle of primary teaching certificate (Grades 5-8) programme	Upper secondary education	Complete secondary
Technical / Vocational education training (TVET level 3)	Upper secondary education	Complete secondary
Technical / Vocational education training (TVET level 4)	Post-secondary non-tertiary education	Incomplete tertiary
Technical / Vocational education training (TVET level 5)	Post-secondary non-tertiary education	Incomplete tertiary
Undergraduate degree (Short)	Bachelor's or equivalent level	Complete tertiary
Secondary Education teacher (long)	Bachelor's or equivalent level	Complete tertiary
Undergraduate degree (Law, pharmacy)	Bachelor's or equivalent level	Complete tertiary
Undergraduate degree (medicine and veterinary science)	Bachelor's or equivalent level	Complete tertiary
Master's degree	Master's or equivalent level	Complete tertiary
Doctorate degree	Doctoral or equivalent level	Complete tertiary

Kenya

Name of the education programme (National language)	ISCED 2011	Education level our category
First stage of primary education (Standards 1 to 3)	Primary education	Incomplete primary
Second stage of primary education (Standards 4 to 6)	Primary education	Complete primary
Second stage of primary education (Standards 7 and 8)	Lower secondary education	Incomplete secondary
Youth polytechnics	Lower secondary education	Incomplete secondary
Secondary education	Upper secondary education	Complete secondary
Youth polytechnics	Upper secondary education	Complete secondary
Technical Vocational Education and Training (TVET)	Post-secondary non-tertiary education	Incomplete tertiary

Pre-Primary Teacher training	Short-cycle tertiary education	Incomplete tertiary
Primary teacher training college	Short-cycle tertiary education	Incomplete tertiary
Teacher training college diploma	Short-cycle tertiary education	Incomplete tertiary
National polytechnics (certificate and diploma)	Short-cycle tertiary education	Incomplete tertiary
National polytechnics (Higher diploma)	Bachelor's or equivalent level	Complete tertiary
Bachelor's degree (Science, Education, Education Science, Arts, Law, Commerce...)	Bachelor's or equivalent level	Complete tertiary
Bachelor's degree (Engineering, Medicine….)	Bachelor's or equivalent level	Complete tertiary
Bachelor's degree (Architecture)	Master's or equivalent level	Complete tertiary
Master's degree	Master's or equivalent level	Complete tertiary
Doctorate	Doctoral or equivalent level	Complete tertiary

Ghana

Name of the education programme (National language)	ISCED 2011	Education level our category
ECD programme	Early childhood education	Incomeplete primary
Kindergarten	Early childhood education	Incomeplete primary
Primary school	Primary education	Complete primary
Junior high school	Lower secondary education	Incomeplete secondary
Senior high school	Upper secondary education	Complete secondary
Technical and vocational education	Upper secondary education	Complete secondary
Polytechnics non-tertiary programmes	Post-secondary non-tertiary education	Incomplete tertiary
Polytechnics Tertiary programmes	Short-cycle tertiary education	Incomplete tertiary
Teacher training diploma	Short-cycle tertiary education	Incomplete tertiary
Professional bodies programmes	Short-cycle tertiary education	Incomplete tertiary
Polytechnics Tertiary programmes	Bachelor's or equivalent level	Complete tertiary
University education - first degree	Bachelor's or equivalent level	Complete tertiary
University education - second degree	Master's or equivalent level	Complete tertiary
PhD programme	Doctoral or equivalent level	Complete tertiary

Liberia

Name of the education programme (National language)	ISCED 2011	Education level our category
Early childhood education	Early childhood education	Incomplete primary
Lower basic education	Primary education	Complete primary
Upper basic education	Lower secondary education	Incomplete secondary
Senior secondary education	Upper secondary education	Complete secondary
Multilateral High School	Upper secondary education	Complete secondary
Technical and Vocational Education	Upper secondary education	Complete secondary
Teacher Training	Post-secondary non-tertiary education	Incomplete tertiary
Polytechnics/ Community Colleges Programmes	Short-cycle tertiary education	Incomplete tertiary
Bachelor's	Bachelor's or equivalent level	Complete tertiary
Bachelor's (Engineering and Medicine)	Bachelor's or equivalent level	Complete tertiary
Master's	Master's or equivalent level	Complete tertiary

Malawi

Name of the education programme (National language)	ISCED 2011	Education level our category
Pre-primary education	Early childhood education	Incomplete primary
Primary (Grades 1 to 6)	Primary education	Complete primary
Primary (Grades 7 and 8)	Lower secondary education	Incomplete secondary
Junior secondary (Forms 1 and 2)	Lower secondary education	Incomplete secondary
Senior secondary (forms 3 and 4)	Upper secondary education	Complete secondary
Nursing and school of health science certificate programmes	Post-secondary non-tertiary education	Incomplete tertiary
Teacher training college programmes	Post-secondary non-tertiary education	Incomplete tertiary
Technical vocational diploma programmes	Short-cycle tertiary education	Incomplete tertiary
Diploma in nursing and in health science programmes (3 years)	Short-cycle tertiary education	Incomplete tertiary
Diploma in nursing and in health science programmes (2 years)	Short-cycle tertiary education	Incomplete tertiary
Diplomas and degree programme in teaching	Short-cycle tertiary education	Incomplete tertiary
Bachelor's degree	Bachelor's or equivalent level	Complete tertiary
Bachelor's of Science in Medicine	Bachelor's or equivalent level	Complete tertiary
Master's degree	Master's or equivalent level	Complete tertiary
Master's of Science in Medicine	Master's or equivalent level	Complete tertiary
Ph.D	Doctoral or equivalent level	Complete tertiary

Nigeria

Name of the education programme (National language)	ISCED 2011	Education level our category
Early childhood care development and education (ECCDE)	Early childhood education	Incomplete primary
Pre-primary education	Early childhood education	Incomplete primary
Primary education	Primary education	Complete primary
Junior secondary	Lower secondary education	Incomplete secondary
Vocational enterprise institutions programmes	Lower secondary education	Incomplete secondary
Innovative enterprise institute	Lower secondary education	Incomplete secondary
Senior secondary	Upper secondary education	Complete secondary
Interim joint matriculation board (IJMB) A - level course	Upper secondary education	Complete secondary
Secondary technical schools programmes	Upper secondary education	Complete secondary
Nigerian certificate in education (NCE)	Short-cycle tertiary education	Incomplete tertiary
National diploma (ND) programme	Short-cycle tertiary education	Incomplete tertiary
Higher national diploma (HND) programme	Short-cycle tertiary education	Incomplete tertiary
School of Nursing	Short-cycle tertiary education	Incomplete tertiary
Bachelor's in Nursing	Bachelor's or equivalent level	Complete tertiary
Bachelor's programme	Bachelor's or equivalent level	Complete tertiary
Post Graduate Diploma Programme	Bachelor's or equivalent level	Complete tertiary
Master's programmes	Master's or equivalent level	Complete tertiary
Master's of philosophy	Master's or equivalent level	Complete tertiary
Ph.D	Doctoral or equivalent level	Complete tertiary

South Africa

Name of the education programme (National language)	ISCED 2011	Education level our category
Grade R	Early childhood education	Incomplete primary
Primary education (Grades 1 to 7)	Primary education	Complete primary
Lower secondary education (Grades 8 to 9)	Lower secondary education	Incomplete secondary
Upper secondary education (Further education training band: Grades 10 to 12)	Upper secondary education	Complete secondary
National Certificate (Vocational) - NC(V) Levels 2 to 4	Upper secondary education	Complete secondary
NATED courses Level 1 to 3	Upper secondary education	Complete secondary
NATED courses Level 4 to 5 (Business)	Upper secondary education	Complete secondary
NATED courses Level 4 to 5 (Engineering)	Upper secondary education	Complete secondary
National Higher certificate	Post-secondary non-tertiary education	Incomplete tertiary
Advanced certificate (AC)	Post-secondary non-tertiary education	Incomplete tertiary
NATED courses Level 6 (Business)	Post-secondary non-tertiary education	Incomplete tertiary
NATED courses Level 6 (Engineering)	Post-secondary non-tertiary education	Incomplete tertiary
National Diploma	Short-cycle tertiary education	Incomplete tertiary
Bachelor's	Bachelor's or equivalent level	Complete tertiary
Advanced diploma (AD)	Bachelor's or equivalent level	Complete tertiary
Honours degree	Bachelor's or equivalent level	Complete tertiary
Post graduate diploma (PGD)	Bachelor's or equivalent level	Complete tertiary
Master's	Master's or equivalent level	Complete tertiary
Doctorate degree / Laureatus in Technology (Technikon)	Doctoral or equivalent level	Complete tertiary

Tanzania

Name of the education programme (National language)	ISCED 2011	Education level our category
Pre primary education	Early childhood education	Incomplete primary
Primary education	Primary education	Complete primary
Ordinary level secondary education	Lower secondary education	Incomplete secondary
National Vocational and Training Level 1 (NVTA 1)	Lower secondary education	Incomplete secondary
National Vocational and Training Level 2 (NVTA 2)	Lower secondary education	Incomplete secondary
National Vocational and Training Level 3 (NVTA 3)	Lower secondary education	Incomplete secondary
University Certificate	Upper secondary education	Complete secondary
University Certificate	Upper secondary education	Complete secondary
Advanced level secondary education	Upper secondary education	Complete secondary
National Technical Award Level 4 (NTA 4)	Upper secondary education	Complete secondary
National Technical Award Level 4 (NTA 4)	Upper secondary education	Complete secondary
Teacher Training (Certificate)	Upper secondary education	Complete secondary
National Technical Award Level 5 (NTA 5)	Upper secondary education	Complete secondary
University Diploma	Post-secondary non-tertiary education	Incomplete tertiary
University Diploma	Post-secondary non-tertiary education	Incomplete tertiary

Teacher Training (Diploma)	Post-secondary non-tertiary education	Incomplete tertiary
National Technical Award Level 6 (NTA 6)	Post-secondary non-tertiary education	Incomplete tertiary
National Technical Award Level 7 (NTA 7)	Short-cycle tertiary education	Incomplete tertiary
Advanced Diploma	Short-cycle tertiary education	Incomplete tertiary
Post Graduate Certificate (PGC)	Short-cycle tertiary education	Incomplete tertiary
Undergraduate	Bachelor's or equivalent level	Complete tertiary
University Bachelor Degree	Bachelor's or equivalent level	Complete tertiary
National Technical Award Level 8 (NTA 8)	Bachelor's or equivalent level	Complete tertiary
University Bachelor Degree (Medicine, Nursing, Pharmacy)	Bachelor's or equivalent level	Complete tertiary
Post Graduate Certificate (PGC)	Bachelor's or equivalent level	Complete tertiary
Post Graduate Diploma (PGD)	Bachelor's or equivalent level	Complete tertiary
Masters	Master's or equivalent level	Complete tertiary
National Technical Award Level 9 (NTA 9)	Master's or equivalent level	Complete tertiary
PhD courses	Doctoral or equivalent level	Complete tertiary
National Technical Award Level 10 (NTA 10)	Doctoral or equivalent level	Complete tertiary

A2 – Internet access

The second required information is the availability of an internet connection by educational level of the household head. For some countries, we can estimate this based on the surveys mentioned above. However, not all of these surveys include questions on internet access. Hence, we retrieve other surveys that include this information.

- **Ghana:** Ghana Living Standards Survey 2016-2017. Link: https://statsghana.gov.gh/gssdatadownloadspage.php.
- **Kenya**: Kenya Integrated Household Budget Survey 2015-2016. Link: https://statistics.knbs.or.ke/nada/index.php/catalog/13.
- **Malawi:** Fifth integrated household survey 2019/2020 and the integrated household panel survey 2019-2020. Link: https://microdata.worldbank.org/index.php/catalog/3818.
- **Nigeria:** General Household Survey, Panel 2018-2019, Wave 4. Link: https://microdata.worldbank.org/index.php/catalog/3557.
- **South Africa:** General Household Survey, 2018. Link: http://www.statssa.gov.za/?p=12180.

For Kenya, Nigeria, and South Africa, we estimate the distribution of internet connection using the questions asking if the household has internet access. It is a binary variable, and all household members respond to this question. We restrict the answers to the household head. The educational categories are equal to those presented before. For Malawi and Ghana, we cannot directly identify the internet connection in each household because there is no such question. In Ghana, we approximate it using questions about the use of internet in the last three months. We define a household as connected to the internet if the household head used the internet in the last three months. In Malawi, we infer internet connection access through information on household expenditures on internet. We define households connected to the internet as those households with positive spending on internet. For

Ethiopia, Tanzania, and Liberia we could not find nationally representative household survey data enabling us to estimate access to the internet by education level.[4] As a result, we decide to impute the data using the information available for other countries and some pre-defined selection criteria. The imputation that we apply works as follows: From the pool of countries with information to estimate internet access, we select two countries with the most similar average internet connection to the country we have to impute using the most updated data point available in the data on overall access to the internet provided by the World Bank.[5]

Table A1 - Average connection differences between countries to be imputed and donors

Donors	Difference in average connection		
	Ethiopia	*Tanzania*	*Liberia*
Kenya	-0.024	0.026	0.006
Nigeria	0.086	0.136	0.116
South Africa	0.432	0.482	0.462
Ghana	0.180	0.230	0.210
Malawi	-0.095	-0.045	-0.065

Source: Own elaboration based on World Bank Data.

Table A1 shows the difference between the average internet connection in the donor country and each country in which the values need to be imputed. For each country, we highlight the cells with the smallest differences. We eventually selected Kenya and Malawi as donors for Tanzania and Kenya and Nigeria for Ethiopia. To decide which of the selected countries is more appropriate, we compare the differences in GDP per capita (at U$ 2017 PPP). The results of these comparisons are shown in Table A2. Again, for each country in the columns, highlighted cells correspond to the most similar country. Kenya is, finally, the donor of the imputed values for Ethiopia. Tanzania and Liberia will be imputed using Malawi's data on internet access.

Table A2 - GDP per capita differences between countries to be imputed and donors

Donors	Distance in GDP per capita		
	Ethiopia	*Tanzania*	*Liberia*
Kenya	2231	1792	2983
Nigeria	2914	n.a	n.a
Malawi	n.a	-1123	67

Source: Own elaboration based on World Bank Data.

As mentioned before, using estimates over multiple dimensions simultaneously might be problematic with household surveys, given that, in general, they are not designed for this

[4] Besides of national household surveys, we considered the phone surveys carried out by the World Bank during 2020 for monitoring the COVID-19 impact. However, we cannot match the educational categories in our samples with those collected in the phone survey. Furthermore, the sample size of the phone surveys is rather small and provides inaccurate estimates by education level for each country. Another alternative that we explored is the RIA ICT Access Survey 2017-2018 carried out by the University of Cape Town. However, the problem in using this survey is the small sample size, which is particularly relevant in the case of the internet connection question.

[5] The data is available here: https://data.worldbank.org/indicator/IT.NET.USER.ZS

purpose. So, the data may exhibit representativity problems. This problem is not easy to tackle because it is not easy to determine the potential bias. Using the levels of internet access produced by the International Telecommunication Union (ITU) published by the World Bank, we correct our estimates to match the aggregate levels of internet connection access.[6] Founded in 1865, ITU facilitates international connectivity in communications networks. ITU collects information regarding information technologies, internet usage, and internet access for households and individuals since 2004. The household indicators about internet access come from an annual questionnaire sent to the Member States.[7] We take this information at the aggregate level. Using this information, we rescale our estimated share of households connected to the internet connection for each educational level by the ratio between the internet connection estimated by ITU and the average connection estimated in the household survey. Formally:

$$IC_{i,j} = IC_{i,j}^{HS} * \left(\frac{\overline{IC}_j^{WB}}{\overline{IC}_j^{HS}} \right)$$

Where IC_i is the adjusted internet access for the educational level i in the country j. IC_i^{HS} is the internet access estimated in the household survey for the educational level i in the country j. $\overline{IC}^{HS}$ is the average internet access estimated in the household survey in country j, and $\overline{IC}^{WB}$ is the average internet access estimated by the World Bank in the country j.

As can be noted, depending on the initial values for the average connections, this process may produce an average internet connection higher than one. When it occurs, we set the categories with a share higher than one to one (i.e. 100% of households with this educational level are connected to the internet). Then, we adjust the remaining internet access to match the ITU estimates. If this second adjustment produces internet accesses higher than one, we repeat the process.

A3 – School closures

We retrieve data on the total duration of school closures in 2020 and 2021 from UNESCO and UNICEF. The data is filed by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) within the Humanitarian Data Exchange (HDX) platform.[8] The data includes for each country the total number of weeks that schools were fully or partially closed during the pandemic and the total number of weeks of academic break. We estimate the total number of weeks of school in a regular year as 52 (the number of weeks in a year) minus the number of weeks of academic break. To convert the values in days, we multiply each item with 5. The data covers the period from mid-February, 2020 to January 31, 2022.

[6] The data is available here: https://data.worldbank.org/indicator/IT.NET.USER.ZS.
[7] To see in detail the definitions and methodology entailed in the question about internet access please see the 2020 manual of ITU see https://www.itu.int/en/ITU-D/Statistics/Documents/publications/manual/ITUManualHouseholds2020_E.pdf
[8] The information is available here: https://data.humdata.org/dataset/global-school-closures-covid19.

Table A3 – Duration of school closures (in weeks) in 2020 and 2021

	Duration of FULL closures	*Duration of PARTIAL closures*	*Duration in two normal years*
Ethiopia	21	41	81
Kenya	28	9	88
Nigeria	18	6	80
South Africa	15	48	86
Ghana	10	29	76
Liberia	15	22	89
Tanzania	11	4	82
Malawi	18	8	82

Source: Humanitarian Data Exchange Platform, own elaboration.

A4 – Educational mitigation policies

To estimate the extent to which governments intervened to respond to the school closures offering offline and online remote learning tools, we use the Survey on National Education Responses to COVID-19 School Closures.[9] This survey was initiated by the United Nations Educational, Scientific and Cultural Organization (UNESCO), the United Nations Children's Fund (UNICEF), the World Bank and the Organization for Economic Co-operation and Development (OECD) to collect information on national education responses to school closures related to the COVID-19 pandemic. The respondents of this survey are, in general, officials of the Ministry of Education at the central or decentralized level in charge of school education.

This survey had three waves of collection. The first wave of data collection started in May and lasted until June 2020. The second wave of data collection lasted from July to October 2020. The third wave of data collection lasted from February to April 2021. We combine the data from the three waves of the survey, and select questions that are stables across waves so that we can maximize the information about mitigation policies in all countries. Using this information, we create two indexes summarizing the extent to which governments supplied online and offline learning resources to teachers and students during the COVID-19 pandemic.

To compute the index for offline learning we use the three following survey items:[10]

- Were TV programs used to teach during COVID-19 pandemic?
- Were radio programs used to teach during COVID-19 pandemic?
- Were paper-based strategies used to teach during COVID-19 pandemic?

Using these questions, we create three versions of the index. The first takes the average of all answers responded by the country's officials. The second takes the average of the most updated answer to each question. The third takes the average of the oldest answer to each

[9] The dataset is available here: https://tcg.uis.unesco.org/survey-education-covid-school-closures/

[10] Each question is not exactly the same across waves, but it was similar enough to use it as the same question. To see the specific details about the questions used in each wave please see *Table 2. Offline learning index: Questions used and methodological decisions.*

question. For our analysis, we use the third version. Results obtained with the other three versions do not differ substantially and are available upon request.

To compute the online learning index, we use questions that show whether online resources were used to provide remote learning, whether teachers were required to work online during the pandemic and whether they were trained to do so. In addition, it includes a question about the monitoring of the use of the resources provided by the government. Specifically, we use the following questions:[11]

- Were online platforms used to provide remote learning during the pandemic?
- Were teachers required to continue teaching online while schools were closed?
- Have teachers been trained and/or supported to use remote learning platforms?
- Was the actual use of distance learning monitored?

Again, we compute three versions of the index, as mentioned above, and choose the third version (oldest answers to this question) for our analysis.

Table A4 and A5 show a description of each question used to compute the indexes and the specific number of each question in the original survey.

[11] Again, the questions here are not exactly the same across waves, but they are similar enough to use it as the same question. To see the specific details about the questions used in each wave please see *Table 1. Online learning index: Questions used and methodological decisions.*

Table A4 - Online learning index: Questions used and methodological decisions

Wave	Block of questions	Question number	Question	Comments
Online learning index learning index				
1	1	6	Which of the following education delivery systems has been deployed as part of the national (or subnational) distance education strategy for different levels of education? (Delivery systems included: Online platform. Options: Yes, No).	If a given resource was provided at any educational level, it is considered as 1 (which means that this resource was implemented)
2	1	13	How effective have online platforms been in maintaining or advancing the levels of learning? (Options: Very Effective, Fairly Effective, Not Effective, Do not know, We do not have such platform).	
3	1	Section 4 Q1	Which distance learning solutions were or are being offered in your country during the pandemic in 2020 and/or 2021? (Select all that apply. Options included: Television, radio, Take-home packages, None)	
1	2	12	Are teachers required to continue teaching while schools are closed?	In the first two waves, the answers are Yes or No. In the third wave, we considered it as Yes if the percentage of teachers required to teach is higher than 0.
2	2	18	Are teachers or were teachers required to teach during school closures?	
3	2	Section 5 Q1	What percentage of teachers (primary to upper-secondary levels combined), approximately, were required to teach (remotely/online) during all school closures in 2020?	
1	3	14	Have teachers been trained to use remote learning platforms?	If any help was provided, we coded it as 1 which means that teachers receive support
2	3	20	How have teachers been supported in the transition to remote learning?	
3	3	Section 5 Q4	How and at what scale were teachers (in pre-primary to upper secondary levels combined) supported in the transition to remote learning in 2020?	
1	4	7	Is the actual use of distance learning monitored?	In the first two waves, the answers are Yes or No. In the third wave, we considered it as Yes if the percentage following distance learning is higher than 0.
2	4	-	Not available	
3	4	Section 4 Q2	What percentage of students (at each level of education), approximately, followed distance education during school closures in 2020?	

Source: Own elaboration based on Survey on National Education Responses to COVID-19 School Closures.

Table A5 - Offline learning index: Questions and methodological decisions

Offline learning index				
Wave	**Block of questions**	**Question number**	**Question**	**Comments**
1	1	6	Which of the following education delivery systems has been deployed as part of the national (or subnational) distance education strategy for different levels of education? (Delivery systems included: Radio, Television, Paper-based take-home materials. Options: Yes, No).	If the system was applied, it is coded equally to 1
2	2	13	How effective has television been in maintaining or advancing the levels of learning? (Options: Very Effective, Fairly Effective, Not Effective, Do not know, We do not have such platform).	If the system is considered Not effective, Fairly Effective, Effective, or Very Effective, it is coded equally to 1, which means it was applied in the country
2	2	13	How effective has radio been in maintaining or advancing the levels of learning? (Options: Very Effective, Fairly Effective, Not Effective, Do not know, We do not have such platform).	
2	2	13	How effective have take-home packages been in maintaining or advancing the levels of learning? (Options: Very Effective, Fairly Effective, Not Effective, Do not know, We do not have such platform).	
3	3	Section 4 Q1	Which distance learning solutions were or are being offered in your country during the pandemic in 2020 and/or 2021? (Select all that apply. Options included: Television, radio, Take-home packages, None)	-

Source: Own elaboration based on Survey on National Education Responses to COVID-19 School Closures.

A5 – COVID-19 cases

We retrieve information on COVID-19 cases and deaths in each country from the Website Our World in Data. We use the number on December 31, 2021. Data on the average number of people living in the same household are used to estimate the probability of infection and death within the household. This data is retrieved from United Nations Department of Economic and Social Affairs, Population Division (2017): Household Size and Composition Around the World 2017.

A6 – Country level earnings and returns to education

We retrieve country level monthly wages for each country from the Statistics of the International Labor Organization (ILOSTAT). Wages are displayed across all occupations and in international dollars PPP. Wage returns to education are collected from the literature review by Psacharopoulos and Patrinos (2018), which includes indicators for returns to education for almost all countries worldwide. These are mostly estimates obtained by Mincer regressions of earnings on years of schooling. Table A6 shows the values that we use in our analysis.

Table A6 – Monthly wages (ILOSTAT) and returns to education (Psacharopoulos and Patrinos, 2018)

	Men (=1) or Women (=0)	Monthly wages in international dollars PPP	Year of reference	Returns to Education	Year of reference
Ethiopia	0	143.43	2013	10.7	2011
Ethiopia	1	210.1	2013	13	2011
Ghana	0	160.4	2017	5.3	2007
Ghana	1	490.12	2017	7	2007
Kenya	0	288.04	2019	13.2	1995
Kenya	1	333.4	2019	13.2	1995
Malawi	0	139.88	2013	13.2	2010
Malawi	1	213.89	2013	11.8	2010
Nigeria	0	500.89	2013	6.5	2011
Nigeria	1	716.96	2013	5.1	2011
South Africa	0	507.25	2019	21.2	2008
South Africa	1	585.4	2019	18.1	2008
Tanzania	0	369.67	2014	14.6	2007
Tanzania	1	460.18	2014	9.2	2007

A7 – Teenage pregnancy

We calculate the proportion of adolescent births over the total adolescent population using data from the World Bank about the adolescent fertility rate (births per 1,000 women aged 15-19), the female population aged 15-19 (% of the female population), and the total female population. To calculate the increase in adolescent birth rates, we proceed in the following way: We derive the total female population ages 15-19 by multiplying the population ages 15-19 female (% of female population) by the total female population. Then, we derive the number of adolescent births using the following formula:

$$Adolescent\ births = \frac{Female\ pop^{15-19} \ * Adolscent\ births\ per\ 1000\ women\ ages\ 15-19}{1000}.$$

Table A7 shows the results of these calculations. Interestingly, the statistics from fertility suggest that both, the total quantity of adolescent births and the proportion of adolescent births as % of the adolescent population, declined in 2020 with respect to 2019. This result is consistent with the statistics presented in some reports of the country aggregates of live, but inconsistent with estimates and anecdotal evidence about increasing teenage pregnancy during the pandemic. This discrepancy could be explained by underreporting of new births or delays in reporting during the pandemic. Hence, we decided to estimate the increase of adolescent births during the pandemic by using the numbers suggested in academic studies, technical reports and media statements included in Table A8.

Table A7 – Fertility, female population and adolescent birth rates

Country Name	Fertility rate, total (births per woman) (1)		Female population (2)		Total births (3)=(1)*(2)	
	2019	*2020*	*2019*	*2020*	*2019*	*2020*
Ethiopia	4.1	4.0	56,009,719	57,446,748	232,216,295	232,601,883
Ghana	3.8	3.8	15,001,772	15,322,946	57,246,762	57,690,892
Kenya	3.4	3.4	26,451,585	27,052,773	90,543,775	91,032,581
Liberia	4.2	4.2	2,456,258	2,515,138	10,431,728	10,520,822
Malawi	4.1	4.1	9,443,471	9,695,918	38,973,205	39,316,947
Nigeria	5.3	5.2	99,131,729	101,669,950	527,083,403	533,563,898
Tanzania	4.8	4.8	29,024,840	29,883,105	140,248,027	142,632,060
South Africa	2.4	2.4	29,698,965	30,092,678	70,713,236	70,958,535

Country Name	Adolescent fertility rate (births per 1,000 women ages 15-19) (4)		Population ages 15-19, female (% of female population) (5)		Total population between 15-19 (6)=(5)*(2)		Adolescent births (7)=[(6)*(4)/1000]		Adolescent births/ Adolescent population (8)=(7)/(6)	
	2019	*2020*	*2019*	*2020*	*2019*	*2020*	*2019*	*2020*	*2019*	*2020*
Ethiopia	63.4	61.8	11.3	11.2	6,343,543	6,432,598	402,324	397,344	6.3%	6.2%
Ghana	65.0	64.1	10.1	10.0	1,508,797	1,533,921	98,008	98,371	6.5%	6.4%
Kenya	73.0	72.0	11.0	11.0	2,917,853	2,988,795	213,032	215,122	7.3%	7.2%
Liberia	135.2	134.8	10.8	10.8	264,243	271,580	35,718	36,602	13.5%	13.5%
Malawi	131.5	130.9	11.2	11.2	1,053,390	1,088,192	138,540	142,490	13.2%	13.1%
Nigeria	103.6	101.7	10.5	10.6	10,420,692	10,758,964	1,079,209	1,093,982	10.4%	10.2%
Tanzania	115.5	114.0	10.6	10.7	3,075,455	3,192,431	355,114	363,963	11.5%	11.4%
South Africa	67.8	67.7	8.1	8.1	2,413,807	2,440,831	163,631	165,318	6.8%	6.8%

Source: Own elaboration based on World Bank Data. Female population: https://data.worldbank.org/indicator/SP.POP.TOTL.FE.IN. Female population between 15-19 as % of the female population: https://data.worldbank.org/indicator/SP.POP.1519.FE.5Y. Adolescent fertility rate (births per 1,000 women ages 15-19): https://data.worldbank.org/indicator/SP.ADO.TFRT. Fertility rate, total (births per woman): https://data.worldbank.org/indicator/SP.DYN.TFRT

Table A8 – Predicted increase in teenage pregnancy during 2020

Country	Change in teenage pregnancy during 2020
Ethiopia	74.7%
Kenya	40%
Malawi	35%
South Africa	58%

Source: Own elaboration based on NGO's and press reports.
Ethiopia: https://www.dovepress.com/getfile.php?fileID=69564.
Kenya: https://www.globalcitizen.org/en/content/rise-in-teenage-pregnancies-during-kenyalockdown/
https://www.africanews.com/2020/06/17/close-to-4000-school-girls-impregnated-in-kenya-during-covid-19-lockdown/
https://www.devex.com/news/dramatic-rise-in-kenya-early-pregnancies-amid-school-closures-irc-data-suggests-97921
Malawi: https://www.unicef.org/esa/media/7626/file/COVID-19-A%20Catastrophe-for-Children-in-SSA.pdf
https://www.researchgate.net/publication/354469739_International_Journal_of_Women's_Health_Care_Teenage_Pregnancy_During_a_Pandemic
South Africa: https://reliefweb.int/report/south-africa/teen-pregnancies-south-africa-jump-60-during-covid-19-pandemic
https://news-decoder.com/teenage-pregnancies-soar-in-africa-as-schools-shut-for-covid/

Table A9 shows adolescent birth rates in 2019 and predicted adolescent birth rates in 2020 in each country. Given that not all countries had information about the increase in teenage pregnancy during 2020, we impute it for the remaining countries using three different criterias:

(i) The increase in each country is imputed using the country with the most similar ratio between adolescent birth and the total adolescent population. Column 2020 (2).
(ii) The increase in each country is imputed using the country with the most similar GDP per capita PPP in 2020. Column 2020 (3).
(iii) The increase in each country (even those where we found information about the increase in teenage pregnancy) is imputed using the increase reported in Ethiopia (75%). Column 2020 (4).

We use scenario (iii) to obtain our main results. The results, which are available upon request, slightly change when the other scenarios are used, but all applications are consistent with the main analysis.

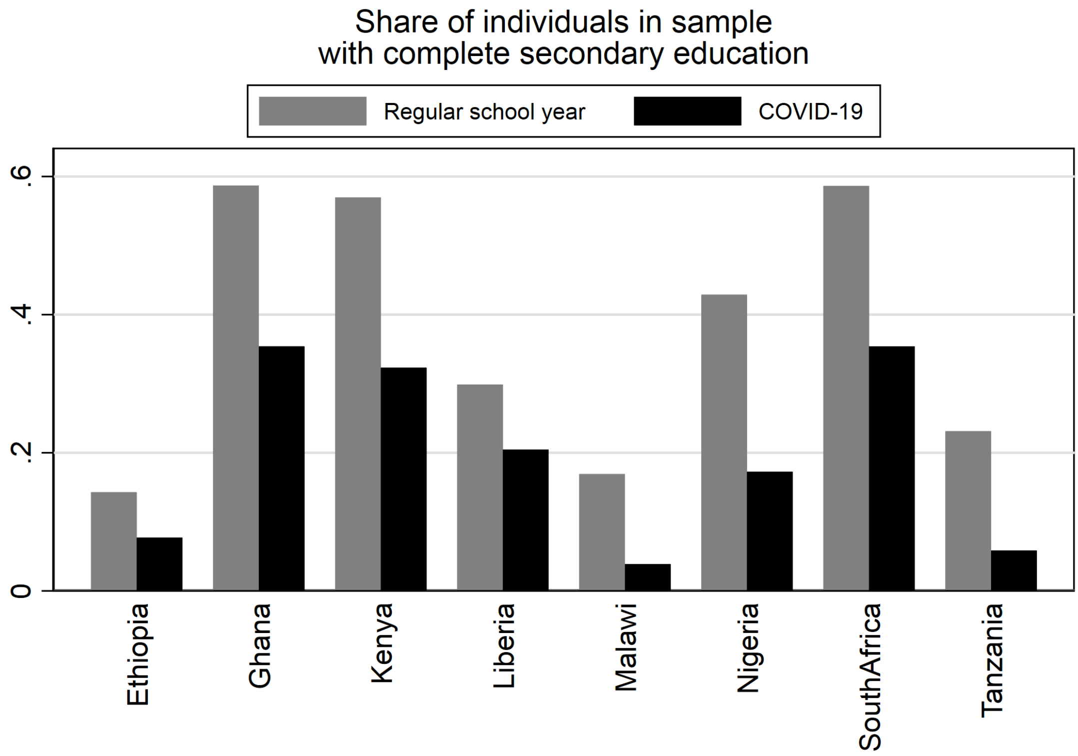
Table A9 – Adolescent birth rates increase

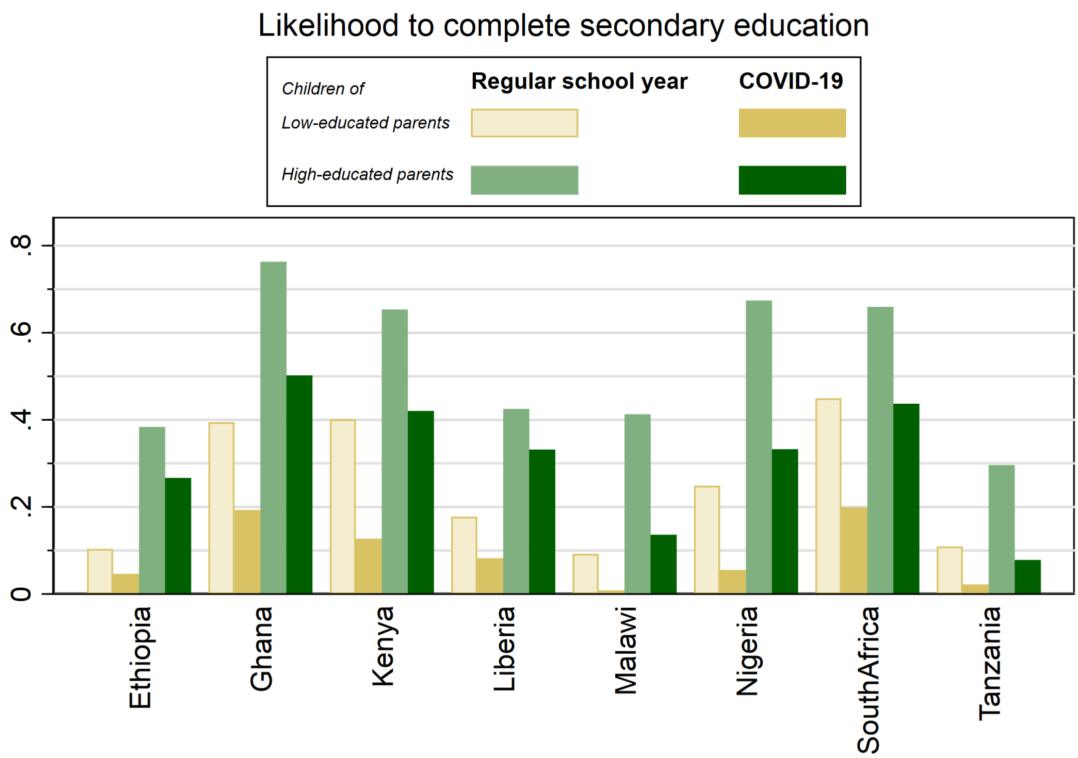
Country Name	Adolescent population		births/Adolescent		Increase in p.p (5)=(2)-(1)	Increase in p.p (6)=(3)-(1)	Increase in p.p (7)=(4)-(1)
	2019 (1)	*2020 (2)*	*2020 (3)*	*2020 (4)*			
Ethiopia	6.3%	10.9%	10.9%	10.9%	4.6	4.6	4.6
Ghana	6.5%	11.2%	8.9%	11.2%	4.7	2.4	4.7
Kenya	7.3%	10.0%	10.0%	12.4%	2.7	2.7	5.1
Liberia	13.5%	17.8%	17.8%	23.0%	4.2	4.2	9.5
Malawi	13.2%	17.2%	17.2%	22.2%	4.0	4.0	9.1
Nigeria	10.4%	13.5%	14.0%	17.5%	3.2	3.7	7.2
Tanzania	11.5%	15.0%	19.4%	19.4%	3.5	7.9	7.9
South Africa	6.8%	10.6%	10.6%	11.7%	3.8	3.8	4.9

Source: Own elaboration based on World Bank Data.

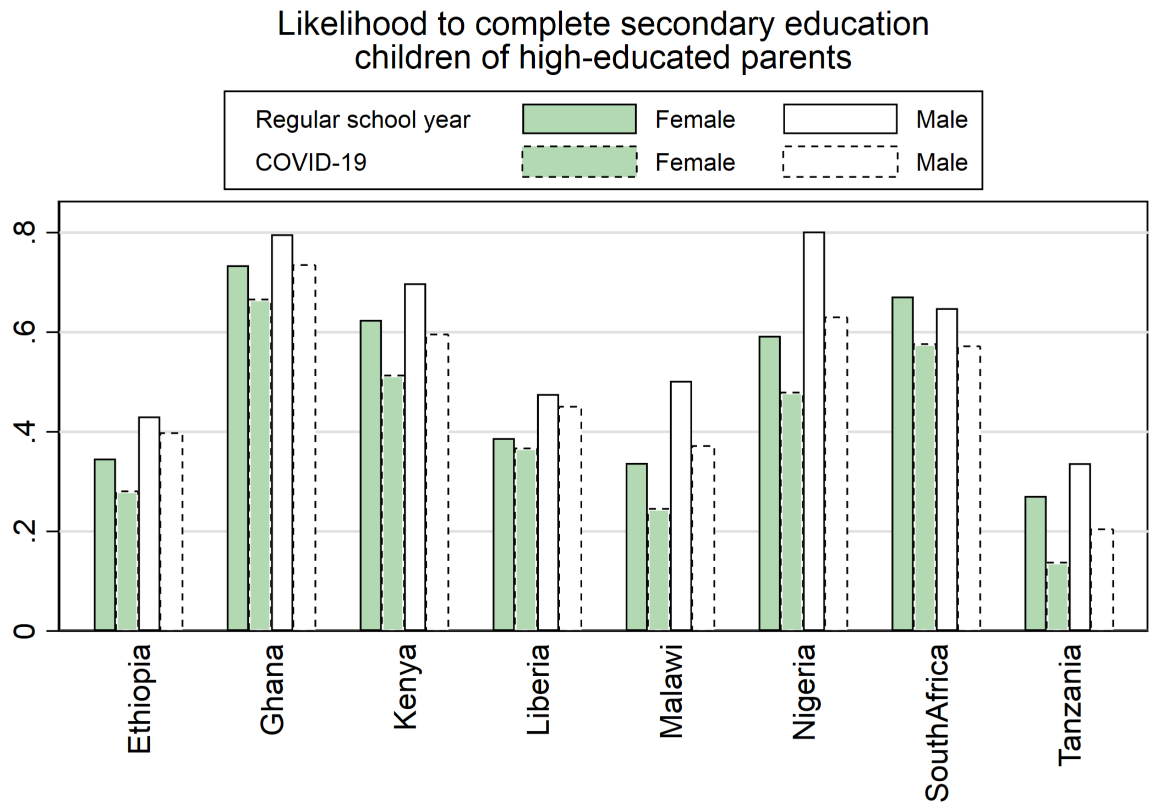
APPENDIX B: Additional analyses and robustness checks

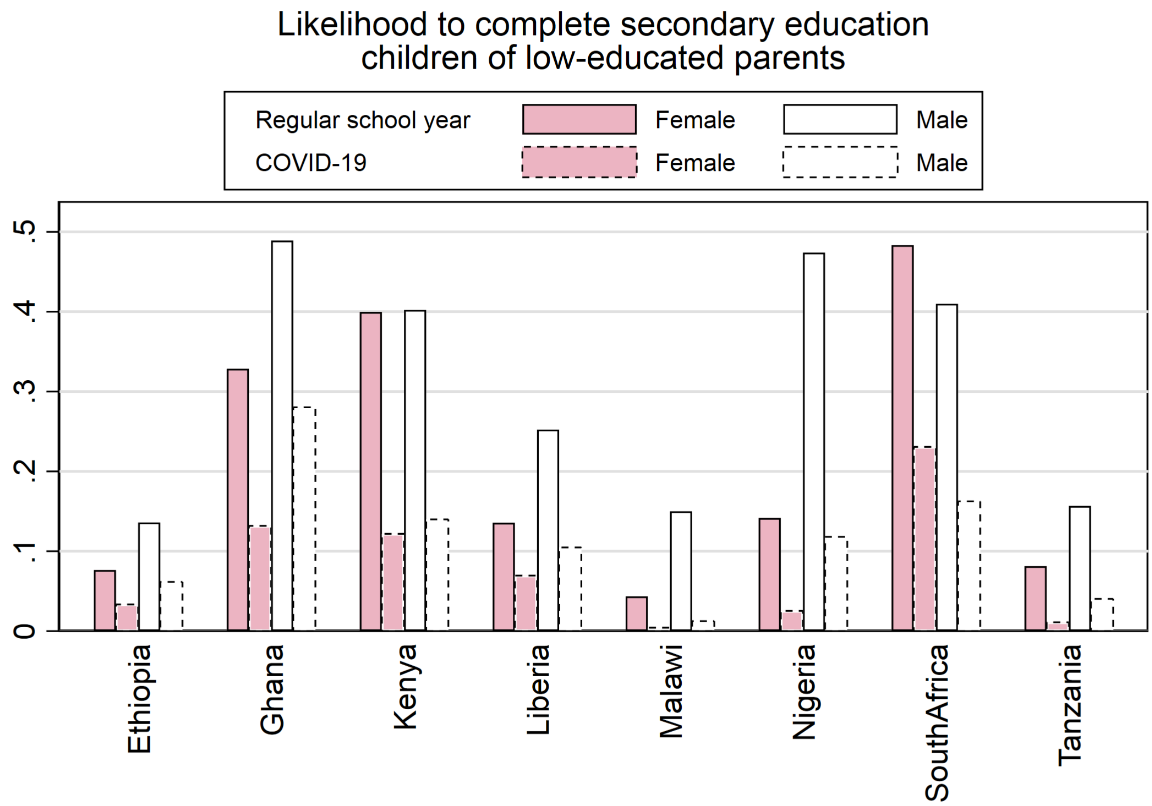
B1 – Higher bound estimates (attributing instructional loss to share equivalent to the parental factor of substitution)



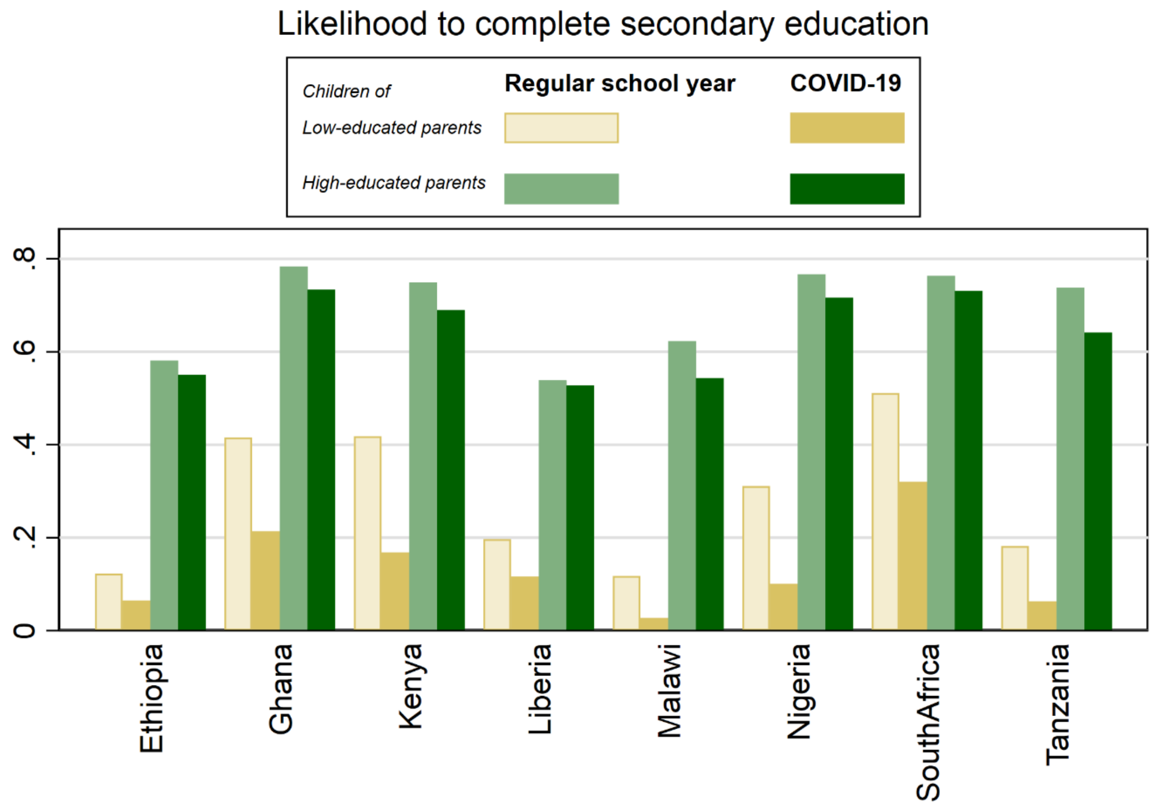


B2 – Gender differences

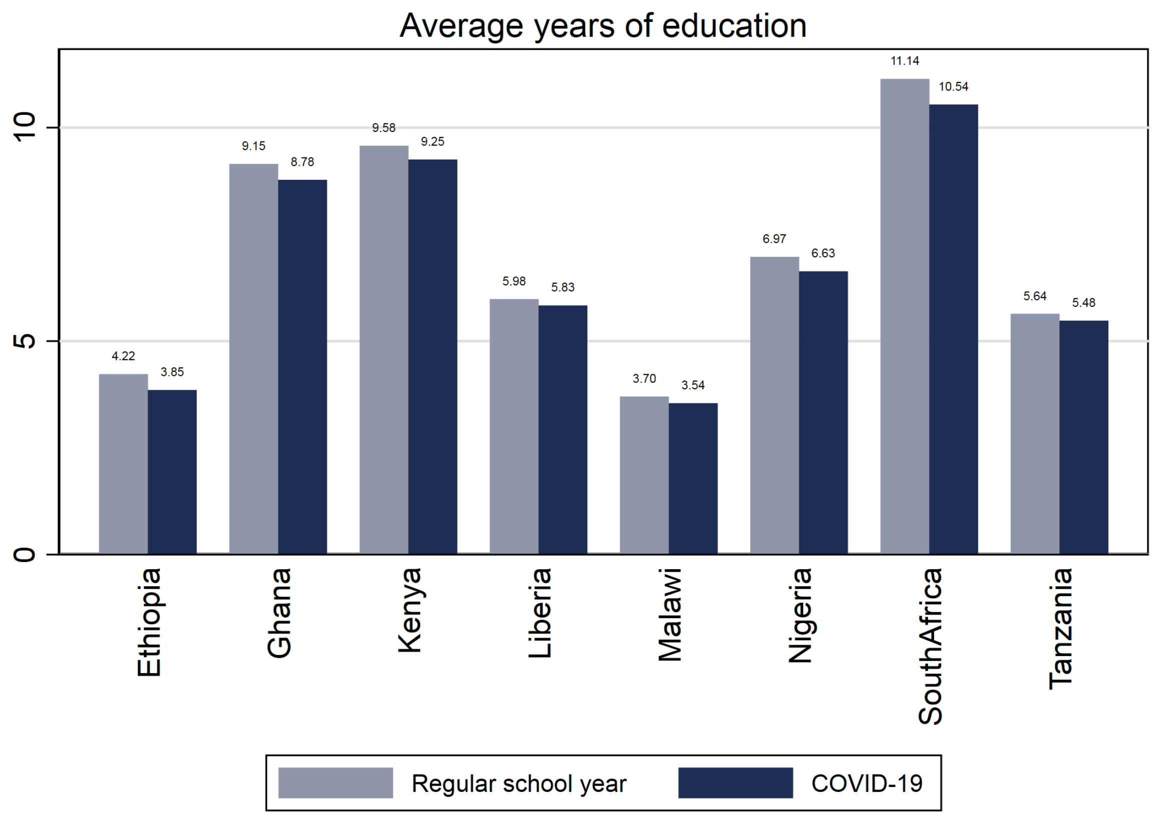




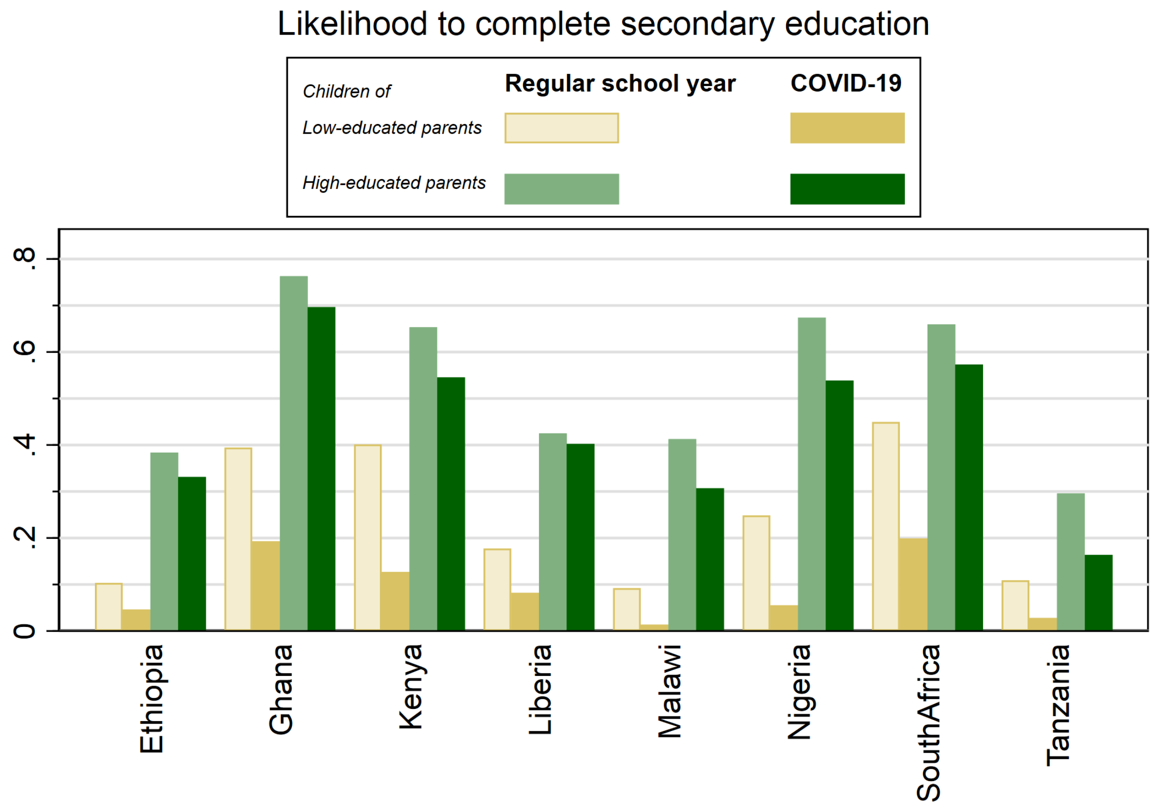
B3 – Different threshold for high-educated parents (completed secondary schooling)

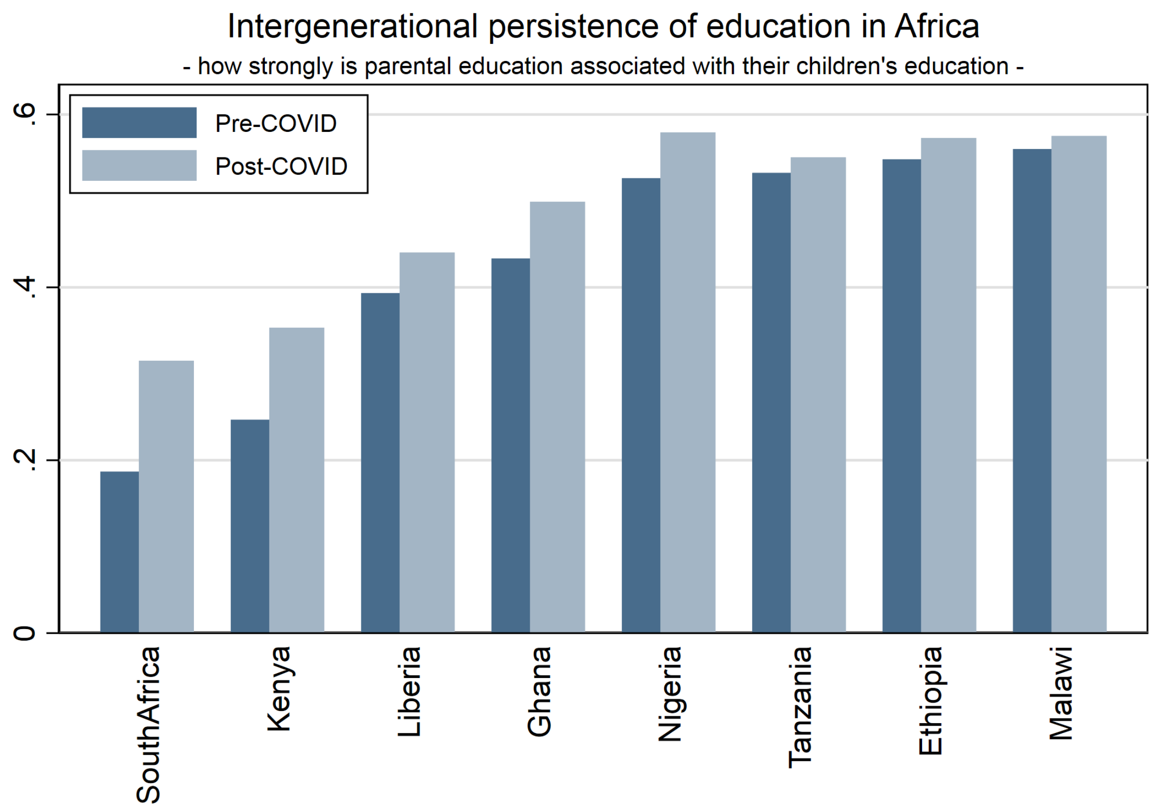


B4 – Change in average years of education following baseline scenario

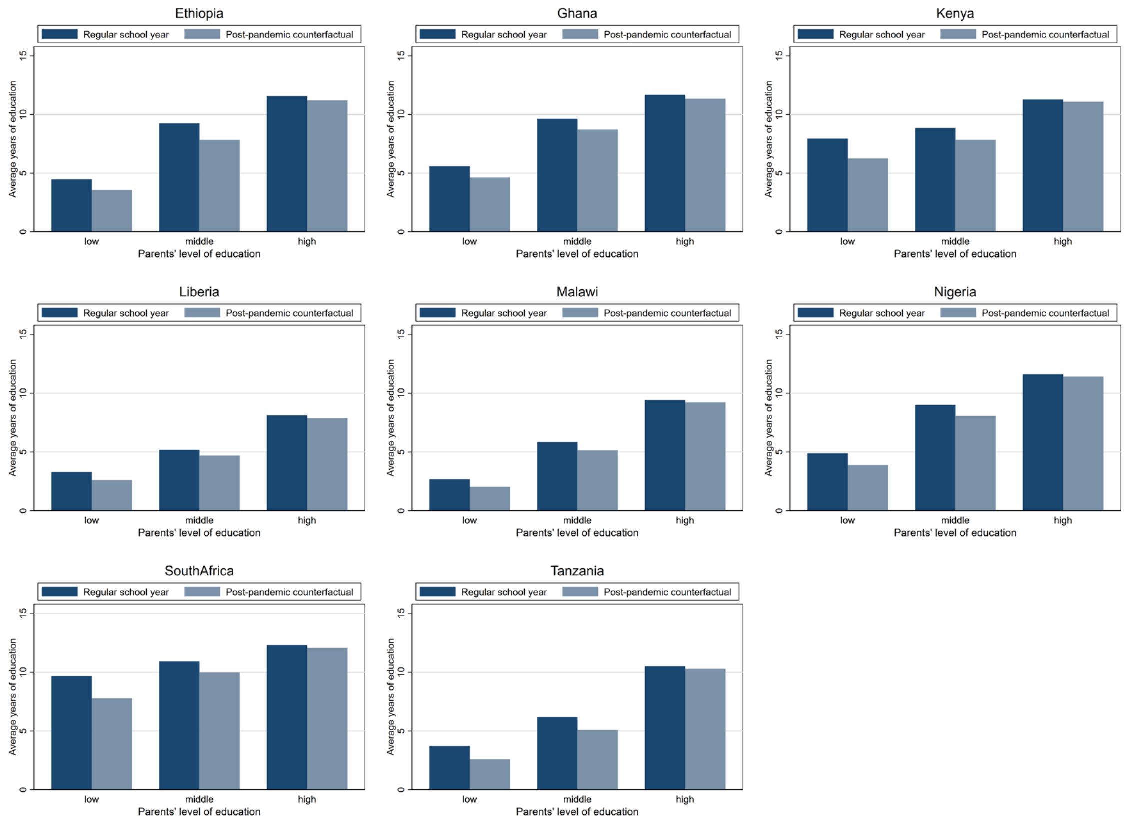


B5 – Intergenerational mobility estimates following the cumulative losses scenario



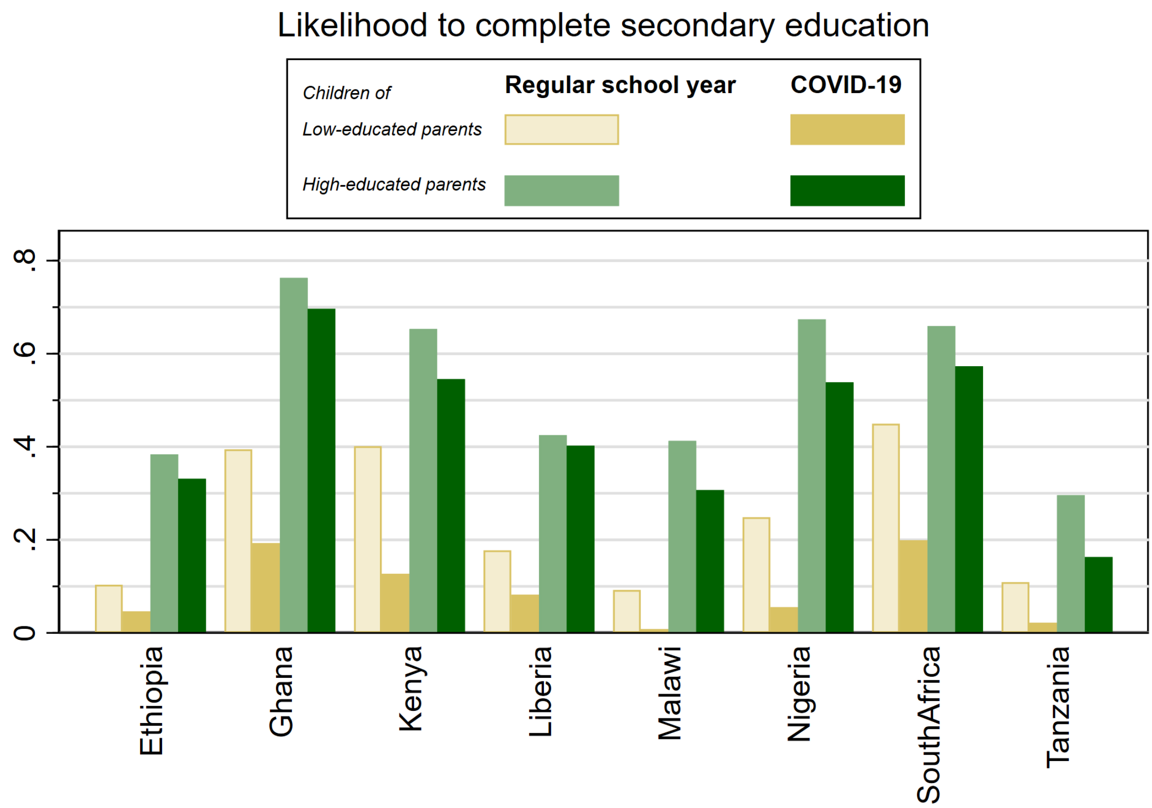


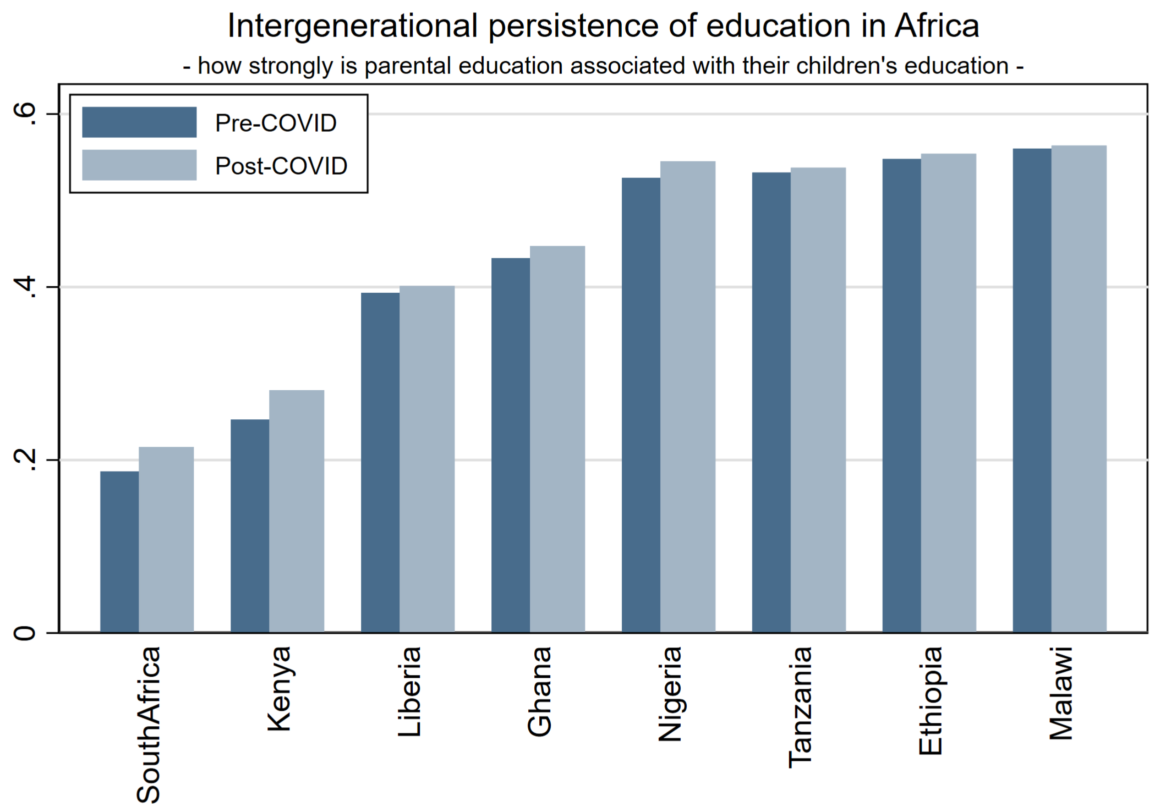
B6 – Impact on average education by parental background (considering cumulative instructional losses)



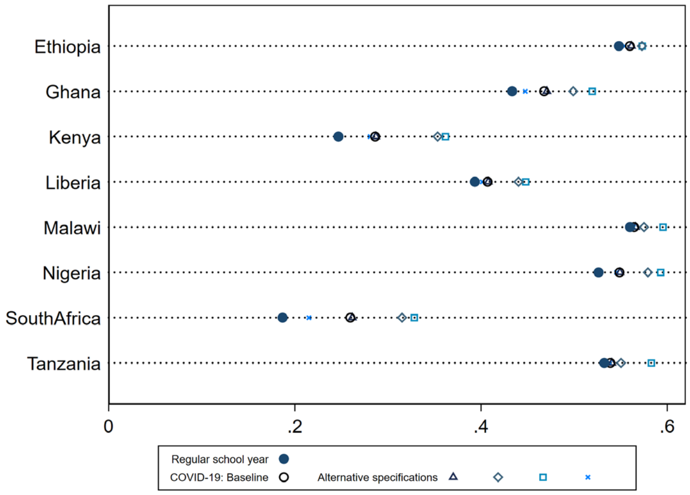
Notes: Parents' level of education is defined as low for parents with incomplete primary education or lower, as middle for parents with a completed primary degree but no completed secondary degree, and high for parents with a completed secondary degree or more.

B7 – Estimates considering only days of full (i.e. nationwide) school closures

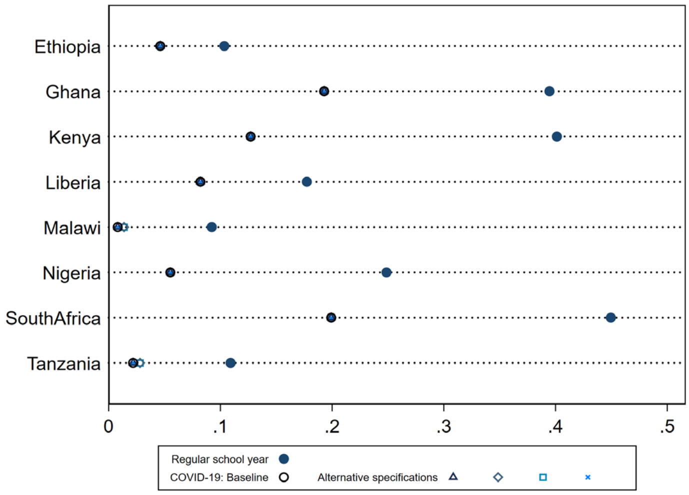




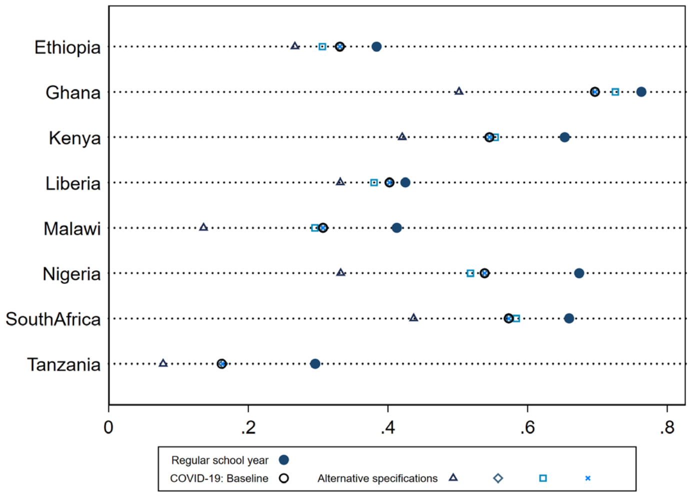
B8 – Point estimates for baseline and alternative estimations



(a) Slope coefficient



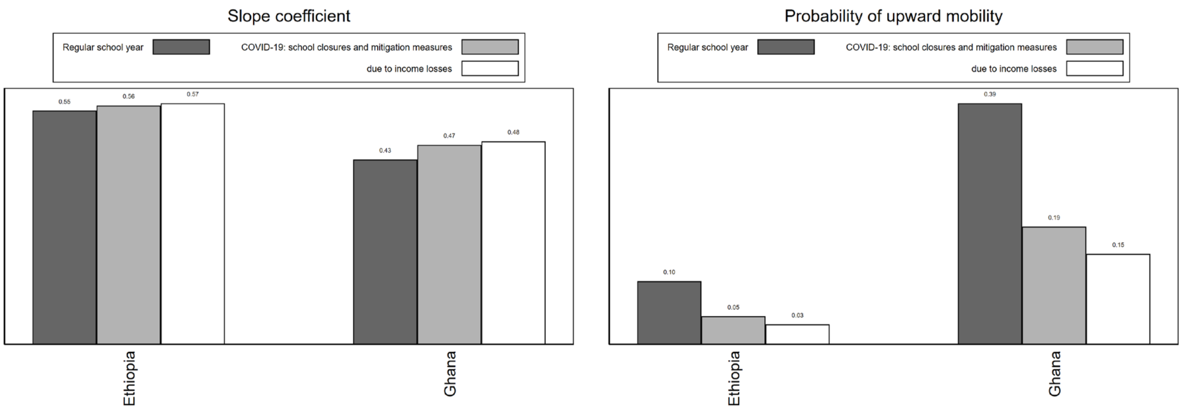
(b) Upward mobility



(c) Top persistence

B9 – Additional impact of income losses on intergenerational persistence

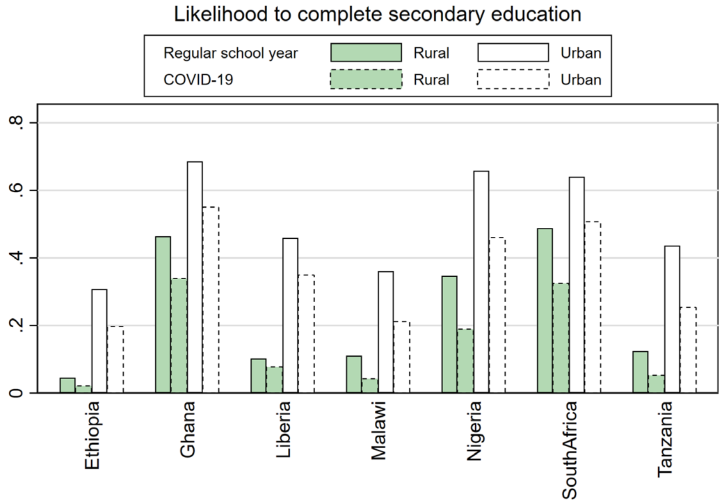
In line with Neidhöfer et al. (2021), we extend our analysis to include additional shocks that might impact investments in human capital among families. Previous research has shown that household income losses may lead to educational drop-out (e.g. Duryea et al., 2007; Cerutti et al., 2019; Thomas et al., 2004). However, the effect of income shocks on educational attainment is not straightforward as an economic crisis can decrease the opportunity cost of leaving school to enter the labor force, resulting in higher educational enrollment (Ferreira and Schady, 2009; Torche, 2010). Therefore, the overall effect of income shocks on educational attainment is ambiguous and may vary depending on family background characteristics. To account for the likelihood of educational drop-out being influenced by parental socioeconomic background, we set the probability of educational drop-out to one minus the parental factor of substitution. In case of income shocks, we assume that the individual drops out of education and, hence, loses two full years of schooling. This provides a counterfactual measure of years of schooling that takes into account the additional impact of household income shocks on educational attainment. To estimate the probability to suffer an income loss, we rely on estimates of the population share affected by income loss from other studies. For Ethiopia, we rely on Geda (2020), who estimates an income loss of 8% among employees, 22.2% among self-employed and 15.6% among members of small and micro-firms. We use the average among these values (15.3%). For Ghana, the estimate of Issahaku and Abu (2020) suggests that 26% of the national population is affected. The estimates for the slope coefficient and the probability of upward mobility are shown in the graph below. These examples suggest that, while the predicted change in intergenerational persistence is mostly driven by the instructional loss caused by school closures, income losses may also contribute to a significant additional rise in intergenerational persistence.

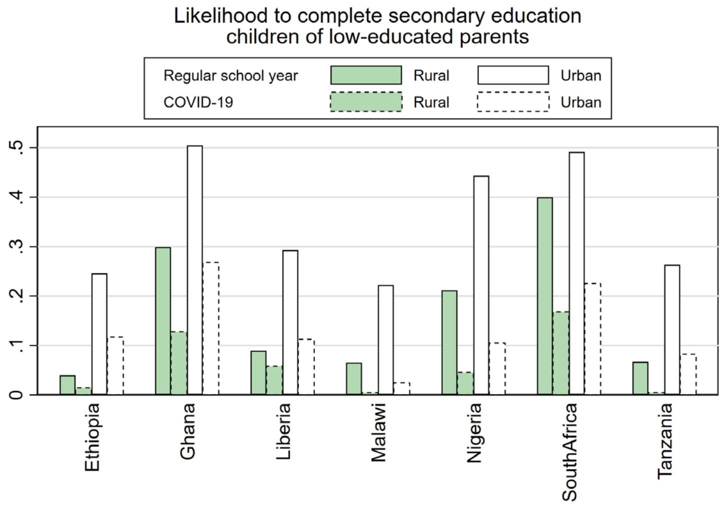


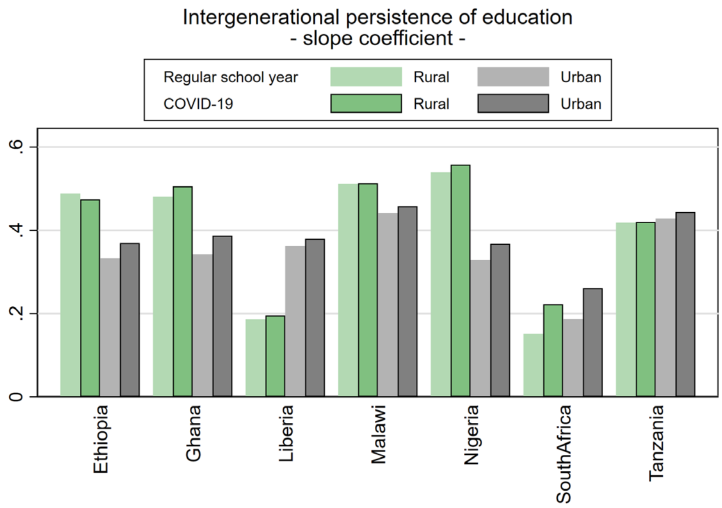
B10 – Simulation results for slope coefficient, intergenerational correlation and rank-rank slope

	Slope coefficient		Intergenerational correlation		Rank-rank slope	
	Baseline	COVID	Baseline	COVID	Baseline	COVID
Ethiopia	0.548	0.560	0.369	0.405	0.416	0.444
Ghana	0.433	0.468	0.444	0.490	0.447	0.542
Kenya	0.247	0.286	0.347	0.400	0.396	0.520
Liberia	0.393	0.407	0.399	0.420	0.353	0.389
Malawi	0.560	0.564	0.480	0.499	0.445	0.481
Nigeria	0.526	0.549	0.515	0.552	0.510	0.592
SouthAfrica	0.187	0.260	0.311	0.414	0.308	0.469
Tanzania	0.532	0.539	0.432	0.443	0.400	0.458

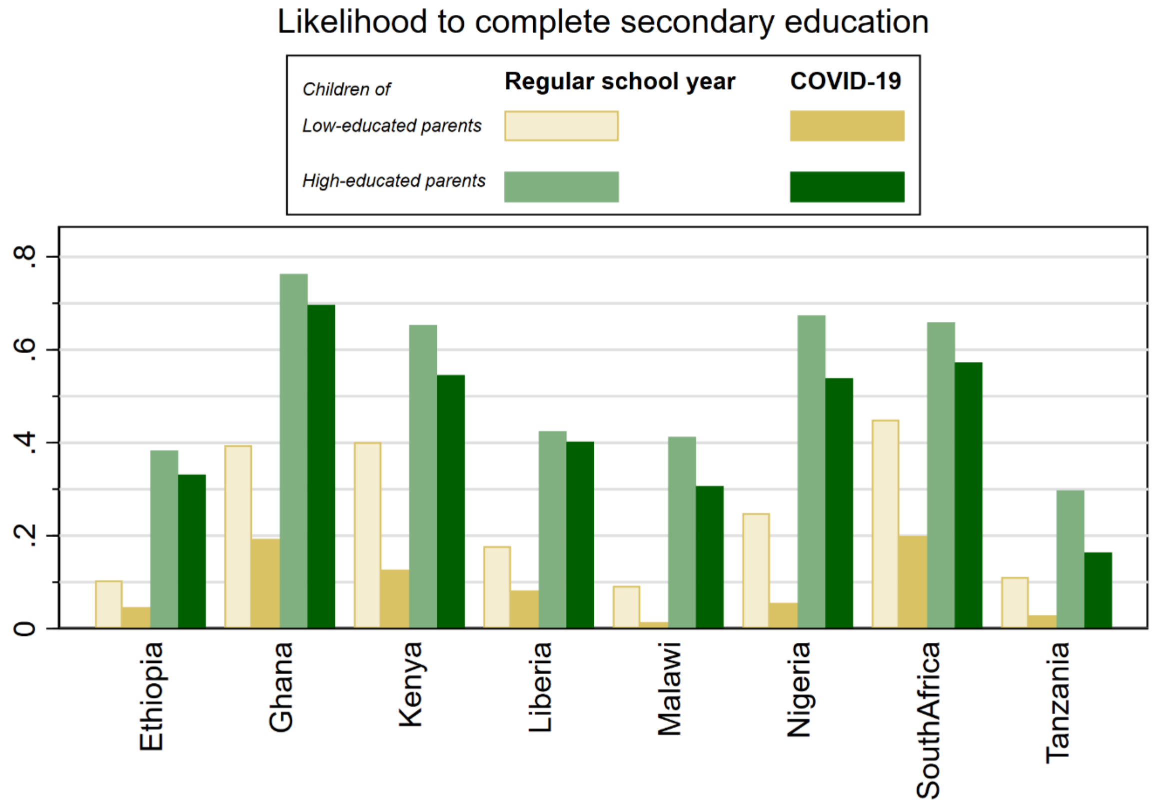
B11 – Results for Rural and Urban Areas

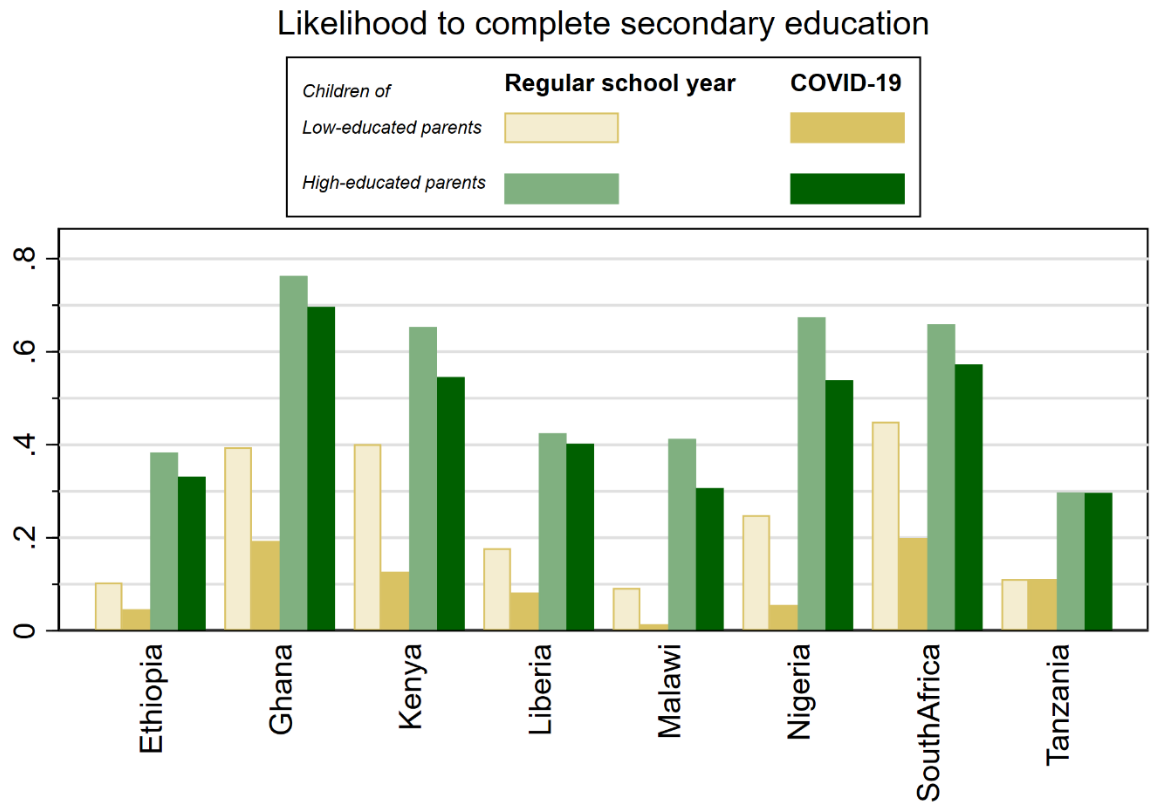






B12 – Estimates considering recovery in case of small instructional losses (upper graph: max. 25% of one school year; lower graph: max. 50% of one school year)





B13 – Reweighting of Sample to Match Composition by Parental Education

To project the impact of the COVID-19 pandemic on instructional loss and intergenerational mobility, we conducted simulations on a sample of individuals aged 18 and above who have completed their education. To ensure the relevance of our analysis to the pandemic's effects on current students, we focused on individuals born between 1987 and 1994. This birth year range represents the youngest respondents who fulfill our chosen age restriction. Given their proximity to the cohort that was in school when the pandemic disrupted education, we assume that the composition of our sampled individuals, in terms of factors such as gender and parental education, does not significantly differ. Furthermore, we assume that any potential differences in these factors do not influence the estimates. Essentially, we treat our sampled individuals as a laboratory, assuming they are the best available proxy for simulating the impact on current students.

Here, we relax this assumption and address the potential disparity in the composition of our sample compared to the cohorts in school during 2020. Specifically, we focus on the composition by parental education. The Sub-Saharan African region has experienced substantial increases in educational achievements, making it unlikely that the composition by parental background in our sample aligns with the composition of individuals currently in school. To address this concern, we performed a robustness check by reweighting the observations in our sample. This reweighting aimed to align the composition of the 1987-1994 cohort by parental education with the presumed composition of the 2002-2014 cohorts.

To initiate the reweighting process, we first defined the potential parents of individuals born between 2002 and 2014 and estimated their distribution of education. We assumed that the typical age range for giving birth falls between 15 and 45 years, covering the reproductive years for most individuals, accounting for cultural, socioeconomic, and individual factors.

Based on this assumption, we calculated the birth year range for the parents as follows:

- For individuals born in 2002:
  - Assuming a minimum age of 15 for the parents, their birth year would be 1987.
  - Assuming a maximum age of 45 for the parents, their birth year would be 1957.
- For individuals born in 2014:
  - Assuming a minimum age of 15 for the parents, their birth year would be 1999.
  - Assuming a maximum age of 45 for the parents, their birth year would be 1969.

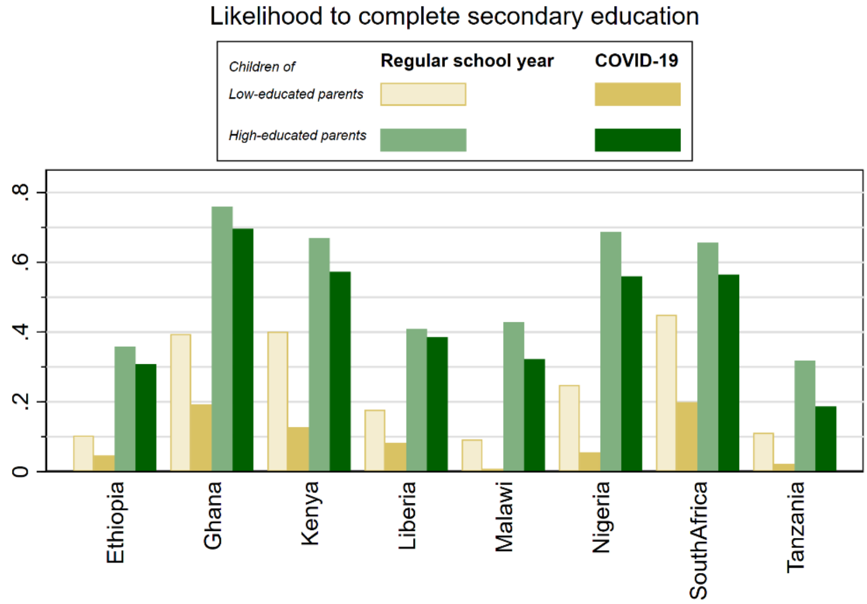
Consequently, the birth year range of the parents of individuals born between 2002 and 2014 in Sub-Saharan Africa is approximately between 1957 and 1999.

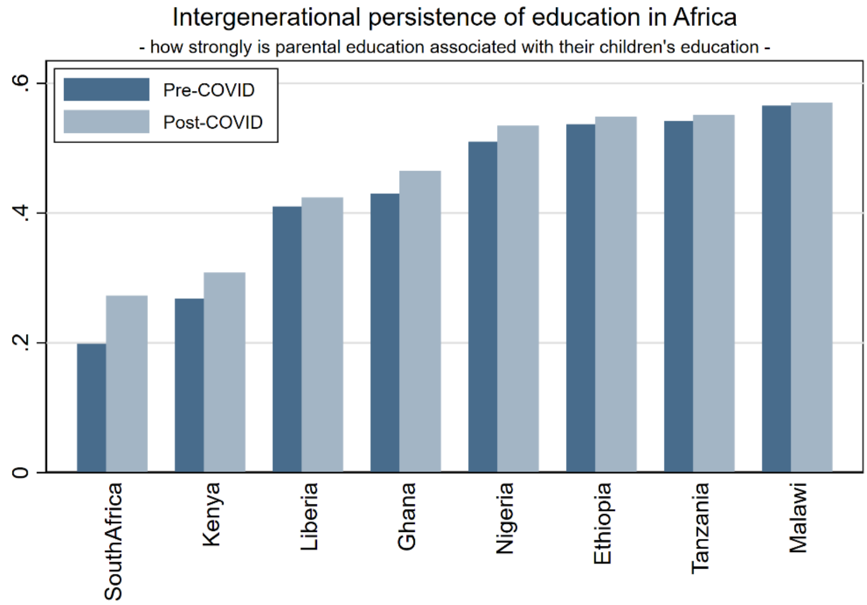
We then examined the data to verify the increase in education. The parents of the 1987-1994 cohort, on average, have lower education levels compared to those born between 1957 and 1999. The table below displays both distributions, confirming the disparity. With these distributions, we calculated reweighting factors as the ratio between the two shares for each parental education category.

Parental Education	Share in Cohort 2002-2014	Share in Cohort 1987-1994	Reweighting factor
No Schooling	34.89	50.64	0.689
Completed Primary	28.69	26.67	1.076
Incomplete Secondary	11.84	5.72	2.070
Completed Secondary	14.85	10.68	1.390
At least some Tertiary	9.72	6.3	1.543

We reweighted the observations by multiplying the survey design weights with the respective reweighting factor and applied the new weighting scheme to estimate the impact on the variables of interest. The figures below demonstrate that the estimates obtained through the reweighted sample are not substantially different from the main estimates. The overall effects indicate a slight weakening: intergenerational mobility decreased by 9% as a consequence of the pandemic, while secondary school completion rates fell by approximately 10%. These results remain consistent with the main analysis.

These reweighting efforts provide an additional assessment of the composition by parental education, ensuring that our analysis captures the potential impact of the pandemic on individuals currently in school. The consistent findings further strengthen the robustness of our main estimates, supporting the validity and reliability of our results.





APPENDIX C: Validation with alternative data

Beyond the fact that, to this day, some sources of information could help test the validity of our predictions, in the cases we consulted, this is not straightforward because the data for African countries is minimal. To the best of our knowledge, no new household surveys are available at the moment in the analysed countries. Furthermore, the information available in phone surveys is limited, and, in general, questions were not stable throughout the different waves. Therefore, we tried to assess the validity of our calibration and model using indirect observable variables available in different data sources, analysing whether the results are qualitatively consistent with our model and estimations. We observe that our simulation of learning losses correlates positively with days of instructional losses and additional health losses, and the opposite occurs with the mitigation policies and the parental factor of substitution, which captures parents' teaching, private tutoring, and any other factor that can affect the likelihood of receiving education during the pandemic (Figure C1-C4).

Given the lack of data on test scores (e.g. PISA) for the pandemic and post-pandemic period, we decided on a different approach relying on the World Bank phone surveys. This surveys provide information on the number of households whose children were involved in activities to recover lost face-to-face classes and allows us to distinguish what types of activities were carried out by households to recover missed school days. Thus, using this information, we can evaluate different features and assumptions of our model and calibration. To the best of our knowledge, the World Bank phone surveys are the only survey data available to date for the countries we analysed. However, this survey data has multiple limitations: The sample size is small, and some questions are not stable through the different waves and between countries. Also, most of the education-related questions were included in the survey only during the first half of 2020. With these potential limitations in mind, the results of our comparisons should be taken with caution.

As a first test, we assess if the percentage of households with children involved in learning recovery actions is correlated with our measure of learning losses. Assuming the former is a proxy of learning losses or positively correlated with it, we should observe a negative slope between these two variables if our model is correctly specified. Figure C1 shows that our estimate of learning losses and the share of households with children involved in learning recovery activities recorded in the phone surveys are (negatively) correlated across countries.

Similarly, if our modelling strategy is accurate, we should observe that the percentage of households not involved in recovery actions correlates with the online and offline learning indices we constructed and the duration of school closures. In Figure C2, the blue dots represent the proportion of households whose children were not involved in any action to recover lost classes. The orange bar indicates the duration of school closures as a percentage of regular school years, while the orange and yellow bars represent the values of the offline and online learning activity indices used in our analysis, respectively.

The indicators shown in the graph indicate that we can distinguish three groups of countries according to the duration of the school closure. In the first group, we find Ethiopia and Kenya with the two highest lost days compared to a regular year. However, these countries differ in the compensatory actions taken to mitigate school closures. While both countries focused on developing offline resources, only Kenya provided online resources to recover classes. This difference is correlated with two different results regarding how many children were not involved in learning recovery activities: While Kenya accomplished the lowest (22%) share of children not involved in learning recovery activities, Ethiopia obtained the highest number (81%). Nigeria is in the second group. It is an intermediate case where fewer actions were taken to mitigate the loss of classes, and the number of lost days was relatively less than in Ethiopia and Kenya. This combination resulted in a proportion of children who did not recover classes greater than Kenya's and less than Ethiopia's. Here, we can see that even though a country is not developing actions to mitigate learning losses, the length of the school closure may also play a role in defining the learning losses. Finally, in the case of Tanzania, more actions were taken to recover class days than in Ethiopia and Nigeria, and the number of lost days was the lowest of the four countries under analysis. This combination resulted in a proportion of people who did not recover classes lower than that of Ethiopia and Nigeria but higher than that achieved by Kenya. In summary, assuming that the proportion of people who are not involved in actions to recover lost classes suffers a learning loss, the results presented suggest that in countries where more school days were lost, the proportion of households not involved in actions to recover classes is higher, as long as the government takes no compensatory action to mitigate it.

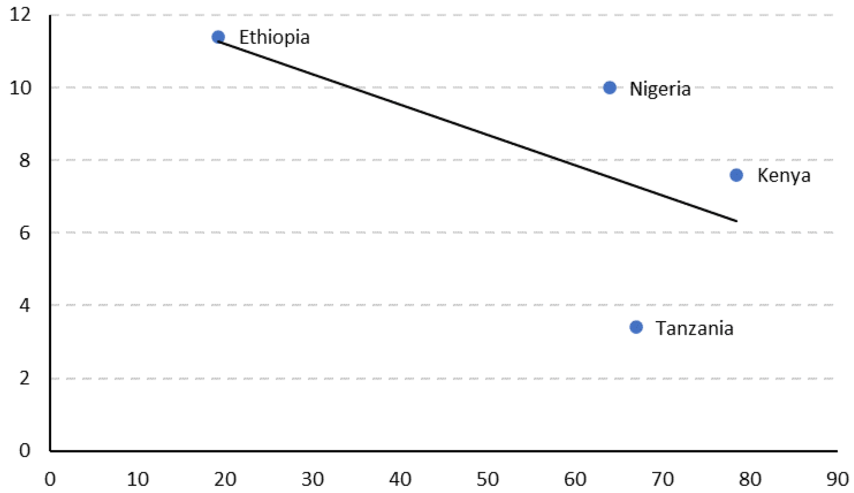
Additionally, in the cases of Kenya, Nigeria, and Tanzania, it is possible to distinguish the relative importance of each learning recovery activity. Consistent with the above, in the case of Kenya, almost 90% of households were involved in activities to recover classes through online learning methods. In the case of Nigeria, given the lack of mitigation actions conducted by the government, most households recovered lost class days through parent teaching schemes. This result emphasizes the relative importance of the parental factor substitution in our model. As Figure C3 shows, Nigeria got a relatively lower percentage of children with learning losses than Ethiopia, even though the mitigating actions were fewer than in this country. Finally, in Tanzania, actions related to offline learning activities were the most important among the households that conducted actions to recover lost classes.

These results suggest two aspects:

1. Our estimation of the online and offline learning indexes are positively correlated with the share of the households that took a given recovery action to mitigate learning losses, which suggests that our indicators are consistent.
2. They also confirm the importance of considering the parental factor of substitutions in our model. In those cases where compensatory actions were not widespread throughout the country, the household solutions to recover the learning losses were more relevant.

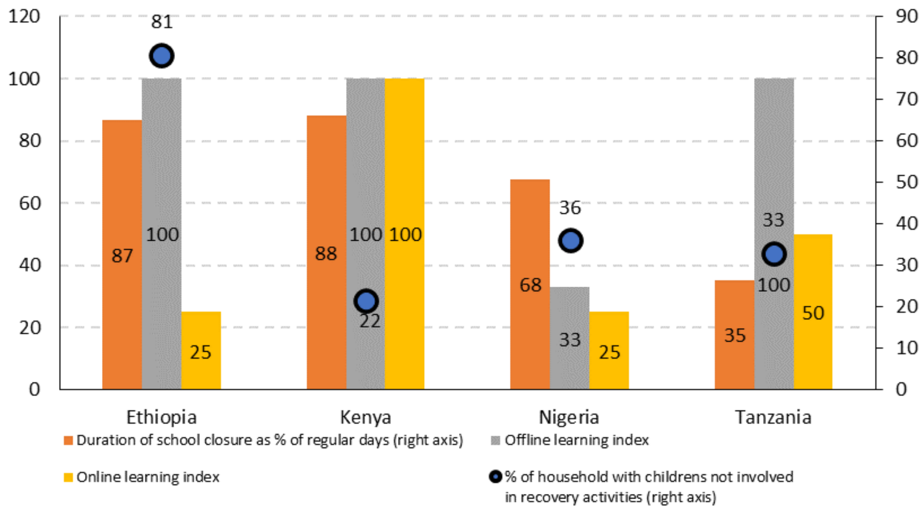
To stress the importance of parental background substitution, we present statistics from Kenya, the only country where we can identify parental education background in the phone surveys. Figure C4 shows the distribution of households by parental education background in each learning recovery activity. It suggests a balanced distribution in terms of access by parental education background (Online and offline learning recovery activities) for those actions where the government can be a supplier. In contrast, in the cases of parents teaching their children, or private tutoring, the results indicate that most households with access to these resources have highly educated parents. Our model captures this feature through the interaction of recovery measures with the parental education background substitution factor.

Figure C1 – Correlation of our estimate for κ and % of households involved in learning recovery activities from World Bank Phone Surveys



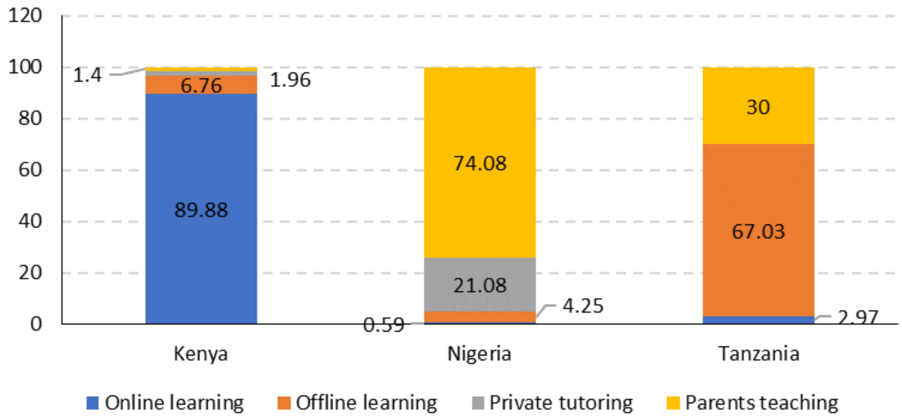
Notes: X Axis: % of households involved in learning recovery actitivies (WB Phone Surveys). Y Axis: Our estimate for κ (average instructional loss in % of two school years). Source: World Bank phone surveys for Ethiopia, Kenya, Nigeria, and Tanzania.

Figure C2 - Correlation of indicators used in our study and households involved in recovery activities from World Bank Phone Surveys



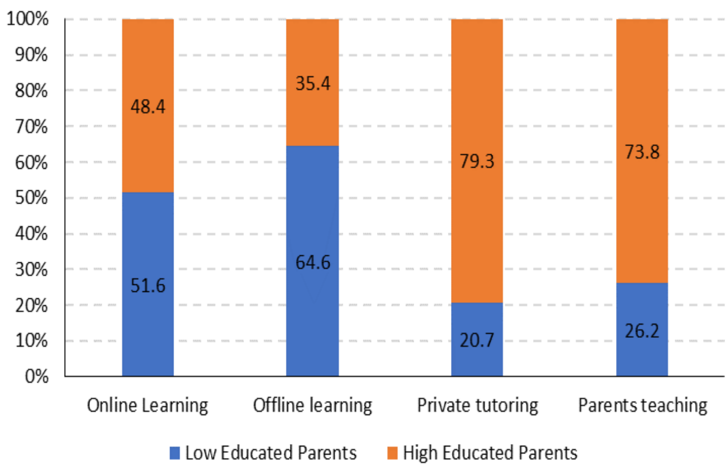
Source: World Bank phone surveys for Ethiopia, Kenya, Nigeria, and Tanzania.

Figure C3 - Distribution of recovery actions by type of action and country



Note: each number represents the share of the households that took a given recovery action over the pool of households whose children were involved in learning recovery action. Source: World Bank phone surveys Kenya, Nigeria and Tanzania.

Figure C4 - Composition of children undertaking learning recovery activities by type of activity and parental educational background



Note: Each bar shows the share of households involved in each learning recovery activity by parental education background. The graph corresponds to Kenya. Source: World Bank phone surveys Kenya.

References

Cerutti, P., Crivellaro, E., Reyes, G., and Sousa, L. D. 2019. Hit and Run? Income Shocks and School Dropouts in Latin America. Labour, 33(4): 533-566.

Duryea, S., Lam, D., and Levison, D. 2007. Effects of economic shocks on children's employment and schooling in Brazil. Journal of development economics, 84(1): 188-214.

Ferreira, F. H., & Schady, N. (2009). Aggregate economic shocks, child schooling, and child health. The World Bank Research Observer, 24(2), 147-181.

Thomas, D., Beegle, K., Frankenberg, E., Sikoki, B., Strauss, J., & Teruel, G. (2004). Education in a Crisis. Journal of Development economics, 74(1), 53-85.

Torche, F. (2010). Economic crisis and inequality of educational opportunity in Latin America. Sociology of education, 83(2), 85-110.