Does sorting matter for learning inequality? Evidence from East Africa

Paul Anand Jere R. Behrman Hai-Anh H. Dang Sam Jones*

December 2019

Abstract

Inequalities in children's learning are widely recognized to arise from variations in both household and school-related factors. While few studies have considered the role of sorting between schools and households, even fewer have quantified how much sorting contributes to educational inequalities in low- and middle-income countries. We fill this gap using data on over 1 million children from three East African countries. Applying a novel variance decomposition procedure, our results indicate that sorting of pupils across schools accounts for at least 8 percent of the total test-score variance, equivalent to half a year of schooling or more. This contribution tends to be largest for children from families at the ends of the socio-economic spectrum. Empirical simulations of steady-state educational inequalities reveal that policies to mitigate the consequences of sorting could substantially reduce inequalities in education.

*Jones (sam.jones@econ.ku.dk, corresponding author) is a Research Fellow with UNU-WIDER, Mozambique; Behrman is the William R. Kenan, Jr. Professor of Economics at the University of Pennsylvania; Dang is an economist in the Analytics and Tools Unit, Development Data Group, World Bank, and is also affiliated with Indiana University, IZA, and Vietnam's Academy of Social Sciences; Paul Anand is a Professor at the Open University and Research Associate at HERC in Oxford University. We would like to thank Jed Friedman, Ha Nguyen, Ron Smith, and participants at the Centre for the Study of African Economies Conference 2019 for helpful comments on earlier versions. We are grateful to the UK Department of International Development for funding assistance through its Strategic Research Program (SRP) program. A previous version of this paper was published as World Bank Policy Research Paper # 8622 under the title "Inequality of Opportunity in Education: Accounting for the Contributions of Sibs, Schools and Sorting across East Africa".

1 Introduction

Education plays a vital role in building human capital, helping individuals make productive contributions to the economy and live fulfilled lives. The international community has recently issued a joint call for all children to have access to free and equitable education through the secondary level by 2030. To achieve this ambitious goal, it is useful to have a better understanding of the complex processes that determine educational achievement, thereby permitting more effective policies to reduce educational inequalities.

A vast literature attests that both household and school factors are key determinants of educational achievement (e.g., Bowles, 1970; Björklund and Salvanes, 2011; Hanushek and Rivkin, 2012). Thus, unequal learning opportunities at either home or school can contribute to inequalities in final educational achievement. Moreover, home and school are unlikely to be independent. Both theoretical and empirical evidence suggests that children from disadvantaged families disproportionately attend lower-quality schools (Nechyba, 2006; Hsieh and Urquiola, 2006). Also, substantial evidence indicates that teachers serving schools in socio-economically deprived areas tend to be less experienced or less qualified, potentially limiting learning outcomes (Jackson, 2009; Clotfelter et al., 2011; Sass et al., 2012; OECD, 2018). Therefore, sorting between households and schools may accentuate educational inequalities. Consequently, policies to mitigate sorting may help narrow educational gaps between rich and poor households.

The vast bulk of evidence regarding educational sorting comes from high-income countries. However, in these contexts the significance of these processes remains unsettled. Kremer (1997) argues that eliminating neighbourhood segregation would decrease long-run educational inequality in the USA by less than two percent. In contrast, Fernández and Rogerson (2001) develop a model where enhanced sorting can have much larger effects on inequality (see also Fernandez, 2003); also, empirical evidence from Canada suggests that elimination of sorting either by home language or by parental schooling could reduce test-score variance as much as 40 percent in some subjects, at least in locations where school segregation is substantial (Friesen and Krauth, 2007).

To our knowledge, hardly any study has sought to quantify the contribution of sorting to educational inequalities in low- and middle-income countries (LMICs). But this is not because educational outcomes are highly equal. In fact, large achievement gaps running along socio-economic lines among pupils within LMICs have been extensively documented (e.g., Watkins, 2012). Also, (low cost) private schools have expanded rapidly in many LMICs over recent decades, increasing possibilities for school choice (Heyneman and

Stern, 2014). To fill the gap in the literature, we show how estimates of the variance contribution of sorting can be derived on the basis of separate estimates of the lower-bound contributions of household and school factors. We present and apply this approach to a large-scale database of test-scores covering over 1,000,000 school-aged children from three East African countries (Kenya, mainland Tanzania and Uganda). Not only do we find a relatively high degree of segregation along socio-economic lines across communities, but we also find the contribution of sorting of children across schools within communities is positive and accounts for around 8 percent of the total test-score variance. As such, almost a fifth of the joint variation in test-scores due to systematic supra-individual circumstances – i.e., due to schools, communities, and households – can be attributed to sorting. Since these circumstances relate closely to the notion of inequality of opportunities (Roemer, 2002), it follows that processes of sorting do have a material role in this domain. Further analysis of heterogeneity suggests that the same sorting contribution tends to be larger among families at both the top and bottom ends of the socio-economic distribution and among those sending their children to private schools, as well as in specific locations.

To explore the long-run implications of these results, we simulate how educational inequality evolves over time. These simulations show that for the average district in the region, the steady-state level of educational inequality would fall by around 10 percent if sorting were to be fully eliminated. While, this may not seem large at first glance, we show this is roughly equivalent to a 15 percent decline in the magnitude of the intergenerational persistence of education. Moreover, in a material number of districts, the reduction in inequality from eliminating sorting would be over 20 percent of the total variance and over 40 percent of inequality of opportunities. As such, this suggests that policies that take sorting into account, and even actively counteract it, merit attention.

The paper consists of six sections. In the next section, we outline a simple conceptual framework that guides our discussion of educational inequalities and points to the primitive components of the corresponding test-score variance decomposition. In Section 3, we show how these primitives can be estimated using standard econometric techniques, incidentally also yielding upper- and lower-bound estimates for the household and school components. For purposes of validation of these estimates we extend the estimation approach of Altonji and Mansfield (2018), which constitutes a more direct strategy to estimate the sorting component based on a set of constructed proxy variables. In Section 4 we describe the data; Section 5 presents the results, showing highly comparable results across the estimation methods; and Section 6 concludes.

2 Conceptual framework

A conventional point of departure for the analysis of inequality in education splits the proposed test-score generating process into the effects of households and schools in an educational production function of the following form:

$$t_{ijk} = f(h_j, s_k) + e_{ijk} \quad (1)$$

where t is a measure of educational achievement (e.g., test-scores), and indexes $i = (1, 2, \ldots, N)$, $j = (1, 2, \ldots, H)$ and $k = (1, 2, \ldots, S)$ refer to individual children, families and schools respectively. Following Bowles (1970) and others, h_j can be considered a comprehensive metric of the contribution of all factors shared by children in the same household (hereafter sibs) to test-scores; and s_k is a comprehensive metric of the contribution of the given school to their learning. Finally, e_{ijk} represents the remaining individual or idiosyncratic variation, which we assume is orthogonal to the household- and school-effects; i.e., $\mathrm{E}(e_{ijk}|s_k, h_j) = 0$.

To make this framework tractable for empirical analysis, two main elaborations are required. The first is to select a specific metric of inequality. While a variety of measures have been employed in the literature (e.g., Hanushek and Wößmann, 2006), we use the test-score variance. As Ferreira and Gignoux (2014) note, unlike other popular inequality measures, the variance is ordinally invariant to standardization procedures often used to express test-scores on a comparable scale. Furthermore, the linear additive nature of the variance makes it straightforward to isolate the contributions of individual factors (Shorrocks, 1982); and the variance also has the attractive property of sub-group decomposability (Chakravarty, 2001).

Second, we must place some structure on $f(\cdot)$, specifying how the school and household factors plausibly affect test scores. With respect to the levels expression in equation (1), we adopt a simple additive linear model. Critically, however, this specification does not pin-down the (co)variance structure without further assumptions. To see this, Table 1 describes four main cases, where each row invokes more specific assumptions about the level *and* variance of t. In the first row we assume the household and school factors make independent, additive contributions to outcomes. So, in terms of the associated variance, this imposes the assumption: $\mathrm{E}(s_k' h_j) = 0$, which rules out any correlation between the two factors.

The zero-covariance assumption embedded in Row 1 is restrictive. With respect to education, various sorting processes, including residential segregation, school choice (by parents) and

Table 1: Summary of alternative test-score data-generating processes

Model		Score level	Score variance	Description
1	Restricted linear	$h_j + s_k + e_{ijk}$	$\sigma_h^2 + \sigma_s^2 + \sigma_e^2$	Independent households & schools
2	Unrestricted linear	$h_j + s_k + e_{ijk}$	$\sigma_h^2 + \sigma_s^2 + 2\Sigma_{hs} + \sigma_e^2$	Correlated household & school factors
3	Household upper-bound	$h_j + \gamma\bar{h}_{jk} + \nu_k + e_{ijk}$	$(1+\gamma)^2\sigma_h^2 + \sigma_\nu^2 + \sigma_e^2$	Household effects partly absorb school effects
4	School upper-bound	$s_k + \theta\bar{s}_{kj} + \omega_j + e_{ijk}$	$(1+\theta)^2\sigma_s^2 + \sigma_\omega^2 + \sigma_e^2$	School effects partly absorb household effects

Note: score variance in Rows 3 and 4 assume $\sigma_h^2 \equiv \text{Var}(h_j) \approx \text{Var}(\bar{h}_{jk})$ and $\text{Var}(s_k) \approx \text{Var}(\bar{s}_{kj})$.

teacher allocation rules, have all been identified as potential determinants of schooling outcomes (e.g., Fernandez, 2003; Nechyba, 2006; Hanushek and Yilmaz, 2007) In such cases, the restriction in Row 1 is untenable, and an unrestricted linear model may apply (Row 2), in which any such sorting is captured via the covariance between household- and school-effects.

An interpretation consistent with the unrestricted linear (sorting) model is that the household and school factors have no direct mutual effects – i.e., both effects are separately pre-determined but become correlated through *ex post* processes of sorting or assortative matching. However, this is not the only mechanism that could generate a correlation between these factors. Some part of the school effect may reflect the causal effect of (average) constituent households, such as when households make direct financial or time commitments to school functioning. This kind of mechanism is also suggested by versions of cream-skimming models, where average peer quality in a school (or class) is driven by household characteristics, which in turn directly influences individual achievement (Walsh, 2009). An extreme version of this is captured in the third row of Table 1, which assumes s can be partitioned into a component that is oblique or parallel to h (by construction) and an orthogonal component ν, with own variance σ_ν^2:

$$s_k = \gamma \frac{1}{N_k} \sum_{\forall j | K=k} h_{jK} + \nu_k = \gamma \bar{h}_{jk} + \nu_k \tag{2}$$

and with $E(\bar{h}'_{jk}\nu_k) = 0$.

Applying this expression to the unrestricted linear model, Row 3 gives a strict upper-bound on the variance contribution due to households. The corollary is given in Row 4, where household-effects are assumed to be (partial) reflections of given school-effects, plus an orthogonal component. Note that in both these cases, the observed covariance between household- and school-effects is attributed wholly to one of the factors and no remaining sorting is allowed (by construction). As such, and as we clarify further below, the models in Rows 3–4 of Table 1 provide bounds on the variance components of interest, including an absolute upper-bound on sorting.

Existing literature has frequently estimated inequality of opportunities via some variant of the household upper-bound model. Concretely, various studies treat family effects as a single fixed unobserved factor and omit any consideration of school-effects. Björklund and Salvanes (2011) describe this approach and show how, under this set-up, the relative variance contribution of households equals the correlation in outcomes between siblings. They summarize estimates of sibling correlations in various developed countries, ranging from 0.24 in former East Germany to over 0.60 in the USA. While many of these estimates are based on grades of completed schooling, Mazumder (2011) estimates sibling correlations across learning domains for children in the USA. His estimates are of the same broad magnitude, ranging from approximately 0.35 to 0.50. For the UK, Nicoletti and Rabe (2013) analyze results from compulsory national tests and find somewhat larger sibling correlations (>0.50). Estimates of sibling correlations in developing countries are scarce, mainly reflecting data constraints. Exceptions are Behrman et al. (2001), who find the sib correlation across Latin American countries in terms of completed years of schooling, lies between around 0.30 and 0.60; and Emran and Shilpi (2015), who find a sibling correlation of around 60 percent in years of schooling among children in India (see also Hertz et al., 2007).

Variation in school (or teacher) effects point to differences in school quality. Many studies seek to assess the magnitude of these effects (e.g., Pritchett and Viarengo, 2015; Sass et al., 2012; Hanushek and Rivkin, 2006), in some cases also controlling for the contribution of family background to avoid confounding. For instance, Freeman and Viarengo (2014) use PISA data to investigate the (sources of) variance in school-effects. They report that a regression of test-scores on school dummies alone (as per the school upper-bound model) explains around two-thirds of the variation in the data, while a limited set of observed family background variables accounts for just one-third, after controlling for school-effects. Dang and Glewwe (2018??) find that schools and communities explain around 40 percent of the variation in either test-scores or years of schooling among Vietnamese children. However, studies of this sort remain rare for LMIC contexts and hardly any explicitly estimate the complete variance contributions of *both* schools and households, including their covariance (for an exception, see Carneiro, 2008).[1] As such, the direction and magnitude of sorting – defined generically as the factor covariance – has been largely neglected.

3 Decomposition

3.1 Methods

To address the gap in the literature we propose a multi-factor variance decomposition. The aim is to identify the main components of educational inequalities, which amounts to apportioning the variance across households and schools without imposing any particular covariance structure *a priori*. In doing so, and following Gibbons et al. (2014), we seek to identify bounds on the variance contributions of households and schools plus the sorting component.

The main analytical insight, from which the empirical methods flow, is that the various models set out in Table 1 can all be (re)stated in terms of a set of primitive quantities. As we explain further below, this means it is not necessary to estimate the correlated variance components (σ_h^2, σ_s^2) directly. To see this, note that in relation to equation (1), the lower-bound or uncorrelated variance contribution of any given factor directly relates to its raw or unadjusted contribution via the pairwise-correlation coefficient (ρ_{sh}). For example, using the upper-bound household model, the lower-bound on the school component can be rewritten as: $\sigma_\nu^2 = (1 - \rho_{hs}^2)\sigma_s^2$ (see Appendix A for derivation). In other words, the uncorrelated or lower-bound contribution due to schooling is proportional to one minus the square of the correlation coefficient between the two factors. An equivalent expression relates the lower-bound and raw-variance contributions of the household factor.

Using this insight, we can then rewrite the unrestricted linear model as:

$$\sigma_t^2 = \frac{\sigma_\omega^2 + \sigma_\nu^2 + 2\rho_{hs}\sigma_\omega\sigma_\nu}{1 - \rho_{hs}^2} + \sigma_e^2 \tag{3a}$$

$$\Leftrightarrow 2\rho_{hs}\sigma_\omega\sigma_\nu = (1 - \rho_{hs}^2)(\sigma_t^2 - \sigma_e^2) - (\sigma_\omega^2 + \sigma_\nu^2) \tag{3b}$$

It follows that once three specific quantities are known – namely, the two uncorrelated variance shares and the total variance jointly attributable to the two latent factors given

[1] By 'complete', we refer to both observed and unobserved aspects of each factor. In this sense, where studies rely on a limited set of observed proxies for any single effect, the observed component only can be expected to represent a part of the overall variance associated with the factor of interest (Ferreira and Gignoux, 2014), thereby constituting a lower-bound.

by $(\sigma_t^2 - \sigma_e^2)$ – then the unknown correlation coefficient can be obtained as the root to equation (3b). Furthermore, and as summarised in Appendix Table D1, *any* of the variance decomposition components of interest, including the contribution of sorting, can be calculated from the three primitives $(\sigma_\omega^2, \sigma_\nu^2, \rho_{hs})$. Thus, the empirical objective of the varaince decomposition is to estimate these primitives.

Before proceeding to implementation, it is important to note that households and schools are encountered in the same locations and, thus, may share a common community (neighbourhood) component. To capture this explicitly, the household and school effects can be operationalised as follows:

$$z_{jkl} = c_l + h_{jl} + s_{kl} \tag{4}$$

where z represents the joint contribution of schools and households in location l; and c is the location-specific effect. Admittedly, the nature and magnitude of these types of effects remains controversial (e.g., Oreopoulos, 2003). However, there is growing evidence that the quality of local environments can have a material influence on child development trajectories (Chetty et al., 2016; Chetty and Hendren, 2018). Moreover, ignoring the contribution of the latter term would mean they are simply absorbed by the upper-bound household or school effect estimates, in turn muddying the interpretation of the variance decomposition components. Consequently, as detailed below, in our empirical implementations we estimate the implied community effects separately.[2]

3.2 Implementation

To estimate the primitives presented above, we first adopt a simple fixed-effects indirect approach (denoted FEi). As suggested by Table 1, the uncorrelated (lower-bound) school and household variance contributions can be obtained from household and school upper-bound models respectively. In fact, following Solon et al. (2000) (also Raaum et al., 2006), a regression of the test-scores on a set of household dummies (only) will capture not just the stand-alone household-effect but also the effect of all unobserved variables correlated with this factor, including sorting – i.e., following equation (2), the household dummies will absorb the joint contribution of h_j and $\bar{h}_{jk}$; and, based on these estimates, residual variation aggregated to the school-level will be orthogonal to all included factors by construction, yielding the uncorrelated (lower-bound) school factor (ν). Using a set of school dummies, the same approach provides estimates for the uncorrelated (lower-bound)

[2] In doing so, we impose that the estimated household/school-effects are orthogonal to the community effects, implying the estimates of sorting will not be confounded by selection of households or schools *across* communities. As such, the estimated contribution of sorting (defined above) should be interpreted in a narrow or within-community sense.

household variance contribution; and the associated contribution of sorting is calculated on the basis of equation (3b). Step-by-step details of this procedure are set out in Appendix B.

A drawback of using such fixed-effects procedures is that estimates of latent factors generally include measurement error, meaning the corresponding raw-variance shares will be upwards biased (Koedel et al., 2015). To address this, we use empirical Bayes shrinkage, which involves adjusting each estimated effect toward a common prior by a factor proportional to the estimated noise-to-signal ratio in the original estimates. Following Stanek et al. (1999), we shrink each estimated fixed effect (e.g., $\hat{h}_j$) toward a global mean as follows:

$$\tilde{h}_j = \bar{h}_j + \psi(\hat{h}_j - \bar{h}_j) \tag{5}$$
$$\text{with shrinkage factor } \; 0 < \psi = \frac{\sigma^2_{\hat{h}}}{\sigma^2_{\hat{h}} + \sigma^2_{\hat{\epsilon}}/N_j} < 1$$

Here, N_j is the effective degrees of freedom available to estimate each of the j effects; $\sigma^2_{\hat{h}}$ is the raw-variance of the estimated household fixed effect; $\sigma^2_{\hat{\epsilon}}$ is the estimated overall residual variance; and $\bar{h}_j$ is the sample fixed-effect mean, typically zero under conventional normalization restrictions.

A second approach to dealing with the presence of sorting is suggested by Altonji and Mansfield (2018) (hereafter, AM18). Motivated by a concern that sorting on unobserved variables may bias estimates of the effects of group-level inputs, such as schools, these authors show that where households with different characteristics also have differing effective preferences (willingness-to-pay) for underlying school or community amenities, group averages of observed household or pupil characteristics can be employed as control functions. Inclusion of these generated variables, henceforth referred to as sorting proxies, in a regression model thus permits identification of a lower-bound on the unique (variance) contribution of school factors. In this set up, the specification of interest becomes:

$$t_{ijkl} = h'_{O,j}\beta + \bar{h}'_{O,k}\gamma + [h_{jl}] + [s_{kl}] + e_{ijkl} \tag{6}$$

where $h_{O,j}$ are observed household-level covariates; $\bar{h}_{O,k}$ are the set of sorting proxies. Treating the terms in brackets as (uncorrelated) household and school random-effects, the specification can be estimated using a linear mixed-effects estimator (denoted MLM; as used by AM18).

An advantage of this more direct procedure is that it relies on a single estimation equation. A drawback is that the random-effects are assumed to be orthogonal to the set of included covariates, implying equation (6) may be misspecified.[3] Furthermore, while the

[3] For households, a 'tighter' lower-bound variance share is obtained by adding the estimated variance of the

variances of the random effects are estimated as model parameters, their associated best-linear-unbiased predictors (BLUPs or conditional modes) typically do not share the same variance-covariance structure as the estimated population-level moments (see Morris, 2002). Consequently, the BLUPs may not be reliable for the purposes of investigating systematic patterns in the random effects (e.g., subgroup heterogeneity).[4] Consequently, as an additional validation procedure, we extend the AM18 approach to incorporate fixed- as opposed to random-effects terms. As described in Appendix B, this involves modifying the indirect estimation procedure whereby the sorting proxies are used directly to account for variation in test-scores before estimation of the (one-way) household and school fixed-effects. This procedure is denoted FEd.

Finally, it merits clarification why we do *not* seek to estimate the (correlated) variance components from the outset (σ_h^2, σ_s^2). While this might seem more straightforward, these factors cannot be identified easily. Treating them as conventional random-effects requires imposing a zero pairwise covariance restriction, ruling out the presence of sorting. This limitation is not shared by two-way fixed-effects models, but a downside of these methods is that the pairwise correlation of the estimated fixed-effects vectors tends to be biased downwards (Abowd et al., 2002). As Andrews et al. (2008) demonstrate, this is driven by a quasi-mechanical relation, whereby if one factor (e.g., household-effects) is over-estimated then on average the other factor (e.g., schools) will be under-estimated (also Andrews et al., 2012).[5] Furthermore, while methods to correct for the negative covariance bias in two-way fixed-effects estimators have been proposed (see Gaure, 2014), they remain work in progress.

4 Data

Since 2010, the Uwezo initiative has undertaken large-scale household-based surveys of academic achievement in Kenya, mainland Tanzania and Uganda.[6] The surveys target children residing in households aged between the official starting-school age and 16 and are representative at both national and district levels. Excluding the initial surveys, five

fitted observed component to the variance of the random component.

[4] Estimation of high-dimensional mixed-effects models also is highly computationally intensive and prone to convergence problems. In the present case we use the `lme4` package in `R` for estimation (Bates, 2010).

[5] Intuitively, this reflects the general problem of model over-parameterization; and the magnitude of bias tends to be larger where fewer observations are available to estimate each effect, which is particularly relevant here as the latent effects are highly granular (e.g., households).

[6] The approach adopted by Uwezo has been inspired by exercises carried out in India by the Assessment Survey Evaluation Research Centre (ASER). For further details and comparison to other regional assessments see Uwezo (2012); Jones et al. (2014).

rounds of the Uwezo surveys are publicly available (2011-2015) and used here. For each household, the surveys collected information covering general characteristics, as well as the demographic and educational details of resident children (e.g., age, gender, whether or not attending school, etc.). Also, all children of school age were individually administered a set of basic oral literacy and numeracy tests, which were tailored to each country and varied by survey round based on a common template to reflect competencies stipulated in the national curriculum at the grade 2 level.

The literacy and numeracy tests (the Uwezo tests) are described in detail in Jones et al. (2014). The literacy tests refer to national languages of instruction in which pupils are tested at the end of primary school – i.e., English and Kiswahili in Tanzania and Kenya; and just English in Uganda. Importantly, the Uwezo tests are not adapted to the children's ages or their completed level of schooling. Given that they focus on basic competencies, it is thus unsurprising there are strong age- related differences, which affect both the level and variance of scores between age cohorts. From the present perspective, this between-cohort variation can be considered unwanted noise (see Mazumder, 2008). As a result, so as to construct an overall metric of achievement, we transform the raw integer scores on the individual tests in two steps. First, we apply a graded response IRT model to the suite of tests answered by each child, yielding an estimated achievement score (the empirical Bayes mean of the latent trait).[7] Next, to place the scores on a comparable scale, we standardize them to take a mean of zero and standard deviation of one within each survey, country and age group. This removes all unwanted variation in the score levels that would add noise to the decomposition; also, since the same Uwezo test forms are administered to children of all ages in each household, it accounts for the fact that the raw score standard deviation is not constant across age groups.

Table 2 reports regional means and standard deviations of the test-scores for the sample used in the present analysis, pooling data from different rounds. Note that due to the relatively basic competency levels assessed by the Uwezo tests, we exclude children above 14 years old, as they tend to perform at the upper end of the tests (and show much lower variance). The analytical sample also excludes observations that can be perfectly predicted using either household or school fixed-effects – i.e., singletons have been removed. The first (column I) of the table reports weighted means of the raw competency tests (ordinal scores); column II reports the IRT scores standardized by country and round, but not age; and column III reports the final measures, including age standardization. As can be seen,

[7] Further details on request. Due to differences in test forms between countries and across years, we estimate these models separately by country and survey round. Also, in a small number of cases the graded response model did not converge and, instead, we used a partial credit model. The test-scores derived from the IRT procedure are extremely highly correlated with those constructed from a conventional standardization.

movement from the second to the third metric constitutes a simple monotone transformation. Also, as per the methodological discussion of Section 3, the analytical focus is on the variance components of the test-score; and there is no evidence to suggest these dropped observations are distributed in a systematic pattern over regions or districts. The (sample) standard deviations of the test-scores are reported in parentheses in the table.[8]

To implement the decomposition procedures, household and school indexes must be defined. The former is trivial – unique indexes are ascribed to all siblings in the same household (in each year).[9] The school-effects are less straightforward. In the present data, limited information about schools is provided. Nonetheless, we know the kind of school attended (none, public or private) and whether or not children attend the main public schools in their local community (catchment area). Consequently, for each enumeration area, we categorise children into four school categories: (1) those not attending school; (2) those attending the specific (known, matched) local public schools; (3) those attending other public schools; and (4) those attending private schools. The advantage of this procedure is that, within each household, children can be associated to different school-effects – i.e., the school and household-effects are crossed. A downside is that for the last two school categories we do not identify specific schools; as such, these effects capture average school quality of a given type.

Further descriptive statistics for the dataset are reported in Table 3. This shows the number of unique children (i), households (j) and schools (k) covered in the dataset. Additionally, the table reports summary statistics, including average child characteristics and schooling status indicators (those out of school, the shares attending the specific matched public schools, and those attending private schools). Overall, these indicate the sample is comprehensive and balanced (by age and gender).[10] It also reveals there are systematic differences in schooling among countries as well as across regions within each country – e.g., in all countries enrolment rates differ substantially across regions (e.g., from 80.0 to 96.5 percent in Kenya) as do the shares of children attending private school.

[8] In line with Ferreira and Gignoux (2014), the rank position of each region according to its test-score variance is largely preserved, regardless of the transformation applied.

[9] The Uwezo surveys are cross-sectional in nature and no attempt is made to track the same children over time.

[10] Average ages are higher in Tanzania as the starting school age is seven, compared to six in the other countries.

Table 2: Metrics of achievement (test-scores), by country & region

Country & Region		Raw score		IRT		IRT std.	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
KE	Central	4.68	(1.63)	0.38	(0.86)	0.45	(0.83)
	Coast	3.71	(2.04)	-0.10	(1.01)	-0.12	(1.01)
	Eastern	4.06	(1.91)	0.04	(0.97)	0.03	(0.95)
	North Eastern	3.21	(2.04)	-0.36	(1.03)	-0.36	(1.13)
	Nyanza	3.87	(1.97)	-0.06	(0.99)	-0.08	(0.95)
	Rift Valley	3.89	(2.01)	-0.03	(1.02)	-0.03	(1.03)
	Western	3.66	(2.02)	-0.15	(1.01)	-0.19	(0.98)
	All	3.95	(1.97)	-0.00	(1.00)	0.00	(1.00)
TZ	Arusha	3.10	(1.81)	0.19	(0.99)	0.22	(1.00)
	Dar Es Salaam	3.45	(1.70)	0.36	(0.95)	0.40	(0.95)
	Iringa	2.92	(1.82)	0.09	(1.00)	0.09	(1.00)
	Kagera	2.67	(1.80)	-0.08	(0.98)	-0.09	(0.98)
	Kigoma	2.28	(1.81)	-0.22	(0.97)	-0.24	(0.96)
	Ruvuma	2.82	(1.77)	0.01	(0.97)	-0.00	(0.96)
	Singida	2.84	(1.84)	0.06	(1.00)	0.07	(0.98)
	Tabora	2.36	(1.84)	-0.23	(1.00)	-0.25	(1.00)
	Tanga	2.86	(1.80)	0.03	(0.98)	0.03	(0.96)
	All	2.77	(1.83)	0.00	(1.00)	0.00	(1.00)
UG	Central	2.93	(1.82)	0.29	(0.95)	0.34	(0.96)
	Eastern	2.17	(1.70)	-0.13	(0.97)	-0.16	(0.94)
	Northern	1.95	(1.66)	-0.28	(0.99)	-0.32	(0.99)
	Western	2.55	(1.81)	0.08	(1.00)	0.09	(0.99)
	All	2.42	(1.79)	0.00	(1.00)	-0.00	(1.00)

Note: achievement refers to literacy and numeracy, combined; regions in Tanzania and Kenya are aggregated for clarity of presentation (see Appendix C); KE is Kenya; TZ is Tanzania (mainland); UG is Uganda; test-scores combine achievement in literacy and numeracy, as described in the text; ‘IRT’ are the achievement scores estimated via IRT models; ‘IRT std.’ are the same scores standardized by age, country and survey round; all survey rounds are pooled.

Table 3: Descriptive statistics, by country & region

Country & Region		Index count i	j	k	Age	Female	School status None	Match?	Private
KE	Central	25,337	11,055	4,756	9.6	52.0	3.5	43.3	26.7
	Coast	34,667	13,349	4,332	9.6	50.3	15.0	40.5	18.5
	Eastern	56,916	23,173	7,597	9.6	50.3	5.7	59.7	6.4
	North Eastern	40,239	14,489	3,031	9.3	44.9	20.0	55.2	3.6
	Nyanza	52,681	20,895	7,487	9.6	50.2	8.5	50.5	13.9
	Rift Valley	119,236	46,896	14,481	9.5	49.4	10.6	50.8	12.0
	Western	57,200	22,063	7,149	9.6	50.2	7.6	52.9	8.0
	All	386,276	151,920	48,833	9.6	50.0	9.0	50.3	13.5
TZ	Arusha	32,535	13,738	4,358	10.0	48.8	10.4	53.4	6.0
	Dar Es Salaam	13,795	5,867	2,032	10.0	51.6	7.7	48.2	5.7
	Iringa	29,351	12,876	4,254	10.0	50.7	13.6	61.9	3.9
	Kagera	32,664	13,103	4,186	10.0	49.9	17.5	54.0	3.8
	Kigoma	20,994	8,700	2,626	10.0	50.4	20.9	52.5	5.4
	Ruvuma	17,514	7,780	2,809	10.1	50.4	10.0	66.2	2.8
	Singida	19,926	8,574	2,603	10.0	50.5	14.7	61.6	3.2
	Tabora	33,127	13,058	3,835	9.9	50.1	22.4	51.0	3.7
	Tanga	24,561	10,336	3,171	10.0	49.4	12.5	61.1	3.3
	All	224,467	94,032	29,874	10.0	50.2	15.3	56.3	4.2
UG	Central	46,474	16,954	6,016	9.5	49.9	5.1	24.3	48.2
	Eastern	86,761	30,557	8,707	9.6	49.8	4.3	44.7	23.9
	Northern	72,306	26,654	6,804	9.5	48.7	13.4	53.1	6.0
	Western	52,746	20,065	6,744	9.6	50.3	6.3	36.2	29.0
	All	258,287	94,230	28,271	9.5	49.7	6.9	38.9	28.0

Note: regions in Tanzania and Kenya are aggregated for clarity of presentation (see Appendix C); KE is Kenya; TZ is Tanzania (mainland); UG is Uganda; i, j, k refer to the number of unique observations for the individual, household and school-grade effects respectively; all other columns are means or proportions; 'match?' indicates the share of children attending the identified main local public school; all survey rounds are pooled.

5 Results

5.1 Spatial clustering

We begin with a preliminary review of the degree of spatial clustering or segregation – i.e., the extent to which similar types of children are found in the same schools or communities. To do so, we compare the correlation between individuals within different aggregate units, which is equivalent to the proportion of the variance attributable to the between-group structure of the data (for similar exercises see Fryer and Levitt, 2004; Friesen and Krauth, 2007, 2010; Lindahl, 2011; Emran and Shilpi, 2015). The magnitude of the correlation across members of the same group, and how quickly this falls as we move to higher levels of aggregation, thus indicates the extent to which variables are spatially segregated (clustered). For instance, if local communities only contained households with the exact same socio-economic status, then the proportion of variance in socio-economic status accounted for by communities would be equal to that accounted for by households, indicating a very high degree of clustering.

Results from this exercise are reported in Table 4, covering a range of variables. As might be expected, assuming child gender is approximately random, the between-group variance share accounted for by households is extremely low. For the remaining variables, however, clustering by households, schools or communities is much higher. For instance, more than two thirds of the variation in access to clean water is accounted for both by schools and by communities. Overall, a little more than half of the total variation in aggregate socio-economic status in the region (SES) is attributable to (average) differences between distinct communities, implying substantial levels of residential clustering or economic segregation.

Turning to the educational outcomes in the bottom of the table, we observe somewhat lower magnitudes of clustering across communities and schools, partly reflecting the presence of variation between children within households.[11] In terms of achievement on the Uwezo tests as measured by the age-adjusted IRT scores, the correlation between siblings is almost 50 percent, which is highly comparable to magnitudes found in a range of other countries (Section 2). More notably, however, the correlation between pupils attending the same schools, as well as between children in the same communities, is only moderately lower. Nearly 40 percent of the overall variation in achievement is accounted for by schools, and 30 percent between distinct residential locations. Although the definition of communities

[11] Of course, enrolment is fully accounted for by school-level variation; but there is within-household variation in enrolment.

Table 4: Percent of between-group variance attributable to alternative grouping structures

Group level →	Hhld.	Sch.	Vill.	District	Region	Ratios	
Variable	(a)	(b)	(c)	(d)	(e)	(b)/(a)	(c)/(b)
Female	3.0	3.0	1.7	0.1	0.0	1.01	0.56
Hhld. has electricity	100.0	54.0	47.5	19.1	7.4	0.54	0.88
Hhld. has clean water	100.0	70.5	66.9	13.4	2.8	0.70	0.95
Hhld. owns phone	100.0	38.9	33.0	10.3	5.4	0.39	0.85
Mother's education	100.0	52.7	46.4	13.6	6.7	0.53	0.88
Aggregate SES index	100.0	58.6	51.8	22.5	9.4	0.59	0.88
Enrolled	40.6	100.0	27.7	8.2	3.2	2.46	0.28
Highest grade	50.2	42.9	31.1	11.8	5.2	0.85	0.72
Math achivement (IRT)	42.6	31.9	25.4	7.3	2.9	0.75	0.80
Literacy achivement (IRT)	45.0	34.4	28.2	10.7	5.0	0.76	0.82
Overall achivement (IRT)	48.8	37.5	30.2	10.6	4.6	0.77	0.81

Note: cells in columns (a)–(e) report the correlation between children within the same grouping unit (e.g., households, schools, etc.), which is the adjusted R^2 from a one-way fixed-effects model, without covariates; for highest grade, this correlation is calculated conditioning on child age; final two columns report ratios.

is not equivalent across studies, these latter magnitudes appear to be a factor larger than encountered in developed countries. For example for pupils in the UK, Nicoletti and Rabe (2013) find a sibling correlation (the household upper-bound variance share) in achievement of around the same order of magnitude, but a neighbour correlation of less than 15 percent (among children at Key Stages 2 and 4); and in the USA, Solon et al. (2000) find a sibling correlation of around 50 percent in educational attainment, but a neighbour correlation of under 20 percent. In comparison, the degree of clustering of both economic and educational outcomes in East Africa appears large.

5.2 Aggregate variance bounds

Turning to the variance decomposition, Table 5 reports estimates of the three primitives required to calculate the within-community sorting component (also shown). Together these form the basis for calculation of the variance bounds, which are set out in Table 6 and cover the household upper-bound model (HUB), the unrestricted linear model (ULM) and school upper-bound model (SUB). In both tables we present estimates on a pooled (all East Africa) basis from each of the three procedures described in Section 3 – namely, our own indirect fixed-effects approach (FEi), as well as the mixed-effects model (MLM) and the direct

Table 5: Primitive components of variance decomposition, all East Africa

		Absolute (in st. dev. units)			Relative (in %)		
		FEi	MLM	FEd	FEi	MLM	FEd
Household	σ_ω	45.1	46.0	45.1	20.3	21.1	20.3
		(0.2)	(0.3)	(0.2)	(0.3)	(0.3)	(0.3)
School	σ_ν	14.3	27.6	21.2	2.1	7.6	4.5
		(0.2)	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)
Correlation	ρ_{hs}	0.41	0.40	0.30	0.41	0.40	0.30
		(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
⇒ Sorting	$2\Sigma_{hs}$	25.3	34.5	24.9	6.4	11.9	6.2
		(0.3)	(1.0)	(0.3)	(0.2)	(0.6)	(0.2)

Note: columns indicate the outcome transformation and decomposition method (fixed-effects indirect, mixed-effects linear model and fixed-effects direct respectively); 'absolute' refers to standard deviation units ($\times$ 100); 'relative' is as a percent of the total variance; standard errors (in parentheses) calculated via a clustered bootstrap procedure.

fixed-effects (FEd) procedures based on AM18.[12] The tables also report results based on absolute variance shares, given in standard deviation units ($\times 100$), and relative variance shares, given in percentages. The relative contribution of the relevant components to overall inequality of opportunity, which effectively refers to the systematic components above the level of the individual, are further illustrated in Figure 1.

Looking at Table 5, we note that the three decomposition methods yield highly consistent results; hence, the different methods yield similar estimates for the various variance bounds (Table 6). Despite some differences in precise magnitudes, discussed further below, this suggests the present results are robust to the method of estimation (e.g., mixed- versus fixed-effects) and, more specifically, that our simple indirect approach (FEi) performs adequately. Furthermore, we find a positive correlation between the household- and school-effects, ranging between 0.30 and 0.41. Consequently, the derived within-community sorting component, shown in the final row, is always larger than the stand-alone (lower-bound) contribution of schools. As such, within-community sorting makes a non-trivial contribution to the total variance in test-scores. This contribution may not appear so large, particularly when viewed in relative terms (between 12 and 6 percent). But this must be placed in context. According to the Uwezo data, an additional year of schooling is associated with an approximate 0.18 standard deviation increase in a child's test-score *ceteris paribus*. Thus,

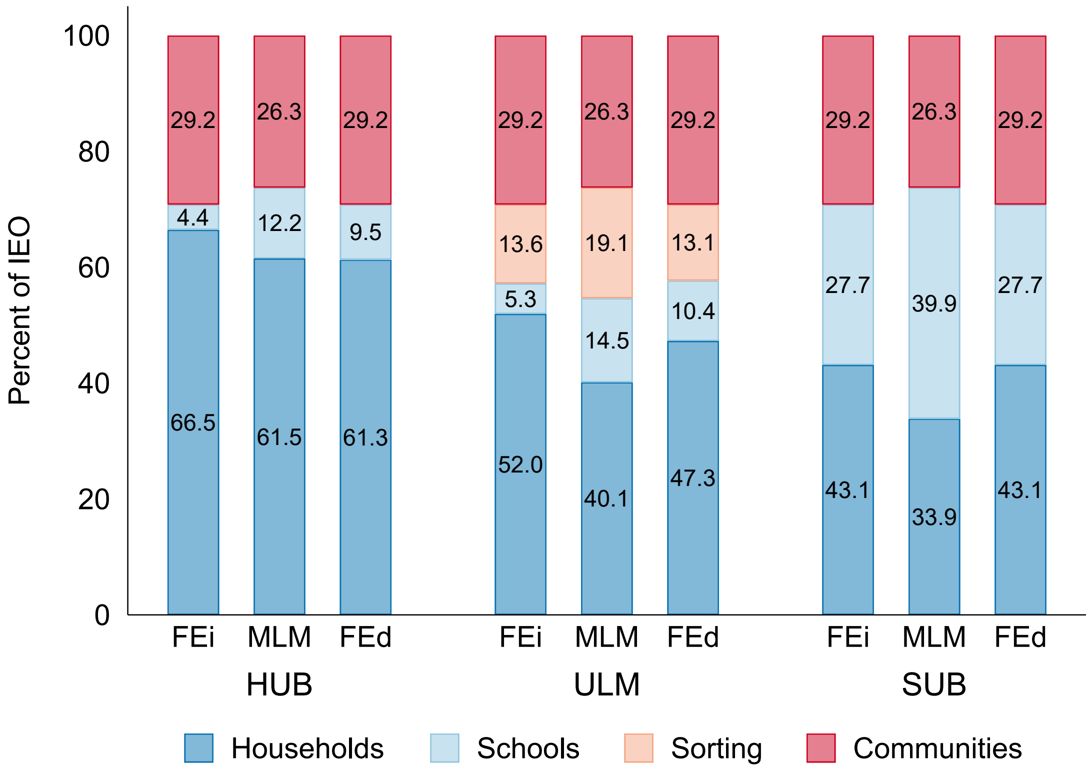
[12] For the MLM results, variance components are calculated from the estimated BLUPs of the random-effects plus their standard error, which match the population second moments. Furthermore, the estimated BLUPs of the random-effects show material cross-correlation, particularly with the community effect. As a result, in order to assure comparability with the fixed-effects procedures in which the school and household primitives are defined as uncorrelated factors, we use only the component of the estimated school and household BLUPs that are orthogonal to the community component. Further details available on request.

Table 6: Decomposition of variation in achievement across East Africa

Method →	FEi			MLM			FEd		
Model →	HUB	ULM	SUB	HUB	ULM	SUB	HUB	ULM	SUB
(a) *Absolute contributions (st. dev. units × 100)*:									
Individual	10.0	10.0	10.0	8.1	8.1	8.1	10.0	10.0	10.0
	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)	(0.1)	(0.2)	(0.2)	(0.2)
Household	55.9	49.5	45.1	62.0	50.1	46.0	53.7	47.2	45.1
	(0.3)	(0.2)	(0.2)	(1.1)	(0.5)	(0.3)	(0.3)	(0.2)	(0.2)
School	14.3	15.7	36.1	27.6	30.1	49.9	21.2	22.2	36.1
	(0.2)	(0.2)	(0.4)	(0.1)	(0.3)	(1.3)	(0.2)	(0.2)	(0.4)
Sorting	-	25.3	-	-	34.5	-	-	24.9	-
		(0.3)			(1.0)			(0.3)	
Community	37.1	37.1	37.1	40.5	40.5	40.5	37.1	37.1	37.1
	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)
Residual	72.0	72.0	72.0	60.7	60.7	60.7	72.0	72.0	72.0
	(1.0)	(1.0)	(1.0)	(0.5)	(0.5)	(0.5)	(1.0)	(1.0)	(1.0)
(b) *Relative contributions (percent)*:									
Individual	1.0	1.0	1.0	0.6	0.6	0.6	1.0	1.0	1.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Household	31.3	24.5	20.3	38.4	25.1	21.1	28.9	22.3	20.3
	(0.6)	(0.4)	(0.3)	(0.9)	(0.3)	(0.3)	(0.5)	(0.4)	(0.3)
School	2.1	2.5	13.0	7.6	9.1	24.9	4.5	4.9	13.0
	(0.1)	(0.1)	(0.4)	(0.1)	(0.1)	(1.0)	(0.1)	(0.1)	(0.4)
Sorting	-	6.4	-	-	11.9	-	-	6.2	-
		(0.2)			(0.6)			(0.2)	
Community	13.7	13.7	13.7	16.4	16.4	16.4	13.7	13.7	13.7
	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)
Residual	51.9	51.9	51.9	36.9	36.9	36.9	51.9	51.9	51.9
	(0.8)	(0.8)	(0.8)	(0.8)	(0.8)	(0.8)	(0.8)	(0.8)	(0.8)

Note: super-columns indicate the decomposition method (fixed-effects indirect, mixed-effects linear model and fixed-effects direct respectively) and sub-columns the assumed variance components model, which are the household upper-bound model (HUB), the unrestricted linear model (ULM) and the school upper-bound model (SUB); standard errors (in parentheses) calculated via a clustered bootstrap procedure.

Figure 1: Variance components (in % of inequality of opportunities), alternative methods and models, East Africa pooled



Note: bars indicate the percentage contribution of each factor to inequality of educational opportunity (IEO), which is the variance accounted for by systematic supra-individual factors (households, schools, communities plus their covariance); abbreviations for methods and models are as per Table 6.

the contribution of sorting to the variance in test-scores is roughly equivalent to around half a year of schooling.

Three further points merit note. First, sources of inequality of educational opportunity outside the household, including sorting, are substantial. Considering the ULM results in Table 6, the combination of schools, communities and sorting is always greater than the household contribution alone; and the same insight derives from the school upper-bound model, where the contributions of schooling and communities exceeds that of households.[13] Second, the stand-alone contribution of schools also is fairly moderate, representing no more than 10 percent of the total variance under the ULM estimates. Necessarily, this limits the overall contribution of sorting; in this sense, were variation between schools to increase, one would expect the contribution of sorting also to be larger (assuming the same

[13] Another way of interpreting the magnitude of the sorting contribution is in terms of the gap between the lower and upper-bound contributions of schools and households. Following Table D1, the smaller the correlation coefficient, the smaller the gap. But for both effects, the gap is relatively wide – e.g., based on the direct method, the school-effect ranges from 4.2 to 11.3 percent of the variation in achievement.

correlation holds). In comparative terms, however, these magnitudes are not especially small. For example, AM18 estimate that school *and* community effects jointly explain around 5 percent of the variation in high school graduation rates (lower-bounds). Here, schools alone account for roughly the same amount of variation in achievement; but, when combined with community effects, the joint contribution is at least 15 percent (Table 6). Similarly, the joint contribution of schools and communities taken from the school upper-bound estimates, which are generally more comparable to estimates found elsewhere (e.g., Freeman and Viarengo, 2014) are hardly trivial. Our estimates for East Africa place this joint effect at around 25 percent (at least), which is similar to the average school-effect variance contribution calculated by Pritchett and Viarengo (2015) for a range of countries. So, while the present results do point to some similarities in the respective variance contributions of different components between East Africa and elsewhere, the material contribution of residential location within the joint school/community effect is noteworthy.

Third, differences in estimates between the three decomposition methods appear to derive from two main sources. On the one hand, some of the differences for the correlation and sorting components simply reflect the sensitivity of calculations based on equation (3b). Recall we estimate ρ_{hs} from the household- and school-effect lower-bounds. Thus, even minor differences here, such as slightly larger values under the MLM approach, are magnified in the estimates for the correlation term (as it is their product which is of interest for sorting). On the other hand, the mixed- and fixed-effects approaches treat the community effects in a slightly different fashion. In the latter case, community effects are removed from the joint fixed effect (households plus schools) in the first of the series of orthogonal projections (see Appendix B). As such, and in similar fashion to the upper-bound models discussed in Section 2, the estimated community variance contribution will capture *both* the direct effect of shared community-level factors on outcomes (e.g., via environmental conditions) plus any covariance with either household or school factors. Thus, any possible sorting of households (or teachers) across different communities will be absorbed by the community effect.[14] In the mixed-effects approach, all terms are estimated simultaneously, meaning that a strict distinction between within- and between-community components is not enforced *ex ante*. Although we implement a correction for this to obtain the uncorrelated school and household components (see above), this methodological difference is likely to explain the (small) disparities in our estimates of the community effect and the uncorrelated variance components.

[14] Since we not observe the same households (or teachers) across different communities, we are not able to distinguish between the direct and sorting effects associated with residential location.

5.3 Subgroup heterogeneity

Averages can often hide substantial heterogeneity. Thus, to look behind the pooled results, we take advantage of the properties of the variance and calculate the decomposition for specific subgroups. These results are presented in Table 7, where we stratify individuals along various dimensions – gender, age group, schooling status, school type and SES quintile. For simplicity, and given the broad consistency across the empirical procedures, hereafter we only present results for the direct fixed-effects method. This is chosen as it gives the most conservative estimates for both the correlation coefficient and sorting component (see Table 5). For reference, Appendix Table D2 presents the identical subgroup decomposition in absolute magnitudes; and Tables D3 and D4 present the same relative and absolute results based on the mixed-effects method, which provides the largest estimates of the sorting component among the three methods.

Focusing on the main insights, there are no systematic differences in the variance components by gender or age. However, there are moderate differences among socio-economic groups. In particular, lower quintiles display comparatively larger contributions due to both households and schools, as well as a higher correlation coefficient, leading to a larger sorting component (5.4 versus 3.7 percent). In other words, IEO is larger among the less advantaged. Even more critical differences emerge between children who are in and out of school. Recall that the school-effect treats all children within each community who are not enrolled as a separate category (effect). Thus, when we undertake the decomposition for *all* children, the school and sorting effects not only refer to specific schools (for those enrolled) but also the effect of the local environment (for children out of school), including the general quality of available schools. In this sense, the overall variance decomposition (e.g., Table 6) refers to both the extensive and intensive margins of engagement in the school system.

Understood in this light, the variance decomposition for the subgroup of enrolled children only captures variation at the intensive margin within the formal school system; and the variance decomposition for the subgroup of non-enrolled children only captures variation at the intensive margin outside of the schooling system. Since we note that the household-school-effect correlation is considerably lower within these two subgroups (<15 percent versus 24 percent overall), mechanically this means that the group-specific averages of both effects must not only diverge considerably from the overall means, but they do so in the same direction as the correlation coefficient. That is, out-of-school children face both lower household *and* lower (general) school-effects. In turn, much of the contribution due to sorting is at the extensive margin, reflecting differences in conditions faced by children attending school and those who do not. Plausibly, some of this may be due to a causal effect of poor-quality local schools faced by less advantaged children on the decision to attend; but we cannot prove this here.

The final subgroup distinction in Table 7 points to differences between school types within the schooling system. In particular, we find a larger household-school-effect correlation and (correspondingly) a larger relative contribution of sorting among children attending private schools compared to those attending public school (16 versus 12 percent). We also note that the contribution of variation between schools is moderately larger in the private sector. This may well reflect differences associated with lower- and higher-cost private schools, where these segments serve children of quite different backgrounds. Also, and consistent with previous literature, this suggests that where there is greater scope for school choice, processes of sorting tend to be larger.

5.4 Spatial heterogeneity

A related form of heterogeneity relates to geographical locations. Arguably, this is particularly relevant from the perspective of policy as it speaks to the possibility of targeted interventions. A spatial perspective is also motivated by educational differentials within each country, as shown in Table 2. These indicate large differences in both the average level and variance of test-scores across regions of each country. Retaining our focus on the unrestricted linear model presentation estimated via the direct (fixed-effects) method, Figure 2 illustrates the relative variance components for each country, and Tables D5–D7 report results at the regional level within each country.

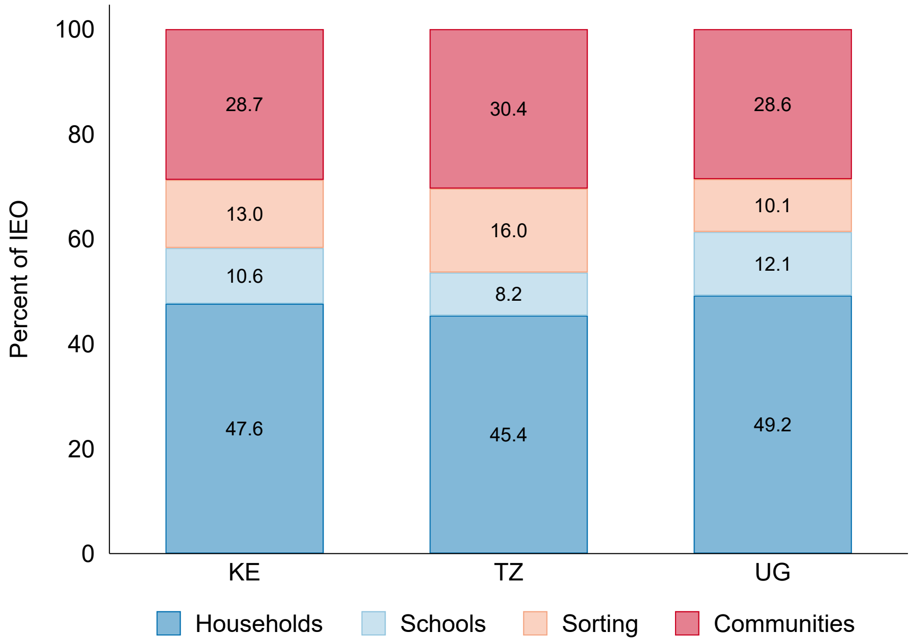
As in the previous subsection, we find substantial heterogeneity in the size and relative importance of the different factors. Between countries, the structure of (relative) IEO appears most distinct in Tanzania where the contribution of communities and sorting is generally larger. Within countries, the absolute variance contribution of different components can vary by a factor of around three. For instance, in Kenya the absolute contribution of sorting in the Central region is 0.11 standard deviation units versus 0.33 in the North Eastern region. Similarly, we see large differences in the variance contribution of schools, ranging from 0.13 (Central) to 0.30 units (North Eastern). In relative terms such differences are somewhat less pronounced; but, even here there remain material differences in the contributions of schools and sorting among regions. We also note that regions containing the capital city in each country (defined here as Central in Kenya and Uganda, Dar Es Salaam in Tanzania) tend to display comparatively lower absolute test-score inequalities, as well as somewhat smaller variation due to schools and sorting. This is consistent with an interpretation that

Table 7: Sub-group variance decomposition (in percent of total variance)

Group	Level	Indiv. σ_x^2	Hhld σ_h^2	Sch. σ_s^2	Sort $2\Sigma_{hs}$	Comm. σ_c^2	Resid. σ_e^2	Total σ_t^2	Correl. ρ_{hs}
Female	No	0.9	22.4	5.0	13.8	6.4	51.6	100.0	0.30
		(0.0)	(0.5)	(0.1)	(0.3)	(0.2)	(1.0)	-	(0.01)
	Yes	0.8	22.2	4.8	13.8	6.0	52.3	100.0	0.29
		(0.0)	(0.4)	(0.1)	(0.3)	(0.2)	(1.0)	-	(0.01)
Age group	6-9	1.4	23.0	5.6	14.3	6.5	49.1	100.0	0.29
		(0.1)	(0.6)	(0.2)	(0.4)	(0.3)	(1.5)	-	(0.01)
	10-13	0.6	21.6	4.4	13.3	5.9	54.3	100.0	0.30
		(0.0)	(0.3)	(0.1)	(0.2)	(0.2)	(0.7)	-	(0.01)
SES quintile	1	1.3	23.5	5.5	14.9	7.5	47.4	100.0	0.33
		(0.1)	(0.3)	(0.2)	(0.3)	(0.3)	(0.6)	-	(0.01)
	2	1.2	22.9	5.5	14.2	7.1	49.1	100.0	0.32
		(0.1)	(0.3)	(0.1)	(0.2)	(0.3)	(0.6)	-	(0.01)
	3	1.0	22.9	5.1	14.1	6.1	50.7	100.0	0.28
		(0.0)	(0.2)	(0.1)	(0.2)	(0.2)	(0.4)	-	(0.01)
	4	1.0	22.3	5.4	14.1	6.2	51.0	100.0	0.28
		(0.0)	(0.2)	(0.1)	(0.2)	(0.2)	(0.5)	-	(0.01)
	5	0.9	20.8	5.1	15.3	6.1	51.7	100.0	0.30
		(0.0)	(0.3)	(0.1)	(0.3)	(0.2)	(0.5)	-	(0.01)
Enrolled	No	1.6	18.0	10.4	14.1	4.4	51.4	100.0	0.16
		(0.1)	(0.6)	(0.3)	(0.5)	(0.4)	(1.0)	-	(0.02)
	Yes	0.3	22.8	4.2	14.6	3.8	54.2	100.0	0.20
		(0.0)	(0.4)	(0.1)	(0.3)	(0.1)	(0.8)	-	(0.00)
Private school	No	0.4	23.7	3.8	15.2	3.5	53.4	100.0	0.18
		(0.0)	(0.2)	(0.1)	(0.3)	(0.1)	(0.4)	-	(0.00)
	Yes	0.3	21.3	6.0	14.1	4.8	53.6	100.0	0.21
		(0.0)	(0.4)	(0.3)	(0.5)	(0.3)	(1.1)	-	(0.01)
All		1.0	22.3	4.9	13.7	6.2	51.9	100.0	0.30
		(0.0)	(0.4)	(0.1)	(0.3)	(0.2)	(0.8)	-	(0.01)

Note: the table reports direct fixed-effects variance decomposition estimates, based on the unrestricted linear model representation; standard errors (in parentheses) are calculated via a clustered bootstrap procedure.

Figure 2: Variance components (in % of IEO), unrestricted linear model, by country

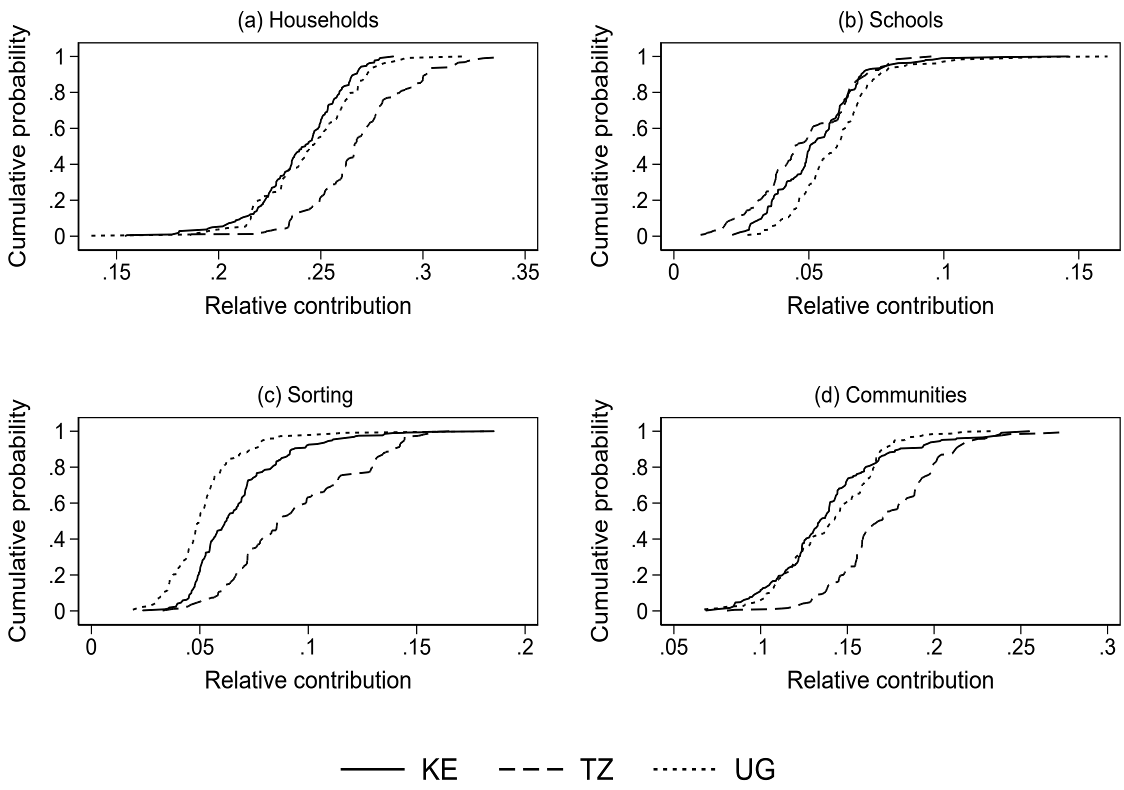


Note: bars indicate the percentage contribution of each factor to inequality of educational opportunity (IEO), which is the variance accounted for by systematic supra-individual factors (households, schools, communities and their covariance); estimates taken from the direct fixed-effects variance decomposition procedure, based on the unrestricted linear model representation; KE is Kenya; TZ is Tanzania (mainland); and UG is Uganda.

capital cities tend to provide more equal access to schools of a similar quality.

An advantage of the Uwezo data is that we can run the same analysis at the district level. Figure 3 plots the cumulative empirical district-level distributions of the relative variance shares for the four main IEO components taken from the preferred estimates, by country. These confirm substantial variations across all components within each country, but country-specific differences continue to be evident – i.e., the distribution functions display (approximate) first-order dominance in all cases. Variation in the magnitude of sorting across countries is confirmed in Figure 4, which shows the cumulative distribution of the estimates for ρ_{hs} across districts, based on both the direct and MLM procedures. This underlines that the correlation between household- and school-effects is substantially larger in Tanzania – e.g., for the median district, the correlation in Tanzania is almost twice that in Uganda. But within each country, movement from the 10th to the 90th percentile of the district-level distribution is associated with at least a 15 percentage point increase in the correlation coefficient. In other words, spatial heterogeneity in sorting is material.

Figure 3: Cumulative distribution of variance components (in %) across districts, by country



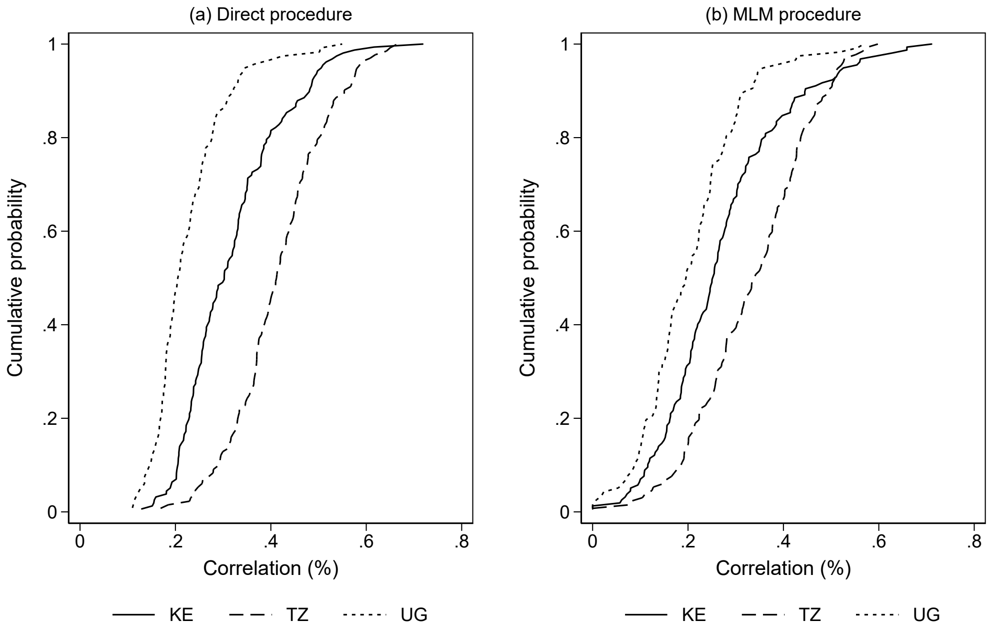
Note: lines indicate the cumulative distribution of of relative contributions of each factor to total test-score variance; estimates taken from the direct fixed-effects variance decomposition procedure, based on the unrestricted linear model representation; KE is Kenya; TZ is Tanzania (mainland); and UG is Uganda.

5.5 Long-run inequalities

Finally, following the spirit of Kremer (1997), we consider the implications of sorting for long-run (steady-state) educational inequalities. The key idea here is that in the presence of positive educational sorting, children from more advantaged families receive a double benefit that can be expected to exacerbate inequalities over time. Such children benefit both from family circumstances and from higher-quality schooling, which has a multiplicative effect across generations.

In order to quantify the potential magnitude of these effects, we move to a dynamic or intergenerational setting. Furthermore, we adopt the simple formulation that household effects (h) are directly proportional to parental achievement, reflecting both parental capacity to support learning and other characteristics that flow from their educational level. Indexing generations by g, and using previous notation, these assumptions imply $h_{jg} \approx h_{j0} + \delta t_{j,g-1}$,

Figure 4: Household-school factor correlation across districts, by country



Note: lines indicate the cumulative distribution of estimates for the household-school pairwise correlation coefficient under different decomposition procedures (MLM and fixed effects direct), based on the unrestricted linear model representation; KE is Kenya; TZ is Tanzania (mainland); and UG is Uganda.

which in turn means:

$$\delta^2 = \frac{\sigma^2_{h_g}}{\sigma^2_{t_{g-1}}} = \frac{\sigma^2_{\omega_g}}{(1-\rho^2_{hs,g})\sigma^2_{t_{g-1}}} = \frac{\bar{\delta}^2}{1-\rho^2_{hs,g}} \tag{7}$$

where $\bar{\delta} = \sigma^2_{\omega_g}/\sigma^2_{t_{g-1}}$ is the sorting-invariant or raw inter-generational persistence parameter. From this, we can then solve for the long-run steady-state level of educational inequality:

$$\sigma^2_{\tilde{t}} = \frac{\sigma^2_\nu + 2\bar{\delta}\rho\sigma_\nu\sigma_{\tilde{t}} + (\sigma^2_c + \sigma^2_e)(1-\rho^2)}{1-\rho^2-\bar{\delta}^2} \tag{8}$$

where σ^2_c is the upper-bound variance contribution associated with residential location (which we treat as exogenous, for simplicity). (See Appendix A for derivation).

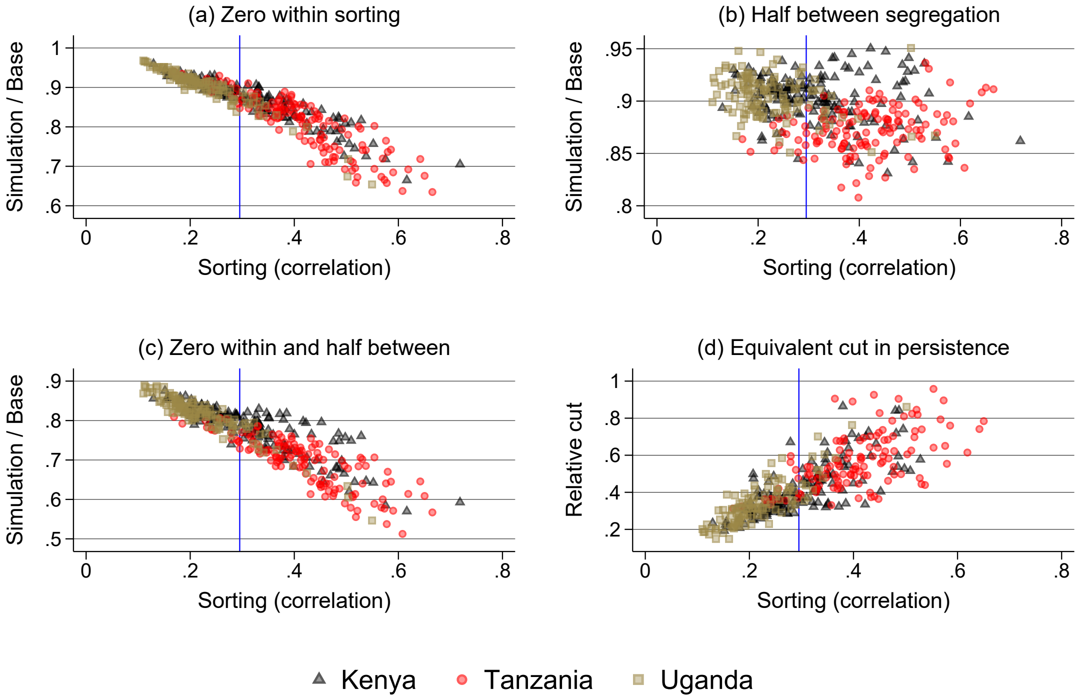
To simulate this model, a value for the raw persistence parameter needs to be chosen. To do so, we opt for a data-driven approach and make the assumption that currently-observed inequality is approximately at the steady-state (i.e., $\sigma^2_t = \sigma^2_{\tilde{t}}$), allowing us to solve for $\bar{\delta}$ using equation (A3c) based on previously-estimated values of the primitives. This gives us the magnitude of the persistence parameter that would maintain achievement inequality at current levels. Fixing $\bar{\delta}$ at this also means that simulations based on changes to (other) primitives in the model can be interpreted directly with reference to the current magnitude of achievement inequality. Appendix Figure E1 plots the resulting district-level cumulative distributions of the inequality-constant persistence parameter, which range from 0.30 to 0.56; and the corresponding pooled overall coefficient is 0.46. Notably, these magnitudes are broadly consistent with estimates of intergenerational regression coefficients in educational attainment found elsewhere.[15] In other words, our baseline assumption is likely to be conservative.

With the constant-inequality persistence parameter in hand, we now use equation (A3c) to simulate the steady-state level of inequality under alternative assumptions for other inputs. First, we set the sorting correlation coefficient ρ to zero, under which the same equation simplifies to $\sigma^2_{\bar{t}} = (\sigma^2_\nu + \sigma^2_c + \sigma^2_e)/(1 - \bar{\delta}^2)$. Figure 5(a) plots the results of this exercise, taking districts as the unit of observation, and where the magnitudes of the primitives entering each unit-level calculation are taken from the earlier fixed-effects direct (FEd) estimates. At the average magnitude of the sorting coefficient, indicated by the blue vertical lines, steady-state inequality falls by about 12 percent relative to the baseline; but in a non-trivial number of districts, the reduction is more than 20 percent. In light of earlier results, this reveals that eliminating educational sorting would lead to a substantially larger reduction in long-run inequality than might be suggested by the relative variance contribution of sorting itself (Figure 3). Indeed, the relative variance contribution due to sorting is roughly only half the expected proportional decline in inequality if sorting were to be reduced.

Eliminating within-community sorting does not address segregation of either households or teachers (schools) by residential location, as captured by the community effect. Our earlier results indicated this term was material, potentially reflecting significant barriers to internal mobility in the East African context. To look at the implications of this more general form of sorting for long-run educational inequality, Figure 5(b) estimates the unique impact of halving the variance contribution due to community effects. For the average district, the reduction in inequality compared to the baseline is similar to the first simulation, falling by 10 percent. And combining these two scenarios equates to more than a 20 percent fall in long-run inequality, on average – see Figure 5(c). To put this magnitude of reduction

[15] For instance, Hertz et al. (2007) estimate an average persistence parameter of 0.42 for a range of mostly developed countries. However, existing estimates for (low income) developing countries typically point to somewhat larger values. For instance, Emran and Shilpi (2015) find a persistence parameter of around 0.50 in India; and in a sample of African countries, Azomahou and Yitbarek (2016) find an average persistence of 0.66. These values are generally larger than our own estimates, implying a higher steady-state level of educational inequality than observed today. Employing the alternative assumption: $\delta = \bar{\delta}\sqrt{1-\rho^2_{hs}} = 0.66\sqrt{1-\rho^2_{hs}}$, we find steady-state inequality is expected to be 40 percent larger than at present (for the region as a whole).

Figure 5: Simulations of steady-state inequality



Note: panels (a)-(c) plot the ratio of long-run inequality to present inequality under different assumptions; in (a) we assume sorting is eliminated; in (b) we halve the variance contribution due to community effects; and in (c) we combine the latter two; panel (d) indicates the relative reduction in the intergenerational persistence parameter required to achieve the fall in long-run inequality indicated in panel (c) (e.g., 0.1 implies a 10 percent reduction); vertical blue line is the pooled correlation coefficient, given in the x-axis.

in inequality in context, it is equivalent to a very substantial fall in the intergenerational persistence of educational achievement. Figure 5(d) reports the percentage reduction in the persistence parameter required to match the effect of both eliminating within-community sorting and halving between-community segregation. On average, the equivalent reduction is over 40 percent, implying a required persistence parameter of 0.27 (versus 0.42 in the baseline).

6 Conclusions

There is continued debate regarding the extent to which sorting exacerbates inequalities, including in educational attainment and achievement. We contribute to this debate by undertaking a variance decomposition, treating sorting as the contribution of the covariance between household- and school-effects to variation in educational outcomes. We quantify the contribution of sorting to learning inequalities among over 1,000,000 children in three

East African countries: Kenya, mainland Tanzania and Uganda.

To estimate the contribution of sorting, we propose three complementary approaches. First, we show how an indirect procedure, based on separate one-way fixed-effects estimators, can identify the lower-bound or uncorrelated variance contributions of both schools and households, which are then used as necessary primitives to derive the sorting component. Second, we adopt the approach of AM18 (Altonji and Mansfield, 2018), who show how group means of observed factors can absorb confounding effects of sorting on unobserved characteristics. This helps isolate the independent contributions of the factors of interest, from which lower-bound variance contributions can be estimated using mixed-effects estimators. Third, we extend the AM18 approach to allow for fixed- as opposed to random-effects, which is both computationally more tractable and better suited to explore sub-group heterogeneity.

Empirically, all three estimation procedures indicate positive sorting of pupils across schools that accounts for up to 8 percent of the total test-score variance and almost a fifth of the joint variation in test-scores due to schools, communities and households. The sorting contribution tends to be larger among families at the top and bottom ends of the socio-economic distribution, among those sending their children to private schools, as well as in specific locations.

To explore the implications of these results, we conduct simulations of how learning inequality evolves over time. These show that for the average district in the region, the steady-state level of educational inequality would fall by around 10 percent if sorting were to be fully eliminated. Moreover, in a number of districts, the reduction in inequality from eliminating sorting would be over 20 percent of the total variance and over 40 percent of IEO. As such, this suggests that policies that take sorting into account, and even actively counteract it, merit consideration.

References

Abowd, J.M., Creecy, R.H. and Kramarz, F. (2002). Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.

Altonji, J.G. and Mansfield, R.K. (2018). Estimating group effects using averages of observables to control for sorting on unobservables: School and neighborhood effects. *American Economic Review*, 108(10):2902–46.

Andrews, M.J., Gill, L., Schank, T. and Upward, R. (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3):673–697.

—— (2012). High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters*, 117(3):824–827.

Azomahou, T.T. and Yitbarek, E.A. (2016). Intergenerational education mobility in Africa: has progress been inclusive? Policy Research Working Paper 7843, The World Bank.

Bates, D.M. (2010). *`lme4`: Mixed-Effects Modeling with R*. New York: Springer.

Behrman, J.R., Gaviria, A. and Székely, M. (2001). Intergenerational mobility in Latin America. *Economía*, 2(1):1–31.

Björklund, A. and Salvanes, K.G. (2011). Education and family background: Mechanisms and policies. In E. Hanushek, S. Machin and L. Wößmann (Eds.), *Handbook of the Economics of Education*, volume 3, chapter 3, pp. 201–247. Elsevier.

Bowles, S. (1970). Towards an educational production function. In W.L. Hansen (Ed.), *Education, Income, and Human Capital*, pp. 11–70. National Bureau of Economic Research, Inc.

Carneiro, P. (2008). Equality of opportunity and educational achievement in Portugal. *Portuguese Economic Journal*, 7(1):17–41.

Chakravarty, S.R. (2001). The variance as a subgroup decomposable measure of inequality. *Social Indicators Research*, 53(1):79–95.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R., Hendren, N. and Katz, L.F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *American Economic Review*, 106(4):855–902.

Clotfelter, C.T., Ladd, H.F. and Vigdor, J.L. (2011). Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education Finance and Policy*, 6(3):399–438.

Emran, M.S. and Shilpi, F. (2015). Gender, geography, and generations: Intergenerational educational mobility in post-reform India. *World Development*, 72:362–380.

Fernandez, R. (2003). Sorting, education, and inequality. In M. Dewatripont, L.P. Hansen and S.J. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume II, chapter 1, pp. 1–40. Cambridge University Press.

Fernández, R. and Rogerson, R. (2001). Sorting and long-run inequality. *The Quarterly Journal of Economics*, 116(4):1305–1341.

Ferreira, F.H. and Gignoux, J. (2014). The measurement of educational inequality: Achievement and opportunity. *The World Bank Economic Review*, 28(2):210–246.

Freeman, R.B. and Viarengo, M. (2014). School and family effects on educational outcomes across countries. *Economic Policy*, 29(79):395–446.

Friesen, J. and Krauth, B. (2007). Sorting and inequality in Canadian schools. *Journal of public Economics*, 91(11-12):2185–2212.

—— (2010). Sorting, peers, and achievement of aboriginal students in British Columbia. *Canadian Journal of Economics/Revue canadienne d'économique*, 43(4):1273–1301.

Fryer, R.G. and Levitt, S.D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, 86(2):447–464.

Gaure, S. (2014). Correlation bias correction in two-way fixed-effects linear regression. *Stat*, 3(1):379–390.

Gibbons, S., Overman, H.G. and Pelkonen, P. (2014). Area disparities in Britain: Understanding the contribution of people vs. place through variance decompositions. *Oxford Bulletin of Economics And Statistics*, 76(5):745–763.

Hanushek, E. and Rivkin, S. (2006). Teacher quality. *Handbook of the Economics of Education*, 2:1051–1078.

Hanushek, E.A. and Rivkin, S.G. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics*, 4(1):131–157.

Hanushek, E. and Wößmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, 116(510):C63–C76.

Hanushek, E. and Yilmaz, K. (2007). The complementarity of Tiebout and Alonso. *Journal of Housing Economics*, 16(2):243–261.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N. and Verashchagina, A. (2007). The inheritance of educational inequality: International comparisons and fifty-year trends. *The BE Journal of Economic Analysis & Policy*, 7(2).

Heyneman, S.P. and Stern, J.M. (2014). Low cost private schools for the poor: What public policy is appropriate? *International Journal of Educational Development*, 35:3–15.

Hsieh, C.T. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, 90(8):1477–1503.

Jackson, C.K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labor Economics*, 27(2):213–256.

Jones, S., Schipper, Y., Ruto, S. and Rajani, R. (2014). Can your child read and count? Measuring learning outcomes in East Africa. *Journal of African Economies*, 23(5):643–672.

Koedel, C., Mihaly, K. and Rockoff, J.E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47:180–195.

Kremer, M. (1997). How much does sorting increase inequality? *The Quarterly Journal of Economics*, 112(1):115–139.

Lindahl, L. (2011). A comparison of family and neighborhood effects on grades, test scores, educational attainment and income – evidence from Sweden. *The Journal of Economic Inequality*, 9(2):207–226.

Mazumder, B. (2008). Sibling similarities and economic inequality in the US. *Journal of Population Economics*, 21(3):685–701.

—— (2011). Family and community influences on health and socioeconomic status: sibling correlations over the life course. *The B.E. Journal of Economic Analysis & Policy*, 11(3):Article 1.

Morris, J.S. (2002). The BLUPs are not "best" when it comes to bootstrapping. *Statistics & Probability Letters*, 56(4):425–430.

Nechyba, T.J. (2006). Income and peer quality sorting in public and private schools. In E.A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, volume 2, pp. 1327–1368. Elsevier.

Nicoletti, C. and Rabe, B. (2013). Inequality in pupils' test scores: How much do family, sibling type and neighbourhood matter? *Economica*, 80(318):197–218.

OECD (2018). *Effective Teacher Policies*. doi:10.1787/9789264301603-en. URL https://www.oecd-ilibrary.org/content/publication/9789264301603-en.

Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. *The Quarterly Journal of Economics*, 118(4):1533–1575.

Pritchett, L. and Viarengo, M. (2015). Does public sector control reduce variance in school quality? *Education Economics*, 23(5):557–576.

Raaum, O., Salvanes, K.G. and Sørensen, E.Ø. (2006). The neighbourhood is not what it used to be. *The Economic Journal*, 116(508):200–222.

Roemer, J.E. (2002). Equality of opportunity: A progress report. *Social Choice and Welfare*, 19(2):455–471.

Sass, T.R., Hannaway, J., Xu, Z., Figlio, D.N. and Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, 72(2):104–122.

Shorrocks, A.F. (1982). Inequality decomposition by factor components. *Econometrica:*, pp. 193–211.

Solon, G., Page, M.E. and Duncan, G.J. (2000). Correlations between neighboring children in their subsequent educational attainment. *Review of Economics and Statistics*, 82(3):383–392.

Stanek, E.J., Well, A. and Ockene, I. (1999). Why not routinely use best linear unbiased predictors (BLUPs) as estimates of cholesterol, per cent fat from Kcal and physical activity? *Statistics in Medicine*, 18(21):2943–2959.

Uwezo (2012). Are our children learning? literacy and numeracy across East Africa. Technical report, Uwezo. URL www.uwezo.net/wp-content/uploads/2012/08/RO_2012_UwezoEastAfricaREport.pdf.

Walsh, P. (2009). Effects of school choice on the margin: The cream is already skimmed. *Economics of Education Review*, 28(2):227–236.

Watkins, K. (2012). The power of circumstance: A new approach to measuring education inequality. Center for Universal Education Working Paper 5, Brookings Institute.

SUPPLEMENTARY MATERIAL

A Derivations

A.1 Correlated and uncorrelated components

To show the relationship between the uncorrelated and correlated school component, we start with the household upper-bound model from Table 1:

$$\sigma_\nu^2 = \sigma_t^2 - (1+\gamma)^2\sigma_h^2 - \sigma_e^2 \tag{A1a}$$
$$= \sigma_h^2 + \sigma_s^2 + 2\Sigma_{hs} - (1+\gamma)^2\sigma_h^2 \tag{A1b}$$
$$= \sigma_h^2 + \sigma_s^2 + 2\rho_{sh}\sigma_h\sigma_s - (1+\gamma)^2\sigma_h^2 \tag{A1c}$$
$$= \sigma_s^2 - \gamma^2\sigma_h^2 \tag{A1d}$$
$$= (1-\rho_{hs}^2)\sigma_s^2 \tag{A1e}$$

where σ_t^2 is the total variance to be decomposed. Movement from the third to the fourth line derives from the definition of the covariance term, using equation (2), as: $\text{Cov}(h_j, s_k) \equiv \Sigma_{sh} = \gamma\text{Var}(h_j)$; and movement from the fourth to the final line uses the fact that γ defines the slope of the (approximate) linear relation between $\bar{h}$ and s: $\gamma = \rho_{hs}(\sigma_s/\sigma_h)$.

A.2 Steady-state inequality

Based on the unrestricted linear model of Table 1, combined with equations (4) and (7), education inequality evolves as follows:

$$\sigma_t^2 = \delta^2\sigma_{t_{g-1}}^2 + \sigma_s^2 + 2\delta\rho\sigma_s\sigma_{t_{g-1}} + \sigma_c^2 + \sigma_e^2 \tag{A2}$$

In long-run equilibrium it must hold that $\forall n > 0:\ \sigma_{t_{g+n}}^2 = \sigma_{t_g}^2 = \sigma_{\tilde{t}}^2$, thus:

$$\sigma_{\tilde{t}}^2(1-\delta^2) = \sigma_s^2 + 2\delta\rho_{hs}\sigma_s\sigma_{\tilde{t}} + \sigma_c^2 + \sigma_e^2 \tag{A3a}$$
$$\sigma_{\tilde{t}}^2(1-\delta^2) = \frac{\sigma_\nu^2}{1-\rho^2} + \frac{2\delta\rho\sigma_\nu\sigma_{\tilde{t}}}{\sqrt{1-\rho^2}} + \sigma_c^2 + \sigma_e^2 \tag{A3b}$$
$$\sigma_{\tilde{t}}^2 = \frac{\sigma_\nu^2 + 2\bar{\delta}\rho\sigma_\nu\sigma_{\tilde{t}} + (\sigma_c^2+\sigma_e^2)(1-\rho^2)}{1-\rho^2-\bar{\delta}^2} \tag{A3c}$$

B Empirical steps

Following Section 3, we recognise that estimates of separate fixed-effects (e.g., household and schools) in a two-way setting are expected to be biased from a mechanical negative pairwise covariance. As such and excluding the distinct issue of measurement error, there is no reason to suspect this concern affects estimates of their joint contribution. In light of this, to undertake the variance decompositions we begin by running a full model with multiple fixed effects. Then we apply a sequence of orthogonal projections to identify the uncorrelated components of the estimated joint effect.

Concretely, for the indirect fixed-effects procedure (FEi) we proceed as follows:

1. use the fully-specified model to obtain the joint contribution of the latent factors (z_{jkl});
2. project $\hat{z}_{jkl}$ on the location indexes to obtain the (upper-bound) location-specific effect;
3. separately project the household- and school-effect indexes on the fitted residual joint factor from Step 2, ($\hat{z}_{jkl} - \hat{c}_l$), in each case obtaining their corresponding upper-bound contributions; and
4. separately project the household- and school-effect indexes on the relevant residuals from Step 3 to obtain their lower-bound contributions (e.g., the school lower-bound is estimated from the residual from the household upper-bound: $\hat{z}_{jkl} - \hat{c}_l - \hat{h}^*_k,\ h^*_k = h_j + \gamma \bar{h}_{jk}$).

For the extension of the AM18 approach to allow for fixed-effects (FEd), we proceed as follows:

1. estimate a fully-specified model, including two-way fixed-effects, to obtain the joint contribution of the latent factors (z_{jkl}, see equation 4);
2. project $\hat{z}_{jkl}$ on the location indexes to obtain the (upper-bound) location-specific effect ($\hat{c}_l$);
3. project the residual from Step 2 ($\hat{z}_{jkl} - \hat{c}_l$) on the set of sorting proxies ($\bar{h}'_{O,k}$);[16]
4. project the school indexes (only) on the residual component of the joint factor obtained from Step 3 ($\hat{z}_{jkl} - \hat{c}_l - \bar{h}'_{O,k}\hat{\beta}$) to obtain the uncorrelated school contribution;

[16] Where, $\bar{h}'_{O,k}$ refer to school-level means derived from observed household- and pupil-level variables.

5. project the household indexes (only) on the final residual taken from Step 4 to obtain the uncorrelated household lower-bound contribution ($\hat{\omega}_k$).

In both the above approaches, the sequential orthogonal projections ensure that the primitives of interest are not only mutually uncorrelated but also uncorrelated with any observed variables that enter the model, including the sorting proxies. Furthermore, any such additional control variables can be entered in a straightforward manner. For instance, individual-specific observed variables would enter the fully-specified model in Step 1 of each procedure. Also, observed household (school) characteristics are added in Step 3 of each procedure, thereby tightening the one-way estimates of the school (household) effects (see Raaum et al., 2006).

C Regional aggregates

Table C1: Definition of aggregated regions

	Region names		
Country	Aggregated	Original	Obs.
KE	Central	Central	22117
KE	Central	Nairobi	3220
KE	Coast	Coast	34667
KE	Eastern	Eastern	56916
KE	North Eastern	North Eastern	40239
KE	Nyanza	Nyanza	52681
KE	Rift Valley	Rift Valley	119236
KE	Western	Western	57200
TZ	Arusha	Arusha	11006
TZ	Arusha	Kilimanjaro	9410
TZ	Arusha	Mara	12119
TZ	Dar Es Salaam	Dar Es Salaam	3236
TZ	Dar Es Salaam	Pwani	10559
TZ	Iringa	Dodoma	10405
TZ	Iringa	Iringa	9312
TZ	Iringa	Morogoro	8060
TZ	Iringa	Njombe	1574
TZ	Kagera	Geita	3276
TZ	Kagera	Kagera	13242
TZ	Kagera	Mwanza	16146
TZ	Kigoma	Katavi	1412
TZ	Kigoma	Kigoma	9261
TZ	Kigoma	Rukwa	10321
TZ	Ruvuma	Lindi	6166
TZ	Ruvuma	Mtwara	3781
TZ	Ruvuma	Ruvuma	7567
TZ	Singida	Mbeya	10862
TZ	Singida	Singida	9064
TZ	Tabora	Shinyanga	16656
TZ	Tabora	Simiyu	3245
TZ	Tabora	Tabora	13226
TZ	Tanga	Manyara	10846
TZ	Tanga	Tanga	13715
UG	Central	Central	46474
UG	Eastern	Eastern	86761
UG	Northern	Northern	72306
UG	Western	Western	52746

D Additional tables

Table D1: Bounds on variance components

Component	Lower-bound	Upper-bound
Household	$\sigma_\omega^2 = (1-\rho_{hs}^2)\sigma_h^2$	$(1+\gamma)^2\sigma_h^2 = \sigma_h^2 + \rho_{hs}^2\sigma_s^2 + 2\rho_{hs}\sigma_h\sigma_s = \dfrac{\sigma_\omega^2 + \rho_{hs}^2\sigma_\nu^2 + 2\rho_{hs}\sigma_\nu\sigma_\omega}{1-\rho_{hs}^2}$
School	$\sigma_\nu^2 = (1-\rho_{hs}^2)\sigma_s^2$	$(1+\theta)^2\sigma_s^2 = \sigma_s^2 + \rho_{hs}^2\sigma_h^2 + 2\rho_{hs}\sigma_h\sigma_s = \dfrac{\sigma_\nu^2 + \rho_{hs}^2\sigma_\omega^2 + 2\rho_{hs}\sigma_\nu\sigma_\omega}{1-\rho_{hs}^2}$
Sorting	0	$2\rho_{hs}\sigma_h\sigma_s = \dfrac{2\rho_{hs}\sigma_\nu\sigma_\omega}{1-\rho_{hs}^2}$

Note: the table shows how the various variance bounds (c.f., Table 6) can be calculated directly from the underling primitives (c.f., Table 5).

Table D2: Sub-group variance decomposition (in std. dev. units $\times 100$)

Group	Level	Indiv. σ_x^2	Hhld σ_h^2	Sch. σ_s^2	Sort σ_c^2	Comm. $2\Sigma_{hs}$	Resid. σ_e^2	Total σ_t^2	Correl. ρ_{hs}
Female	No	0.09	0.47	0.22	0.25	0.37	0.72	1.00	0.30
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
	Yes	0.09	0.47	0.22	0.24	0.37	0.72	1.00	0.29
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Age group	6-9	0.12	0.47	0.23	0.25	0.37	0.68	0.97	0.29
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)
	10-13	0.08	0.48	0.21	0.25	0.37	0.75	1.02	0.30
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
SES quintile	1	0.11	0.49	0.24	0.27	0.39	0.69	1.00	0.33
		(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	2	0.11	0.47	0.23	0.26	0.37	0.69	0.99	0.32
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
	3	0.10	0.46	0.22	0.24	0.36	0.68	0.96	0.28
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
	4	0.10	0.45	0.22	0.24	0.36	0.68	0.95	0.28
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
	5	0.09	0.42	0.21	0.23	0.36	0.66	0.92	0.30
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Enrolled	No	0.13	0.43	0.32	0.21	0.38	0.72	1.01	0.16
		(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
	Yes	0.06	0.46	0.20	0.19	0.37	0.71	0.97	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)
Private school	No	0.06	0.46	0.19	0.18	0.37	0.70	0.95	0.18
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
	Yes	0.05	0.44	0.23	0.21	0.35	0.69	0.94	0.21
		(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
All		0.10	0.47	0.22	0.25	0.37	0.72	1.00	0.30
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)

Note: the table reports direct fixed-effects variance decomposition estimates, based on the unrestricted linear model representation; standard errors (in parentheses) are calculated via a clustered bootstrap procedure.

Table D3: Sub-group variance decomposition (in percent), MLM procedure

Group	Level	Indiv. σ_x^2	Hhld σ_h^2	Sch. σ_s^2	Sort σ_c^2	Comm. $2\Sigma_{hs}$	Resid. σ_e^2	Total σ_t^2	Correl. ρ_{hs}
Female	No	0.7	24.9	9.0	16.4	11.8	37.2	100.0	0.39
		(0.0)	(0.3)	(0.1)	(0.3)	(0.6)	(0.9)	-	(0.02)
	Yes	0.6	25.3	9.1	16.5	12.1	36.4	100.0	0.40
		(0.0)	(0.4)	(0.1)	(0.3)	(0.7)	(1.0)	-	(0.02)
Age group	6-9	0.9	27.1	10.1	17.2	14.0	30.6	100.0	0.42
		(0.0)	(0.3)	(0.1)	(0.3)	(0.6)	(0.8)	-	(0.02)
	10-13	0.3	23.4	8.2	15.7	10.2	42.2	100.0	0.37
		(0.0)	(0.4)	(0.1)	(0.3)	(0.5)	(0.9)	-	(0.01)
SES quintile	1	0.8	22.0	8.7	18.0	9.2	41.3	100.0	0.34
		(0.0)	(0.3)	(0.1)	(0.4)	(0.4)	(0.7)	-	(0.02)
	2	0.8	22.9	9.0	17.6	9.7	40.0	100.0	0.34
		(0.0)	(0.3)	(0.1)	(0.3)	(0.5)	(0.7)	-	(0.02)
	3	0.7	22.7	8.7	17.5	7.6	42.8	100.0	0.27
		(0.0)	(0.3)	(0.1)	(0.3)	(0.3)	(0.6)	-	(0.01)
	4	0.7	23.7	9.3	17.3	8.6	40.4	100.0	0.29
		(0.0)	(0.3)	(0.1)	(0.3)	(0.4)	(0.7)	-	(0.01)
	5	0.6	25.6	10.0	18.0	10.2	35.5	100.0	0.32
		(0.0)	(0.5)	(0.1)	(0.4)	(0.5)	(1.0)	-	(0.01)
Enrolled	No	1.1	21.1	11.8	18.8	7.2	40.0	100.0	0.23
		(0.1)	(0.5)	(0.3)	(0.6)	(0.7)	(1.0)	-	(0.02)
	Yes	0.2	24.4	8.3	17.2	7.6	42.2	100.0	0.27
		(0.0)	(0.3)	(0.1)	(0.3)	(0.7)	(0.9)	-	(0.02)
Private school	No	0.3	23.7	8.2	18.2	5.8	43.8	100.0	0.21
		(0.0)	(0.2)	(0.1)	(0.3)	(0.4)	(0.6)	-	(0.01)
	Yes	0.2	27.0	9.4	15.6	8.9	39.0	100.0	0.28
		(0.0)	(0.5)	(0.2)	(0.4)	(0.8)	(1.3)	-	(0.02)
All		0.6	25.1	9.1	16.4	11.9	36.9	100.0	0.40
		(0.0)	(0.3)	(0.1)	(0.3)	(0.6)	(0.8)	-	(0.02)

Note: the table reports mixed-effects variance decomposition estimates, based on the unrestricted linear model representation; standard errors (in parentheses) are calculated via a clustered bootstrap procedure.

Table D4: Sub-group variance decomposition (in std. dev. units $\times 100$), MLM procedure

Group	Level	Indiv. σ_x^2	Hhld σ_h^2	Sch. σ_s^2	Sort σ_c^2	Comm. $2\Sigma_{hs}$	Resid. σ_e^2	Total σ_t^2	Correl. ρ_{hs}
Female	No	0.08	0.50	0.30	0.34	0.41	0.61	1.00	0.39
		(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)
	Yes	0.08	0.50	0.30	0.35	0.40	0.60	1.00	0.40
		(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)
Age group	6-9	0.09	0.51	0.31	0.36	0.40	0.54	0.97	0.42
		(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)
	10-13	0.05	0.49	0.29	0.33	0.41	0.66	1.02	0.37
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
SES quintile	1	0.09	0.47	0.29	0.30	0.43	0.64	1.00	0.34
		(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
	2	0.09	0.47	0.29	0.31	0.41	0.62	0.99	0.34
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)
	3	0.08	0.46	0.28	0.27	0.40	0.63	0.96	0.27
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
	4	0.08	0.46	0.29	0.28	0.40	0.61	0.95	0.29
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
	5	0.07	0.47	0.29	0.29	0.39	0.55	0.92	0.32
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Enrolled	No	0.10	0.46	0.35	0.27	0.44	0.64	1.01	0.23
		(0.00)	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
	Yes	0.05	0.48	0.28	0.27	0.40	0.63	0.97	0.27
		(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)
Private school	No	0.05	0.46	0.27	0.23	0.41	0.63	0.95	0.21
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
	Yes	0.04	0.49	0.29	0.28	0.37	0.59	0.94	0.28
		(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
All		0.08	0.50	0.30	0.35	0.40	0.61	1.00	0.40
		(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)

Note: the table reports mixed-effects variance decomposition estimates, based on the unrestricted linear model representation; standard errors (in parentheses) are calculated via a clustered bootstrap procedure.

Table D5: Regional variance decomposition for Kenya

Region	Indiv. σ_x^2	Hhld σ_h^2	Sch. σ_s^2	Sort σ_v^2	Comm. $2\Sigma_{hs}$	Resid. σ_e^2	Total σ_t^2	Correl. ρ_{hs}
(a) *Absolute contributions (st. dev. units × 100)*:								
Central	0.08	0.39	0.15	0.21	0.27	0.62	0.83	0.38
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Coast	0.09	0.48	0.24	0.27	0.37	0.72	1.01	0.31
	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Eastern	0.08	0.45	0.19	0.23	0.35	0.70	0.95	0.32
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
North Eastern	0.08	0.50	0.32	0.37	0.53	0.71	1.13	0.42
	(0.00)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Nyanza	0.08	0.46	0.22	0.23	0.33	0.68	0.95	0.26
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Rift Valley	0.08	0.48	0.24	0.26	0.39	0.75	1.03	0.30
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)
Western	0.08	0.50	0.22	0.22	0.34	0.70	0.98	0.22
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
All	0.08	0.47	0.22	0.24	0.36	0.73	1.00	0.29
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
(b) *Relative contributions (percent)*:								
Central	0.8	22.4	3.2	6.5	10.5	56.6	100.0	0.38
	(0.0)	(0.4)	(0.1)	(0.5)	(0.8)	(0.7)	-	(0.03)
Coast	0.7	22.8	5.6	6.9	13.4	50.6	100.0	0.31
	(0.0)	(0.8)	(0.4)	(0.3)	(0.9)	(1.7)	-	(0.01)
Eastern	0.7	22.2	3.9	5.8	13.6	53.8	100.0	0.32
	(0.0)	(0.8)	(0.2)	(0.3)	(0.5)	(0.8)	-	(0.01)
North Eastern	0.6	19.8	7.9	10.5	22.0	39.3	100.0	0.42
	(0.0)	(0.6)	(0.4)	(0.7)	(0.9)	(1.1)	-	(0.02)
Nyanza	0.8	23.9	5.5	5.9	12.0	52.1	100.0	0.26
	(0.0)	(0.4)	(0.2)	(0.2)	(0.2)	(0.6)	-	(0.01)
Rift Valley	0.6	21.4	5.4	6.4	14.3	51.9	100.0	0.30
	(0.0)	(0.6)	(0.2)	(0.4)	(0.3)	(0.7)	-	(0.02)
Western	0.7	26.3	5.2	5.2	12.0	50.6	100.0	0.22
	(0.0)	(0.3)	(0.2)	(0.1)	(0.4)	(0.4)	-	(0.01)
All	0.7	21.7	4.9	5.9	13.1	53.7	100.0	0.29
	(0.0)	(0.3)	(0.1)	(0.2)	(0.3)	(0.6)	-	(0.01)

Note: the table reports direct fixed-effects variance decomposition estimates, based on the unrestricted linear model representation; standard errors (in parentheses) are calculated via a clustered bootstrap procedure.

Table D6: Regional variance decomposition for Tanzania, unrestricted linear model

Region	Indiv. σ_x^2	Hhld σ_h^2	Sch. σ_s^2	Sort σ_v^2	Comm. $2\Sigma_{hs}$	Resid. σ_e^2	Total σ_t^2	Correl. ρ_{hs}
(a) *Absolute contributions (st. dev. units × 100)*:								
Arusha	0.11	0.49	0.19	0.25	0.39	0.70	1.00	0.35
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Dar Es Salaam	0.10	0.45	0.17	0.24	0.36	0.68	0.95	0.38
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Iringa	0.12	0.47	0.21	0.30	0.42	0.67	1.00	0.45
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)	(0.02)
Kagera	0.13	0.50	0.22	0.30	0.40	0.63	0.98	0.40
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
Kigoma	0.14	0.50	0.24	0.33	0.40	0.58	0.96	0.44
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)
Ruvuma	0.11	0.50	0.15	0.27	0.36	0.65	0.96	0.50
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)
Singida	0.13	0.49	0.20	0.28	0.42	0.64	0.98	0.41
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Tabora	0.15	0.51	0.24	0.32	0.40	0.63	1.00	0.43
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.03)
Tanga	0.12	0.48	0.19	0.28	0.39	0.65	0.96	0.41
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
All	0.13	0.49	0.21	0.29	0.40	0.68	1.00	0.41
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
(b) *Relative contributions (percent)*:								
Arusha	1.3	23.9	3.5	6.4	15.4	49.6	100.0	0.35
	(0.1)	(0.7)	(0.2)	(0.4)	(1.1)	(1.6)	-	(0.02)
Dar Es Salaam	1.2	22.8	3.3	6.6	14.6	51.6	100.0	0.38
	(0.1)	(0.8)	(0.3)	(0.5)	(0.7)	(1.7)	-	(0.02)
Iringa	1.5	22.2	4.4	8.9	17.4	45.5	100.0	0.45
	(0.2)	(1.0)	(0.5)	(0.9)	(1.0)	(2.6)	-	(0.02)
Kagera	1.9	25.7	5.1	9.3	16.5	41.6	100.0	0.40
	(0.2)	(0.7)	(0.5)	(0.9)	(0.7)	(1.6)	-	(0.02)
Kigoma	2.1	27.0	6.4	11.5	17.1	35.9	100.0	0.44
	(0.2)	(1.4)	(0.7)	(1.1)	(1.0)	(2.5)	-	(0.03)
Ruvuma	1.3	26.9	2.5	8.1	14.5	46.8	100.0	0.50
	(0.1)	(0.9)	(0.3)	(0.7)	(1.0)	(1.6)	-	(0.03)
Singida	1.7	25.2	4.1	8.3	18.4	42.3	100.0	0.41
	(0.2)	(0.9)	(0.5)	(1.0)	(1.4)	(2.2)	-	(0.02)
Tabora	2.1	25.9	5.6	10.3	16.4	39.7	100.0	0.43
	(0.2)	(0.7)	(0.3)	(0.8)	(1.2)	(1.6)	-	(0.03)
Tanga	1.6	25.1	3.9	8.2	16.0	45.2	100.0	0.41
	(0.2)	(1.0)	(0.4)	(0.6)	(1.0)	(1.9)	-	(0.02)
All	1.6	24.0	4.3	8.4	16.0	45.6	100.0	0.41
	(0.1)	(0.5)	(0.2)	(0.4)	(0.4)	(1.2)	-	(0.01)

Note: the table reports direct fixed-effects variance decomposition estimates, based on the unrestricted linear model representation; standard errors (in parentheses) are calculated via a clustered bootstrap procedure; regions have been combined for simplicity (see Appendix C).

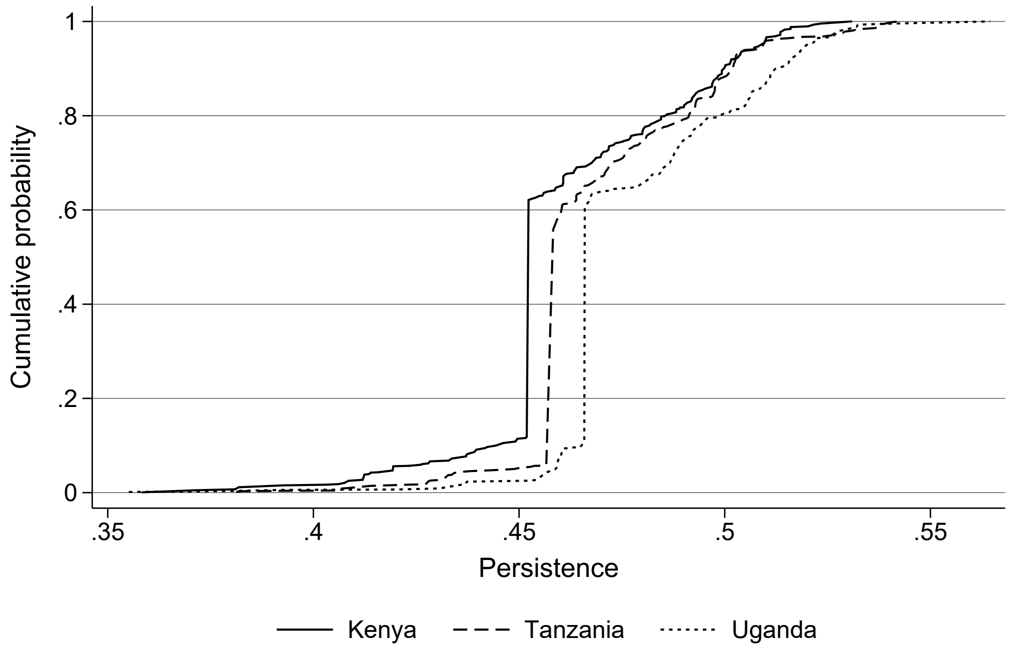
Table D7: Regional variance decomposition for Uganda, unrestricted linear model

Region	Indiv. σ_x^2	Hhld σ_h^2	Sch. σ_s^2	Sort σ_v^2	Comm. $2\Sigma_{hs}$	Resid. σ_e^2	Total σ_t^2	Correl. ρ_{hs}
(a) *Absolute contributions (st. dev. units $\times$ 100)*:								
Central	0.09	0.43	0.23	0.21	0.35	0.71	0.96	0.22
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Eastern	0.08	0.47	0.22	0.19	0.33	0.68	0.94	0.18
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Northern	0.13	0.48	0.24	0.23	0.37	0.70	0.99	0.22
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Western	0.09	0.49	0.25	0.23	0.38	0.69	0.99	0.23
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
All	0.10	0.47	0.23	0.21	0.36	0.74	1.00	0.21
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
(b) *Relative contributions (percent)*:								
Central	0.8	20.3	5.6	4.8	13.1	55.5	100.0	0.22
	(0.1)	(1.1)	(0.4)	(0.4)	(0.6)	(2.2)	-	(0.01)
Eastern	0.7	25.1	5.2	4.2	12.2	52.5	100.0	0.18
	(0.0)	(0.5)	(0.3)	(0.2)	(0.6)	(0.5)	-	(0.01)
Northern	1.6	23.2	5.9	5.2	14.1	50.0	100.0	0.22
	(0.2)	(0.8)	(0.6)	(0.6)	(0.6)	(1.2)	-	(0.02)
Western	0.8	24.2	6.4	5.6	14.9	48.1	100.0	0.23
	(0.1)	(0.4)	(0.3)	(0.3)	(0.4)	(0.6)	-	(0.01)
All	0.9	21.8	5.4	4.5	12.7	54.7	100.0	0.21
	(0.1)	(0.8)	(0.3)	(0.3)	(0.4)	(1.7)	-	(0.01)

Note: the table reports direct fixed-effects variance decomposition estimates, based on the unrestricted linear model representation; standard errors (in parentheses) are calculated via a clustered bootstrap procedure.

E Additional figures

Figure E1: Distribution of estimates of raw persistence parameter, by districts



Note: lines give the cumulative distribution of estimates for the inequality-constant intergenerational persistence parameter (see equation 7).

Source: own calculations.