What matters for learning in East Africa?
A comparison of education production functions between and within countries

Preliminary draft – not for citation

Sam Jones

University of Copenhagen

November 6, 2012

Abstract

East African education systems have undergone rapid transformation over recent decades. Nonetheless, little is known about what matters for learning across the region. Based on a unique micro-dataset incorporating test scores for almost 350,000 school-aged children in Kenya, Tanzania and Uganda (enrolled and not enrolled), this paper provides reduced-form estimates of educational production functions at the regional, country and sub-group levels. The possibility of parameter heterogeneity is admitted at the outset and motivates development of a sample-weighted mean-group estimator that ensures consistent estimation of average marginal effects across heterogeneous groups, such as geographical units. The findings confirm that parameter heterogeneity is substantial. For instance, the average child in Uganda acquires basic skills much later than in Kenya and Tanzania, holding all other inputs constant. However, there are also regularities. The contribution of family background factors to learning is most crucial. School inputs are statistically significant but play a moderate role on average. Amongst these, teacher-pupil ratios and the size of schools are both associated with the largest positive gains in learning.

1 Introduction

The role of education in supporting social and economic development is widely recognized. Wage or income regressions almost invariably show a premium to education, at least beyond very low levels (Psacharopoulos and Patrinos, 2004). Spill-overs from education to fertility and health, among other things, are also potentially large (Schultz, 1999; Kravdal, 2002). Aside from individual benefits, a growing body of research suggests that population-wide improvements in skills can boost economic growth on aggregate (Sianesi and Reenen, 2003; Hanushek and Wößmann, 2008). However, this literature also suggests that the macroeconomic contribution of education is often poorly captured by traditional measures of educational performance. Rather, acquired cognitive skills play a key role in boosting aggregate income.

Despite this evidence, there is less consensus about what matters for the production of education, both within and between countries. Economic theory suggests that a wide range of inputs are likely to affect education outcomes. These include family background, school resources and other local environmental factors. The workhorse empirical model used to understand the role of these different inputs is an educational production function. The small number of comparable estimates of production functions from different countries point to similarities and important differences in the role of inputs. Family background factors are frequently found to be crucial, but not by the same amount. For instance, in his review of educational production across East Asia, Wößmann (2005) concludes that family background is a much stronger predictor of individual performance in Korea and Singapore than in Hong Kong or Thailand. Similarly, in a review of 35 European countries, Wößmann (2003) concludes that neither differences in school resources nor background factors can adequately account for (large) disparities in student attainment. Differences in institutional incentives appear to fundamental.

High quality evidence about education production in the developing world is both scarce and indicates few consistent findings (for reviews see Glewwe, 2002; Glewwe and Kremer, 2006, *inter alia*). This is especially the case for low income Africa. In large part this reflects the fact that representative data about learning outcomes, measured as acquired cognitive skills, has been rare (in the public domain). Moreover, existing studies typically employ small and/or highly specific samples, casting doubt on the wider applicability of the findings. For instance, the extensive SACMEQ exercise, analysed by Michaelowa (2001) among others, provides observations for an average of only about 3,000 students enrolled in the sixth grade per country. Given that a large proportion of students in low income Africa either never enroll, start late, drop-out, or repeat grades, this is hardly informative of the wider school age population.

The absence of previous research motivates a deeper understanding of what matters for educational production in East Africa. This agenda is also warranted on additional grounds. First, education systems in the region have undergone rapid transformation over recent decades. Economic difficulties during much of the 1980s and 1990s severely affected the provision of public goods and services (from a low base). As late as 1999, official figures compiled by UNESCO (2011) show that net enrolment rates in primary education were as low as 49% in Tanzania and 63% in Kenya.[1] However, as fiscal positions improved, growth returned and donor governments shifted attention to social sectors, rapid expansion of access to education has been achieved. Bolstered by the abolition of user fees (in Uganda in 1997, Tanzania in 2002 and Kenya in 2003), the vast majority of children now attend primary school in these countries. By 2009, official estimates of net enrolment were 83% in Kenya, 97% in Tanzania and 92% in Uganda (UNESCO, 2011).

Second, expansion of access has been accompanied by unease regarding quality. Nishimura et al. (2008), for example, notes that in Uganda between 1997 and 2004, primary school enrolment increased by 141% while the number of teachers and schools increased by 41%. Negative impacts such as larger class sizes, inadequate class rooms and poorly educated teachers are frequently attributed to the introduction of universal primary education (UPE).[2] Thus, widely praised measures of performance in the sector – namely, access rates and reported grade attainment – may tell us little about educational production in the sense of attained cognitive skills. Moreover, 'free' primary education does not necessarily translate to equality of opportunity. As Mugo (2007) notes from the Kenyan experience, a policy focus on access has tended to neglect issues around improving affordability and reducing educational disparities. Put differently, reducing the nominal price of schooling represents a change in one input to educational production. This may not always be a critical constraint to learning, and could be offset by changes in other inputs (e.g., in the home). A multivariate analysis of educational production functions can help shed light on these issues.

Finally, East Africa is marked by substantial diversity in economic, social and cultural conditions. This raises the question whether a single or homogeneous educational production function is appropriate. Nubé and Sonneveld (2005), for instance, document distinct sub-regional variations in the prevalence of underweight children across sub-Saharan Africa. Consumption poverty

[1] No data is available for Uganda from the same source and year.

[2] One author cites a local Ugandan brick layer: "UPE emphasizes promotion rather than efficiency. It is so bad that children in UPE schools can neither read nor write their names, yet they keep on being promoted to higher classes. UPE promotes failures, for example, a child who scores 80 marks out of 400 can take the 12th position out of 600 pupils. These are all failures and yet they are promoted to the next class." (Okuni, 2003, p. 34).

and other forms of deprivation (e.g., asset ownership; rates of enrolment) typically follow clear geographic patterns, with rural areas, conflict zones and marginal agro-ecological zones frequently worst affected (e.g., see Bird et al., 2010). This has a direct bearing on the environment in which children learn, including their readiness to absorb new cognitive skills if and when they begin primary school. It also is important to recall that large swathes of the adult population remain illiterate, meaning that many children are not able to benefit from extensive parental assistance during their basic education. The point is that educators face very different sets of challenges. Thus, variation in what matters for learning may be present.

The broad objective of this paper is to characterize education production functions in East Africa. Attention is placed on identifying similarities *and* differences in the role of alternative inputs into the 'production process'. The analysis is made possible by a new micro-dataset, comprising household-level interviews of almost 350,000 children aged 6 to 16 in Kenya, mainland Tanzania and Uganda. The data, collected by the independent Uwezo initiative in 2011, is representative at the district level and incorporates child-level scores on basic literacy, comprehension and numeracy from oral tests administered by the survey team when visiting each household. The dataset also includes information about the child's family background and local school environment based on school surveys undertaken at the village level. Importantly, the samples encompass children that are attending school and those who are not currently attending. Consequently, possible biases from incorporating only children who are enrolled at a specific grade can be avoided.

As a point of departure, the analysis does not presume that a homogeneous education production function is likely to be particularly informative. Not least for the reasons given above, substantial parameter heterogeneity is to be expected *a priori*. This view bolstered by evidence from other areas of economics, such as the macroeconomic growth literature, which frequently casts doubt on the simplifying assumption of parameter homogeneity (e.g., see Temple, 2002). On closer scrutiny this viewpoint has profound implications, even for empirical estimates of education production functions at the aggregate level. It can be shown that under highly plausible conditions in grouped cross-sectional data, standard least squares estimators frequently yield parameter estimates that have no meaningful economic interpretation. As such, even fixed effects estimators – frequently employed to address potential bias from omitted variables – will be inconsistent for policy-relevant summary parameters. In turn, this motivates less restrictive approaches. One of these, presented here, is the sample-weighted counterpart of the mean-group estimator due to Pesaran and Smith (1995).[3] This estimator provides a formal means to test for

[3] See also Lee et al. (1997); Haque et al. (1999).

parameter heterogeneity as well as the consistency of the more restrictive fixed effects estimator. As such, it provides a robust basis on which to investigate and characterize what matters for education in East Africa.

The rest of the paper is structured as follows: Section 2 gives an overview of the survey data, including the test scores. Section 3 presents the empirical framework, based on a reduced-from education production function. In particular, threats to validity including parameter heterogeneity are discussed. Section 4 presents the results in three stages, moving from the aggregate to the particular. Specifically, regional reduced-from estimates of what matter for learning are reported, followed by country-specific estimates, followed by estimates for sub-groups of interest. The latter is made possible by the richness of the data and allows, among other things, an appreciation of differences in the magnitude and importance of inputs into learning across different socio-economic groups. In reporting the results, however, it should be noted that whilst due attention is made to parameter heterogeneity, space considerations limit the extent to which the full range and diversity of findings can be discussed. Section 5 concludes.

2 The Uwezo surveys

2.1 Overview

In response to growing concerns about the quality of education across East Africa, Uwezo – an independent initiative that promotes access to information, citizen agency and improved service delivery – has assessed the acquired cognitive skills of a large number of children of school age across East Africa.[4] To date, two rounds of the Uwezo surveys have been completed. The first was undertaken in 2009 in Kenya, and in 2010 in mainland Tanzania and Uganda. While the scope of the first round (Uwezo 1) was substantial, encompassing over 150,000 children across the three countries, the second round (Uwezo 2, undertaken in 2011) was designed to ensure a more thorough coverage of administrative districts. The Uwezo 2 survey provides representative samples from 75% of all Kenyan districts, 100% of all Tanzanian districts and all but one district in Uganda.[5] Due to the enhanced coverage of the second round surveys, as well

[4]Their approach is inspired by a similar exercise carried out in India by the Assessment Survey Evaluation Research Centre (ASER). Since 2005, ASER has annually surveyed the literacy and numeracy abilities of over 700,000 children.

[5]Due to frequent changes of administrative boundaries, the administrative divisions from each countries' most recent Population and Housing Census were used.

as small changes in the questionnaire between rounds (see below), the Uwezo 2 data is in focus here. Nonetheless, the analysis in Uwezo (2012) indicates that aggregate (national) results from the Uwezo 1 and Uwezo 2 surveys are extremely similar.

In terms of the questionnaire design, in each primary sampling unit (PSU, typically a village), information was gathered in four steps. First, a single local government primary school, randomly pre-selected by the Uwezo district coordinator, was surveyed. This involved administering a series of questions to the most senior staff member available (ideally the head teacher) and direct observation of pupil and teacher attendance numbers, as well as school conditions according to a simple questionnaire. Second, the chief or administrative head of the PSU was visited. Aside from establishing permission to visit individual households, as well as confirming the validity of the sample frame for the village, he or she was asked a series of simple questions about the PSU (e.g., is there access to clean water?). Third, the selected households were visited. In each case, the head of the household was asked a short set of simple questions and details of the children in the household were recorded (e.g., age, gender, whether or not attending school etc.).

Table 1 provides descriptive statistics from the dataset used hereafter. Panel (a) summarises the coverage of the surveys (after data cleaning). For each country more than 2,000 individual schools, 35,000 households and 100,000 children are included. Encompassing a total of almost 350,000 children, this represents one of the largest and geographically most comprehensive non-census surveys of its kind, especially for a low income African country. Panel (b) summarises information from the survey of government primary schools. Pupil (teacher) attendance rates reflect the ratio of the number of pupils (teachers) observed on the day of fieldwork to the number of formal or officially registered pupils (teachers) according to information provided by the interviewed staff member. In a small number of cases the ratio exceeds one (100%), meaning that there are in fact more children in school than are officially enrolled. This may be due to measurement errors, but in some instances may also reflect difficulties faced by households in formally registering their children as enrolled. Other measures, such as the number of books per pupil are calculated in a similar fashion. It should be highlighted that these variables are constant at the PSU-level (i.e., they do not vary by child or household). This assures consistency across the survey instruments and has the advantage of avoiding bias from within-school sorting or selection effects (Ammermüller et al., 2005). However, it may also contribute to attenuation bias.

Panel (c) summarises some of the principal variables that vary at the household level. These include household composition and it's socio-economic status. The latter is calculated from

Table 1: Selected descriptive statistics from the Uwezo 2 dataset

		All	(s.d.)	Kenya	(s.d.)	Tanzania	(s.d.)	Uganda	(s.d.)
(a)	No. districts sampled	320		122		119		79	
	No. schools sampled	9,317		3,474		3,728		2,115	
	No. households sampled	152,524		55,118		60,372		37,034	
	No. children sampled	348,167		129,665		116,232		102,270	
(b)	Total number of enrolled pupils	560.6	(326.5)	470.1	(292.3)	570.6	(310.8)	683.4	(357.2)
	Pupil attendance rate	84.5	(13.1)	90.7	(8.4)	82.7	(14.3)	77.8	(17.3)
	Teacher attendance rate	84.8	(32.4)	86.6	(17.5)	79.9	(38.4)	90.1	(37.7)
	Teachers per 100 pupils (reported)	2.7	(4.2)	3.2	(1.8)	2.7	(6.4)	2.1	(0.7)
	Teachers per 100 pupils (observed)	3.2	(11.0)	3.5	(7.4)	3.3	(15.9)	2.3	(2.7)
	No. of books for every 100 pupils	28.1	(25.9)	39.7	(20.3)	18.8	(28.3)	25.8	(22.2)
(c)	Household is poor	43.8	(24.6)	37.8	(23.5)	47.2	(24.9)	45.4	(24.8)
	Household is ultra-poor	7.4	(6.9)	8.8	(8.0)	6.5	(6.1)	7.3	(6.8)
	Number of children in household	2.9	(1.5)	2.9	(1.4)	2.5	(1.2)	3.5	(1.7)
	Number of other household members	4.2	(2.9)	3.6	(2.0)	4.7	(2.9)	4.0	(3.4)
(d)	Age of mother	36.0	(9.8)	34.7	(10.9)	37.9	(8.8)	35.0	(9.5)
	Mother has primary education	62.4	(23.5)	53.7	(24.9)	72.0	(20.2)	58.9	(24.2)
	Mother has secondary education	13.2	(11.5)	20.2	(16.1)	6.0	(5.6)	15.2	(12.9)
	Child's age	10.9	(3.1)	10.6	(3.2)	11.2	(2.9)	10.6	(3.1)
	Child reports to be enrolled in school	86.6	(11.6)	86.2	(11.9)	84.0	(13.4)	89.9	(9.1)
(e)	Reading test score (%)	56.3	(39.6)	69.2	(34.8)	56.6	(41.1)	44.3	(38.0)
	Comprehension test score (%)	34.6	(22.6)	44.6	(24.7)	37.7	(23.5)	22.1	(17.2)
	Numeracy test score (%)	60.9	(40.2)	72.3	(37.1)	62.6	(39.6)	48.5	(40.2)
	Combined test score (std.)	0.0	(100.0)	32.0	(92.1)	3.3	(99.6)	-32.7	(97.2)

Source: author's calculations from the Uwezo 2 data.

Notes: statistics in panel (a) indicate survey coverage; in the rest of the table statistics are sample means; variables in panel (b) are calculated at the PSU level, remaining variables are calculated over all children of school age; unless indicated, test scores are rescaled to percentages of the maximum mark; literacy and comprehension tests refer to the predominant national language of instruction at primary school; the combined test score, based on the sum of the literacy, comprehension and numeracy tests, is standardized at the regional-level and multiplied by 100.

the count of observed physical and human capital assets of the household, as per the Alkire-Foster multidimensional poverty headcount index (Alkire and Foster, 2011). Due to limits on the number of candidate variables, only six welfare dimensions are used (given equal weight) – access to electricity, access to piped water, ownership of a phone, ownership of a radio, ownership of a TV, and mother's education. Deprivation with respect to the access/ownership categories is defined as absence of that item in the household; deprivation on the final category is defined as the mother having no formal education. A household thus is defined as 'ultra-poor' if it is simultaneously deprived in all dimensions. If a household is not 'ultra-poor', then it is either defined as 'poor' if is deprived in any four of these dimensions, and 'non-poor' otherwise. Admittedly, these distinctions are somewhat arbitrary; however, this measure is transparent and applies without modification across all countries.

Panel (d) describes a number of key variables that vary at the lowest level of aggregation – i.e., children. In each country the target population are children of school age, up to 16. In Uganda and Kenya, the minimum age for starting primary school is six. In Tanzania, primary school starts at seven, meaning that six year-olds were not assessed on their learning outcomes and thus are not included in the dataset. This explains the slightly higher average age of children (and mothers) in the Tanzanian sample. It should also be noted that there is some variation within the household as regards parental characteristics. This occurs for various reasons including where multiple family units reside in the household (defined as taking meals in common) and if children are orphaned or fostered out.

2.2 Uwezo tests

Within each surveyed household, all children of school age were administered a set of basic oral literacy and numeracy tests. For simplicity, the literacy tests included in the tests refer to the national languages of instruction in which pupils are tested at the end of primary school – English and Kiswahili in Kenya and Tanzania; and only English in Uganda. For these languages, the Uwezo literacy tests evaluated simple reading skills in order of increasing difficulty. Based on pre-prepared test cards, children were asked to: recognise a letter from the alphabet, read a word, read a sentence, and read a paragraph (story). Provided the child was able to read at the story level, she was further asked at least one question to assess whether she also comprehended the story. In the numeracy tests, children were asked a set of questions (also from pre-prepared cards) starting with simple arithmetic and either increasing in difficulty, if they were successful, or decreasing in difficulty if not. The numeracy skills assessed in each country thus covered:

number recognition, counting, and the performance of basic calculations with numbers of up to two digits (addition, subtraction and multiplication). The numeracy test was administered in the predominant language of instruction (English in Kenya and Uganda; Kiswahili in Tanzania). In each test the child was given a score indicating the maximum skill level achieved.

Three aspects of the tests can be highlighted. First, the tests were designed by local education experts to reflect competencies stipulated in the national curricula at the Standard 2 level. That is, they test skills that should be achieved by the majority of pupils after two years of completed primary schooling. Therefore, they are set at a low level.[6] Second, whilst all the tests had the same objectives and format, their precise content differed from country to country. This is largely because of differences in the national curricula, which are slight but somewhat more material at the highest levels of ability (e.g., as regards expected vocabulary). In the present analysis, these differences have been ameliorated by only including in the test scores those skill levels that were assessed in all three countries. Nonetheless, the same mark in Kenya cannot be considered as representing precisely the same absolute level as in Uganda or Tanzania. Indeed, the point of the tests is not to measure underlying cognitive skills in a uniform way, but rather to assess children against the skills they should master by a given grade of primary school.

Third, the internal reliability of the Uwezo tests is high. Taking the literacy, comprehension and numeracy tests together, measures of Cronbach's alpha are around 0.87, whether calculated for each country individually or for all countries together. As Tavakol and Dennick (2011) explain, this suggests that the different tests measure the same concept or construct (i.e., competence at a Standard 2 level) and, thus, can be combined into a single overall score. Results from the various Uwezo tests are summarised in Panel (e) of Table 1, shown as percentages of the maximum possible mark. The combined score, which only employs scores for the predominant language of instruction in the literacy and comprehension components, is also shown in the table – reported on a standardized basis. The latter is frequently used in learning assessments, one of the reasons being that estimated marginal effects then represent standard deviation units. The same practice is adopted here and, unless otherwise indicated, the focus is on the standardized combined test score, which takes a mean of zero and standard deviation of 100 (calculated on a pooled basis, across East Africa).

[6] Examples of the tests are available on the Uwezo website.

2.3 Test outcomes

Before proceeding to a multivariate framework, it is helpful to get a quick sense of what the data reveals about learning in East Africa. Appendix Figure A1 plots the share of children in each country at each age that reach the maximum combined test score (i.e., they pass each of the literacy, comprehension and numeracy tests).[7] Two points stand out. First, a small proportion of children under the age of ten are able to pass the tests. This suggests that basic competencies are learnt slowly, and much later than would be expected assuming a child enters primary school at the correct starting age and progress 'normally'. Equally, a significant share of children aged 16 remain unable to pass the tests and therefore are unlikely to be functionally literate or numerate. Second, there is a distinct ordering of the countries in terms of average performance, evident even at young ages. Kenyan children of a given age tend to outperform children in Tanzania, who in turn outperform those of Uganda. Recall that although the tests differ slightly in content, the interpretation is that more children of a given age in Tanzania attain core Standard 2 level skills (as per their country's curricula) versus children of the same age in Uganda.

Aggregate results hide significant local variation. This is substantiated from the large standard deviations on the test scores shown in Table 1. The degree of geographic variation is indicated by Figure A2, which plots district-level average pass rates on the combined test score for all children aged ten and above. While the broad ranking of the three countries is preserved, there is very substantial distributional overlap. In Kenya, for example, there are a number of very poorly performing districts where less than 40% of 10-16 year olds are able to pass the tests. These largely correspond to the more challenging arid zones. There are also a substantial number of very poorly performing Ugandan districts, the worst of which correspond to areas of greatest insecurity. Overall, this motivates a more rigorous analysis of what matters in East Africa, including the extent to which the characteristics of educational production functions vary across the region.

[7]The overall pass rate is chosen here for expositional purposes only. Identical qualitative findings arise from the same analysis employing the (standardized) test score. Also see the analysis in Uwezo (2012).

3 Methodology

3.1 Framework

This subsection introduces the general framework employed to analyse the Uwezo data. Following a rich existing literature (e.g., Todd and Wolpin, 2003), the starting point is a familiar education production function. Empirically, this is implemented as:

$$t_{ijklm} = C_i^{'}\beta_1 + H_j^{'}\beta_2 + S_k^{'}\beta_4 + L_l^{'}\beta_5 + \lambda_m + \epsilon_i \tag{1}$$

where i indexes the child, j her household, k the school she attends (if any, see below) and l her PSU (village). t is a measure of acquired cognitive skills (learning) for a given child, C contains observed factors that vary at the level of the child (her age, gender, birthorder etc.), H contains exogenous factors that principally vary at the level of the household (parent's education, parent's age, household size, socio-economic status etc.), S contains school effects (e.g., number of teachers per student, numbers of books per student, whether the school has a feeding programme) and L other local conditions (whether the village has a primary school etc.). The set of parameters $\lambda_n,\ n \in (i, j, k, l, m)$ represent unobserved effects at the indicated level of aggregation, which are candidates for inclusion as fixed effects. Following the exposition in Aturupane et al. (2011), in the minimum these are expected to include information regarding education preferences and prices, for which there are no direct analogues in the Uwezo data. Finally, ϵ_i represents residual white noise error.

Various aspects of this empirical model merit discussion. First, equation (1) should be understood as a reduced form relationship, encountered widely in the applied literature (e.g., Jimenez and Sawada, 1999; Glick et al., 2011). As such, endogenous choice variables that would enter a structural form model (the education production function 'proper'), such as purchased household-level resources or grade attainment, have been solved out.[8] The reduced form includes only exogenous or prior inputs that are not contemporaneously amenable to manipulation by the child or household in order to influence her acquired level of skills. Thus, the partial derivatives associated with equation (1), capture total marginal effects due to the RHS variables.

Second, the hierarchical structure of the dataset is material. As implied by Section 2, individual

[8] For elaboration of the distinction between reduced and structural form models see Glewwe et al. (2004). Omitted endogenous variables include enrolment in school, grade attainment and indicators of household assistance with school work.

children are nested in households, which are nested in PSUs (where a single school was surveyed), which in turn are nested in administrative districts. The implication is that the inclusion of fixed effects at any given level would absorb all observed and unobserved variables in (1) at the same or higher levels of aggregation. For instance, inclusion of household fixed effects (λ_j) would eliminate all variables and parameters indexed by j, k, l or m.

Third, the model is specified in contemporaneous levels form. That is, the achieved level of learning at the time of the survey is modelled as a function of various background characteristics as well as the contemporaneous (observed) learning conditions. Neither past levels of achievement nor past inputs are included. This is not because they are immaterial. As Todd and Wolpin (2003) clarify, rather stringent assumptions are necessary to obtain consistent estimates of contemporaneous inputs on skills acquisition. However, data constraints mean that alternative specifications of the education production function, such as a value-added model, are not feasible here. This constitutes a limitation. Thus, consistent with much of the existing literature (Wößmann, 2003; Ammermüller et al., 2005; Aturupane et al., 2011; Glick et al., 2011), estimates based on equation (1) represent a linear approximation to the conditional expectation function for learning outcomes. Caution should be exercised, therefore, in giving a causal interpretation to the results.

The challenge of correctly specifying the education production function raises the more general issue of omitted variables bias (OVB). As discussed extensively elsewhere (e.g., Hanushek and Rivkin, 2006; Glewwe and Kremer, 2006), many determinants of learning outcomes are often unobserved. These might include the innate ability of a child, the cognitive endowment of her parents, and the effective quality of her teachers. Where these are correlated with the observed covariates, these will lead to biased estimates on the latter. Two approaches to addressing OVB are frequently encountered. When the aim is to precisely identify specific parameters, instrumental variables techniques can be deployed. These have attractive theoretical properties. However, in practice it is frequently difficult to find: (i) a sufficient number of instruments for numerous explanatory variables; or (ii) instruments that are unambiguously valid for a broad range of population sub-groups. This limits the feasibility of these techniques to situations, such as the present, where one is interested in estimates for parameters across the model. Moreover, if the strength and validity of chosen instruments does not hold, their use can exacerbate bias and imprecision in parameter estimates (Bound et al., 1995).

A second, rather more general strategy, is to use fixed effects estimators. By construction, these absorb all linear variation due to fixed characteristics at the chosen level, regardless of the vector of variables to hand. This is particularly advantageous where units are observed over time (e.g., in a panel setting). However, use of fixed effects is not a panacea. First, they can be inefficient due to a potentially large reduction in model degrees of freedom. Second, specifying fixed effects at low levels of aggregation can sweep-out key variables of ultimate policy interest. This is pertinent here as we do not observe information from different schools for the same household, nor do have reliable measures of school inputs that vary within individual schools (see above). The implication is that results based on household- or school-level fixed effects, although potetnially less prone to bias from omitted variables, might provide little substantive insight as to which factors determine variation in learning outcomes. Third, the attenuating effects of classical measurement error tend to be exacerbated by the inclusion of fixed effects (see Ashenfelter and Krueger, 1994; Bound et al., 2001). This implies that addressing one source of bias (from omitted variables) may come at the expense of inflating others.[9] Fourth, a less well known property of fixed effects estimators is that they only provide meaningful and relevant estimates under fairly restrictive conditions. This is apparent from a consideration of the question of parameter heterogeneity, to which I now consider in more detail.

3.2 Parameter heterogeneity

At the outset, it is important to note that parameter heterogeneity can be viewed as a type of specification problem. As Zietz (2001) notes, if we precisely observe all relevant aspects of economic behaviour and know the appropriate functional form by which a given outcome is determined, then (unobserved) parameter heterogeneity would be irrelevant. This is rarely the case, if ever. In the education literature, the majority of variation in test scores typically remains unexplained, even when the highest quality data and robust techniques are employed (see references in Glewwe et al., 2011). Put another way, various aspects of education production are almost always poorly measured, omitted or incompletely specified in any given empirical application. These are likely to give rise to parameter heterogeneity, especially where they vary systematically across groups. This is supported by Figlio (1999), who finds that a less restrictive functional form yields statistically significant evidence that school inputs are associated with student performance in the USA. In this light, explicit recognition of parameter heterogeneity should not be considered a comprehensive solution to these challenges. This can only be achieved by dealing with specification errors directly (e.g., via instrumental variables methods,

[9] In light of these issues, Deaton (1997) recommends that the trade-off between bias and efficiency in employing fixed effects is resolved on a case-by-case basis.

or using an alternative functional form). However, in the absence of such an unambiguously 'better' specification, failure to take parameter heterogeneity into account can produce parameter estimates that are meaningless and potentially misleading. Thus, in the present case where underlying conditions are highly diverse and data constraints mean that a fairly basic specification is employed, addressing parameter heterogeneity is crucial.

To appreciate how and where bias from unmodelled parameter heterogeneity can arise, a simple exposition is informative. Consider the following general model incorporating group-specific intercepts and slopes:

$$Y = \lambda_g + (X \circ \beta_g)\iota_K + \varepsilon \tag{2}$$

where Y is an $N \times 1$ column vector, X is an $N \times K$ matrix of covariates, λ_g is a column vector of constants that vary with the group index $g \in G$ to which the individual i belongs, and ε is a white noise error term, $\mathrm{E}(X'\varepsilon) = 0$.[10] β_g is an $N \times K$ parameter matrix. Without loss of generality, the rows of the latter can be thought of as being made-up of K fixed parameters that are the same for all individuals, and K group-specific additive factors: $\beta_{ig} = \beta + \alpha_{ig}$, and $\mathrm{E}(\alpha_{ig}) = 0$ by definition. Following Wooldridge (2005), a natural model summary parameter is the $K \times 1$ vector of sample-weighted average (treatment) effects: $\beta^* = \mathrm{E}(\beta_{ig}) = \beta + \mathrm{E}(\alpha_{ig}) = \beta$.

Ignoring slope heterogeneity, the standard fixed effects (FE) estimator can be derived by applying the within transform across groups and running a pooled OLS regression on the following equation:

$$\tilde{Y} = \tilde{X}\beta + \upsilon \tag{3}$$

where $\upsilon = (\tilde{X} \circ \alpha_g)\iota_K + \tilde{\varepsilon}$, and

$$\tilde{X}_g = \left(I_g - \frac{\iota_g \iota_g'}{N_g} \right) X_g, \; \sum_{g \in G} N_g = N$$

Denoting the estimated unconditional variance-covariance matrix of X as $\hat{\Sigma} \equiv (\tilde{X}'\tilde{X})/N$, the familiar FE estimate for β is thus:

$$\hat{\beta}_{FE} = \hat{\Sigma}^{-1} \left(\frac{1}{N} \tilde{X}'\tilde{Y} \right) \tag{4a}$$

$$= \beta + \hat{\Sigma}^{-1} \left(\frac{1}{N} \tilde{X}' \left[(\tilde{X} \circ \alpha_g)\iota_K + \tilde{\varepsilon} \right] \right) \tag{4b}$$

[10] Note that '$\circ$' denotes the Hadamard or pointwise product operator.

The relevant question is when will $\hat{\beta}_{FE}$ be consistent for β^*? Aside from standard regularity assumptions, two general conditions apply. The first is apparent from equation (4b). As per a random coefficients framework (e.g., Wooldridge, 2005), if the group-wise slopes are uncorrelated with the demeaned covariates, $\mathrm{E}(\alpha_g|\tilde{X}_g) = 0$, then it follows that the terms in parentheses in equation (4b) will evaluate to zero in expectation , implying $\mathrm{E}(\hat{\beta}_{FE}) = \beta$. This essentially corresponds to the requirement that any non-linearities and interaction terms that would give rise to marginal effects that differ along the values of the covariates are already incorporated in the model.

Second, if the variance of the covariates is the same across all groups, then $\hat{\beta}_{FE}$ remains consistent. This can be seen from the following decomposition:

$$\hat{\beta}_{FE} - \hat{\Sigma}^{-1}\left(\frac{1}{N}\tilde{X}'\tilde{\varepsilon}\right) = \beta + \hat{\Sigma}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_{ig}x'_{ig}\alpha_g\right) \tag{5a}$$

$$= \beta + \hat{\Sigma}^{-1}\left(\frac{1}{N}\sum_{g=1}^{G}\alpha_g\sum_{i\in g} x_{ig}x'_{ig}\right) \tag{5b}$$

$$= \beta + \hat{\Sigma}^{-1}\left(\sum_{g=1}^{G}\frac{N_g}{N}\alpha_g\left(\hat{\Sigma}\mid g\right)\right) \tag{5c}$$

where movement from the first to the second line is made possible by noting that α_g is fixed for each $g \in G$. Equation (5c) shows that $\hat{\beta}_{FE}$ will be consistent for the average effect when the conditional group-wise variances and covariances are the same – i.e., $\forall g : \left(\hat{\Sigma}\mid g\right) = \hat{\Sigma}$. If not, then the estimate for β_{FE} will be dominated by groups with higher conditional (co)variances.[11] Gibbons et al. (2011) further show this is not a mere technical quibble. Rather, based on replications of papers published in a leading journal, they find substantial differences between the sample-weighted average effect and reported fixed effects results. Similarly, Zietz (2001) presents Monte Carlo evidence that in linear regression models applied to cross-sectional data, the neglect of parameter heterogeneity typically leads to significant bias and renders estimated regression coefficients economically meaningless. The key implication is that where parameter heterogeneity is present, fixed effects results may provide a misleading characterisation of even the general properties of an educational production function.

Following the above, addressing parameter heterogeneity involves two challenges. The first is to test for the extent of parameter variation (across relevant groups). Second, where appropriate,

[11] For similar derivations see Angrist (1998); Juhl and Lugovskyy (2010); Gibbons et al. (2011). Note this result is not dependent on the application of the within transformation.

an alternative and consistent estimator for average effects should be employed. With respect to the first challenge, preliminary insight comes from auxiliary regressions (tests). Following Zietz (2001), there is a direct link between heteroscedasticity and parameter variation. Thus, as recognised early on by Breusch and Pagan (1979), tests of the former can be interpreted as tests for the latter. Monte Carlo evidence provided by Zietz (2001) further suggests that general specification tests, such as that of Ramsey (1969), are also sensitive to unmodelled parameter heterogeneity, especially where parameters are not drawn from the same underlying distribution. These tests are readily available in standard econometric packages and therefore provide a simple means to gauge whether or not heterogeneity is likely to be present.

Specification tests represent indirect evidence of neglected parameter variation. A direct approach, which permits a more complete characterization of parameter heterogeneity, comes via Chow-type tests. Broadly speaking, this involves explicitly including group-specific slopes in a regression model and, subsequently, testing for their significance. These tests can be implemented in various technically equivalent ways. Where there are a small number of groups across which heterogeneity is suspected and/or attention is restricted to heterogeneity in a few specific parameters, then inclusion of relevant dummy variables and their interactions in a single model is appropriate (see Zietz, 2006; Gibbons et al., 2011). However, in more general cases this approach becomes impractical. Thus, deploying separate regressions for each group constitutes a general and flexible approach, as per the original contribution of Chow (1960). Either way, the null hypothesis of parameter homogeneity is given by: $H_0 : \beta_g = \beta \ \forall g \in G$, whilst the alternative is $H_1 : \beta_j \neq \beta_k$ for a non-zero fraction of pairwise slopes $j \neq k; j, k \in G$. A straightforward test of the null is an F-test of a restricted model, including only group fixed effects, versus the unrestricted model:

$$F = \frac{N - G(k+1)}{k(G-1)} \left[\frac{S_{FE} - \sum_{g \in G} S_g}{\sum_{g \in G} S_g} \right] \tag{6}$$

where S_m are the residual sum of squares associated with model m.

The group-specific regressions associated with the flexible version of the Chow test will often be of interest *per se*. However, the average effects can provide a useful overall summary of the relation between the outcome and covariates. An estimator which simply delivers the sample-weighted average effects is the cross-sectional counterpart to the mean-group (MG) estimator due to Pesaran and Smith (1995). The latter was developed for the time-series panel context as a means to address parameter heterogeneity, but can be easily adapted. Specifically, employing group-specific estimates of equation (2), the weighted mean group (WMG) estimator is given by:

$$\hat{\beta}_{WMG} = \sum_{g \in G} \frac{N_g}{N} \hat{\beta}_g = \sum_{g \in G} w_g \hat{\beta}_g \tag{7}$$

which bears a clear resemblance to equation (5c). Two alternative estimates of the variance of β_{WMG} are apparent. The first is the sample-weighted analogue of the non-parametric estimator of the asymptotic variance of the unweighted MG estimator suggested by Pesaran and Smith (1995). This would be:

$$\text{Var}(\hat{\beta}_{WMG}) = \frac{1}{G(1 - \sum_{g \in G} w_g^2)} \sum_{g \in G} w_g \left(\hat{\beta}_g - \hat{\beta}_{WMG} \right)^2 \tag{8}$$

A more direct alternative employs the variance estimates from the G individual regressions. Under the assumption that the draws of β_g are independent across groups, this is given by:

$$\text{Var}(\hat{\beta}_{WMG}) = \text{Var}\left(\sum_{g \in G} w_g \hat{\beta}_g \right) = \sum_{g \in G} w_g^2 \text{Var}(\hat{\beta}_g) \tag{9}$$

In contrast to equation (8), which requires at least a moderate number of groups, the latter estimate is appropriate for G of any size.

The WMG estimator permits a further test of the null hypothesis of slope homogeneity. Under the null, the fixed effects estimator is both consistent and efficient. Under the alternative, the WMG estimator is consistent for the sample-weighted mean. This suggests a Hausman-type test:

$$(\beta_{WMG} - \beta_{FE}) \left[\text{Var}(\beta_{WMG}) - \text{Var}(\beta_{FE}) \right]^{-1} (\beta_{WMG} - \beta_{FE})' \sim \chi^2(K) \tag{10}$$

where K denotes the number of slope parameters estimated in the restricted model. As Pesaran and Yamagata (2008) note, this test lacks power if parameter heterogeneity takes the form of random coefficients (i.e., α_g are mean-zero random draws from the same distribution). However, as per the previous discussion, these are precisely the conditions where there is no difference between the WMG and FE estimates in expectation. Thus, for the present purposes, this limitation to the test is not germane.

3.3 Empirical strategy

The main goal of the paper is to shed light on what matters for learning in East Africa taking due account of regularities and differences, both within and between countries. Cognizant of the potential bias due to parameter heterogeneity, the empirical analysis proceeds in three steps, starting at the aggregate and moving down to the particular. First, I consider the broad regularities in educational production across the region. This corresponds to what Fuchs and Wößmann (2007) refer to as an 'international' education production function, which is of interest because it represents a parsimonious general summary of what matters for learning. Moreover, it indicates the extent to which the observed explanatory variables are able to account for differences in student performance on aggregate.

Second, based on evidence that parameter heterogeneity is material, it is useful to compare results for each country individually. These correspond to estimates of national (reduced-form) educational production functions (for similar exercises see Ammermüller et al., 2005; Wößmann, 2005). Third, attention is turned to population sub-groups within each country. These are of interest because geographic differences may not be the only source of parameter heterogeneity. Such estimates are also relevant for policy-makers. For instance, they can inform discussions about the possible distribution of benefits from general versus targeted educational interventions.

Before proceeding to the results, a few practical issues deserve comment. Due to the nature of the data collection process, there is no guarantee that surveyed children who attend primary school actually attend the same school covered by the Uwezo survey. In other words, the Uwezo data does not assure a one-to-one match between (all) children and their schools. In the minimum, this is likely to contribute to measurement error in the school-level variables, generating attenuation bias. One way to address this would be to restrict the sample. As the data flags whether the child attends the surveyed school, a sub-sample of children could be perfectly matched to the sampled Uwezo school. However, by definition this would exclude children either not attending school or attending non-government schools. Thus the sample would be unrepresentative of the broader population of school age children. Consequently, the focus of the analysis remains on the full sample.[12]

Given the above, a key difficulty is how to associate school characteristics to children that either are out of school or are enrolled in private school. To the author's knowledge this problem has not been confronted in the literature. To start, it should be acknowledged that whether or not children attend a government primary school, the characteristics of local schools enter their (and their parents') information set. For instance, the existence of a particularly poor quality school in the locality could lead parents to conclude that attending school would not generate

[12] In a more detailed background paper, I run sensitivity tests to consider the effect of different specification and sampling assumptions. The results show only moderate sensitivity, but the author concludes that the full sample estimates are most robust and informative.

higher expected future earnings. Thus, school characteristics influence learning directly, via their contribution to skills acquisition, and indirectly, via enrolment and attendance decisions. In this light, a simple (and easily interpretable) procedure is to associate metrics of hypothetically-available schooling conditions with children not enrolled in government schools. Parameter estimates for such schooling variables would thus combine direct and indirect influences of such conditions on skill acquisition.[13]

Adopting this approach, the issue is how to operationalize 'hypothetically-available' schooling conditions. To do so, the following procedure is employed: (i) for all children attending government schools, their schooling conditions are those of the surveyed school in the PSU where they reside; (ii) for children who are not enrolled in school, their 'hypothetically-available' schooling conditions are as least as good as the 'worst' schools in the district; and (iii) for children enrolled in private primary schools, their schooling conditions are likely to be at least as good as the 'best' government schools in the district. Specifically, for the non-enrolled group, schooling conditions are approximated by the first decile of values for each observed schooling variable within the corresponding district. For those attending private schools, the ninth decile of values is employed.[14] This approach has the advantage of ensuring moderate variation in schooling conditions at the PSU-level.

Finally, as with almost all survey data, there are missing values on relevant variables for a non-negligible number of observations (individuals). Analysis also suggests that these are unlikely to be missing at random. As such, results would be biased if these cases were excluded from the analysis. To address this problem I follow a similar procedure to that of Wößmann (2003). Specifically, missing school-, household- and child-level values are imputed using multiple regression methods on core variables. Additionally, dummy variables are included at the household- and school-level that represent the count of variables for each unit that have been imputed. Additionally, in a small number of cases no school was found in the locality. Regression techniques cannot be applied here; thus missing school data is imputed using district-level average school characteristics. An additional dummy variable is included in the model to flag these observations.

[13] An alternative would be to model the enrolment decision separately (in a first stage) and then estimate learning outcomes conditional on the enrolment decision in a second stage. Whilst feasible, this approach adds substantial complexity to the estimation procedure and interpretation of results, especially in the presence of fixed effects. Moreover, the validity of estimates in a second stage typically depends on a non-parametric identification procedure – i.e., inclusion of variables in the first stage that explain enrolment but not outcomes. These variables are hard to find and specification errors in the first stage can seriously contaminate the second stage. For these reasons, a more simple and direct approach is favoured here.

[14] Maxima and minima are not used to avoid contamination from extrema.

4 Results

4.1 Regional production

The first step of analysis focuses on regional reduced-form educational production function estimates. Table 2 presents the results. Movement horizontally across the columns moves from more to less restrictive estimators. Column (I) is a pooled ordinary least squares (OLS) estimator, which ignores any potential heterogeneity in either intercepts or slopes across countries. Column (II) adds country fixed effects; and Column (III) is the sample-weighted mean group estimator, applied using the countries as the grouping variable. The parameters here represent the weighted average of estimates from separate regressions in which the intercepts and parameter slopes differ at the country-level.[15] Columns (IV) and (V) focus on variation within districts. The former is a simple fixed effects estimator, which assumes slope coefficients are the same across districts. The latter is the sample-weighted mean group estimator from 320 separate district-level regressions.

What do we learn from these results? A general point is that there are both statistically significant and economically meaningful differences between the models – i.e., econometric choices matter. This is most apparent with respect to school resource inputs, suggesting that parameter heterogeneity is most acute amongst these variables. For instance, the parameter on class size (teachers per 100 pupils) is estimated to be around one in the OLS and fixed effects (FE) regressions. In contrast, the WMG estimates are larger by a factor of five (in the case of column III) or ten (column V). This is a prominent example of the result, discussed in Section 3.2, that in the presence of parameter heterogeneity, pooled OLS and fixed effects estimators often yield estimates that are inconsistent for any meaningful summary parameter. Additionally, as indicated, material differences emerge depending on the geographical unit within which identification is derived (e.g., countries versus administrative districts). This applies generally across different types of explanatory variable, suggesting that there are significant differences within each country as regards the means of the covariates and, possibly, unobserved factors. For the latter reason, district-level results are likely to be less biased from the point of view of more precisely identifying the average effect of inputs to educational production. This conclusion is supported by the parameter heterogeneity tests, discussed in Section 3.2. These are summarised

[15]Note that the employed sample weights correspond to appropriately scaled population expansion factors (for children of school age). Thus, the countries are not given an equal weighting. Rather, these estimates can be seen as referring to the average child in the region taken as a whole.

Table 2: Estimates of the reduced-form regional educational production function

Estimator →	OLS (I)	FE (II)	WMG (III)	FE (IV)	WMG (V)
Child's age	29.78***	28.92***	30.02***	28.66***	29.48***
	(0.3)	(0.3)	(0.3)	(0.2)	(0.2)
Child's age squared	-1.31***	-1.23***	-1.34***	-1.22***	-1.29***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Child's birthorder	-4.22***	-4.20***	-4.67***	-4.31***	-4.29***
	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)
Child is female	5.16***	5.04***	5.07***	5.03***	5.25***
	(0.4)	(0.4)	(0.4)	(0.4)	(0.3)
Father is not in household	22.00***	10.94***	10.33***	10.23***	7.59***
	(1.7)	(1.6)	(1.6)	(1.5)	(1.6)
Mother is not in household	25.17***	19.73***	20.51***	17.90***	12.79***
	(2.0)	(2.0)	(2.0)	(1.9)	(1.7)
Age of mother	0.43***	0.34***	0.34***	0.32***	0.28***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Age of father	0.17***	-0.06*	-0.04	-0.03	0.01
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Mother has primary education	6.41***	4.71***	4.84***	4.61***	4.87***
	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)
Father has primary education	4.73***	6.54***	5.90***	6.33***	4.08***
	(0.9)	(0.8)	(0.8)	(0.8)	(0.8)
Mother has secondary education	19.14***	17.92***	18.55***	15.93***	17.17***
	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)
Father has secondary education	16.63***	19.57***	19.45***	18.14***	16.17***
	(1.1)	(1.0)	(1.0)	(0.9)	(1.0)
No. children in household	1.10***	3.08***	3.38***	3.82***	3.67***
	(0.3)	(0.2)	(0.2)	(0.2)	(0.2)
No. other household members	-0.62***	-0.79***	-0.80***	-0.42***	-0.51***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Household is poor	-15.98***	-15.08***	-13.94***	-12.09***	-10.44***
	(0.6)	(0.5)	(0.5)	(0.5)	(0.5)
Household is ultra-poor	-27.42***	-26.61***	-24.50***	-21.77***	-20.52***
	(1.2)	(1.1)	(1.1)	(1.0)	(1.3)
Household is aware of Uwezo	20.66***	0.20	0.28	0.01	1.74*
	(0.4)	(0.6)	(0.8)	(0.6)	(0.7)
Missing obs. on core household vars.	-5.89***	-5.55***	-5.60***	-5.32***	-4.94***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
School size (log.)	2.27**	5.92***	8.59***	12.71***	17.30***
	(0.8)	(0.8)	(0.7)	(0.8)	(0.8)
Teachers per 100 pupils (reported)	1.02*	0.92*	5.23***	0.92**	9.85***
	(0.5)	(0.4)	(0.5)	(0.3)	(0.6)
Teacher attendance rate	0.04***	0.07***	0.11***	0.06***	0.15***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Table 2: Estimates of the reduced-form regional educational production function

Estimator →	OLS (I)	FE (II)	WMG (III)	FE (IV)	WMG (V)
No. of books for every 100 pupils	0.16***	0.11***	0.09***	0.09***	0.10***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Number of toilets for every 100 pupils	0.20	0.15	0.10	0.12	0.83**
	(0.1)	(0.1)	(0.1)	(0.1)	(0.3)
School provides subsistence to pupils	4.34***	5.99***	4.93***	5.12***	2.00*
	(0.8)	(0.8)	(0.7)	(0.8)	(1.0)
School has access to clean water	5.26***	9.07***	7.99***	4.37***	0.34
	(0.9)	(0.8)	(0.8)	(0.8)	(0.9)
Head teacher present	9.20***	9.45***	8.79***	10.52***	7.19***
	(0.8)	(0.8)	(0.8)	(0.8)	(0.8)
Missing obs. on core school vars.	0.78***	0.95***	0.96***	0.57***	1.00***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.3)
No school-level data available	-7.49**	-5.11*	-2.56	-4.27*	0.46
	(2.6)	(2.2)	(2.4)	(2.1)	(1.0)
No. of primary schools in village	2.70***	1.59***	1.27***	1.10***	1.30***
	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)
No. of secondary schools in village	3.40***	1.87**	2.22***	1.31*	1.55**
	(0.6)	(0.6)	(0.6)	(0.5)	(0.5)
Village has access to clean water	-2.90***	0.18	0.07	1.10	0.60
	(0.8)	(0.7)	(0.7)	(0.7)	(0.7)
Village has access to electricity	9.64***	7.52***	6.29***	4.05***	3.39***
	(0.8)	(0.7)	(0.7)	(0.7)	(0.7)
Grouping variable	-	Countries	Countries	Districts	Districts
No. groups	-	3	3	320	320
Obs.	348,167	348,167	348,167	348,167	348,167
Estimated parameters	33	33	99	353	10,560
R-squared (centered) adj.	0.460	0.440	0.470	0.420	0.490
R-squared (uncentered) adj.	0.460	0.477	0.505	0.491	0.553
RMSE	73.59	72.25	70.68	71.03	66.99
RESET stat.	3,879.3	2,328.7	3,497.9	1,176.1	89.7
RESET prob.	0.00	0.00	0.00	0.00	0.00
Chow stat.			245.8		5.3
Chow prob.			0.00		0.00
Hausman stat.			752.2		1,065.9
Hausman prob.			0.00		0.00

significance: * 5% ** 1% *** 0.1%.

Source: author's calculations from the Uwezo 2 data.

Notes: dependent variable is the standardized combined test score; selected coefficients omitted; estimator is indicated by the column headings; standard errors (given in parentheses) are robust and adjust for clustering at the PSU-level; specification / parameter heterogeneity tests are as described in the text, the null hypothesis being that slope parameters are the same across the groups (as indicated); Hausman statistics compare the FE and WMG estimates, based on the same grouping variables.

at the bottom of the table. They include Ramsey's (1969) general specification RESET test, a Chow test for parameter equality across the separate regressions deployed to form the WMG estimates, and a Hausman-type test of the difference between the slope parameter of a fixed effects model versus the WMG estimator (e.g., Column IV vs. Column V). All of these tests are statistically significant and point to a rejection of the null of parameter homogeneity. Thus, only the WMG estimator is consistent for region-wide average effects.

Notwithstanding the above, there remain similarities across the models. This is particularly true with respect to their overall explanatory power, as well as the statistical significance and broad magnitude of the majority of individual parameters. First, the goodness-of-fit of even the most restrictive model is over 40% (see the adjusted R-squared statistic for column I). Compared to estimates in the literature, which often incorporate many more explanatory variables, these constitute strong results. In developed countries, R-squared statistics for international and national education production functions often range from around 10% to 30% (e.g., Wößmann, 2003; Ammermüller et al., 2005; Fuchs and Wößmann, 2007). In developing countries, estimates tend to hover around 30% (e.g., Fehrler et al., 2009; Aturupane et al., 2011; Glick et al., 2011). Additionally, the majority of explanatory power is derived from the household- and child-level covariates. This can be seen by adding the variables in sequence, estimating the model at each step. An upper bound estimate of the explanatory power of school and contextual factors is found when they enter the model first. As summarised in Table 3, an OLS estimate (as per column I of Table 2) yields an R-squared statistic of just 3% when the contextual variables are entered, increasing to 7% when the schooling variables are added, and 15% when factors that only vary at the household level are added.[16] Further adding child-level variables, which include parental education indicators, constitutes the full model. The conclusion is that differences in background factors explain more than a third of variation in acquired cognitive skills across the region.

A related result is that the country-specific average test score rankings, indicated in the descriptive analysis (Section 2), are preserved in the multivariate context without the need for country-fixed effects.[17] The implication is that variation in observed factors accounts for much of the unconditional average differences in test scores between countries. A closer look suggests that

[16]The latter set of covariates correspond to variables summarised in panel (c) of Table 1. Note that they do not include parental characteristics as these vary within the household in some instances. Somewhat lower R-squared results are found when alternative estimators are employed which account for country- or regional- fixed effects (not shown).

[17]Use of country-specific intercepts (via fixed effects) would assure this result automatically.

Table 3: Summary of sequential-build of regional education production function

	Specification	Fitted test score (mean) Kenya	Tanzania	Uganda	R-sq.
1.	L	2.16	2.66	-5.06	0.03
2.	$L + S$	9.00	-4.85	-2.43	0.07
3.	$L + S + H$	24.50	-7.73	-13.03	0.15
4.	Full	24.25	-2.55	-18.85	0.46
Unconditional mean		31.99	3.34	-32.71	-

Source: author's calculations from the Uwezo 2 data.

Notes: the table summarises the country-average fitted values and the (adjusted, centered) R-squared statistic corresponding to OLS regression results including terms in sequential fashion; specifications refer to variable sets indicated in equation (1) – i.e., L refers to location/context effects, S school effects, H household effects; the full model corresponds to Table 2, Column (I); country-specific intercepts not included.

these differences are also driven by household factors. This can be seen from Table 3, which reports the average predicted test score values for each country, based on different models. The results show there is almost no expected test score differential between countries based solely on the contextual factors (row 1); and a small difference based on school and contextual factors combined (row 2).[18] Once household factors, such as her socio-economic status and size, are included (row 3) the average test score differential between countries more closely replicates those of the unconditional means (in the final row).

Turning to the individual parameters, the estimates for the household-level variables lie in a direction that is consistent with previous literature and are of an economically meaningful magnitude. Critically, there are significant and large gains to learning that accrue with age, holding fixed all other inputs. These reflect conditional average age-related improvements across the sample (hereafter described as 'predicted year increments'). The negative parameter estimate of the square of the child's age shows that younger children improve on the tests disproportionately faster than older children. This is to be expected given that the Uwezo tests are set at a low level. Importantly, these gains do not purely reflect school-based learning. Gains in cognitive abilities are expected to occur naturally with age among young healthy children (for discussion see Eccles, 1999). Also, general numeracy and literacy skills learned outside the school environment (e.g., from parents or siblings) would contribute to higher scores.

[18]Notably, although these are not unbiased estimates, the latter would suggest that Kenyan schools are somewhat 'better' and Tanzanian schools somewhat 'worse'.

Consequently, the ratio of other estimated coefficients to the combined age effect (calculated at the approximate sample mean of 11 years) provides an intuitive comparative measure of their economic contribution to learning. For example, from the preferred estimates of Table 2 Column (V), the positive average effect of being female is equal to around 5% of a standard deviation in test scores or a difference of about 0.3 predicted year increments.[19]

Viewed in this light, the sheer magnitude of the effects of background factors on predicted test scores is large. For instance, the marginal increase in a child's predicted test score if her mother and father have a secondary education is over 30% of a standard deviation or 2 predicted year increments. Equally, a child from an ultra-poor family is expected to perform around 20% of a standard deviation lower than if she were from a non-poor family (*ceteris paribus*), equivalent to around 1.25 predicted year increment. Thus, holding all other things the same, the expected learning gap for children of the same age (at any age) between children born in ultra-poor households to uneducated parents, versus those in non-poor households to highly educated parents, equals more than 3 predicted year increments. These findings give credence to concerns surrounding the intergenerational transmission of poverty and inequality (see Harper et al., 2003; Bird, 2007).

An interesting result is the positive estimated coefficients on the (dummy) variables which take a value of one if the biological mother or father of the child does not reside in the household. If we were to assume all these children are orphans, taken in by relatives, then this result would run counter to expectations and previous literature (Case et al., 2004). However, this may well be driven by other phenomena. Parental absenteeism is not uncommon – e.g., due to single-parent families or out-migration to seek work. Also, anecdotal evidence suggests that more able children are not infrequently fostered out to (better-off, urban) relatives in order to attend school. It is impossible to distinguish between these separate hypotheses (mechanisms) from the present data. The point is that a positive effect of biological parental absenteeism remains plausible.

Finally, parameter estimates capturing school inputs are generally positive and significant. This implies that (on average) better resourced-schools produce higher test scores. To the extent that this is not purely an artifact of upward bias from omitted variables (across the underlying regressions), this is consistent with findings from the sparse developing country literature that school inputs have a role in enhancing learning outcomes. At the same time, the magnitude of most of these effects is moderate, particularly from a cost-benefit perspective. For example,

[19]Recall the dependent variable is standardized and multiplied by 100; thus estimated coefficients represent percentages of a standard deviation.

taking the results in Column (V) of Table 2, adding one book per 100 pupils is associated with an expected test score that is higher by less than 1% of a predicted year increment. Similarly, a 20 percentage point increase in the teacher attendance rate is also associated with a test score gain of under 1% of a predicted year increment.

Nonetheless, two school inputs consistently appear most important – (i) class sizes, given by the number of teachers to pupils; and (ii) school size, measured as the natural logarithm of the number of enrolled pupils. Taking the same WMG estimates, adding just two more teachers per 100 children raises fitted average test scores by more than one predicted year increment. According to Hanushek and Wößmann (2011), based on international evdience, this effect tends to be found where teacher quality is low. With respect to school size, a doubling of the number of enrolled pupils is associated with around a one year predicted increment in test scores (per pupil). On the one hand, this result may be due to unobserved factors such as the degree of urbanization or teacher quality (e.g., if the best teachers are attracted to the larger, well-known schools). On the other hand, smaller schools face particular constraints including insufficient staff to undertake administrative/management duties or cover for teacher absences; also, by definition, teachers in smaller schools are subject to less monitoring (by other staff). Evidence from elsewhere in Africa further suggests that smaller (more remote) schools typically receive the least experienced head teachers (De Grauwe et al., 2005). Indeed, in the present data the observed presence of the head teacher appears to be important for learning. This therefore provides suggestive evidence that school management plays an important role in the learning process.

4.2 National production

In light of the evidence for parameter heterogeneity, it is informative to look at the reduced-form educational production functions for each country separately. Appendix Table B1 summarises a series of country-specific regressions. Following the comparative approach of the previous section, OLS, FE and WMG results are presented (using districts as the grouping variable). Similar to the region-wide results, an immediate finding is that differences across the models are material for each country individually. Broadly speaking, coefficients on the household- and child-level covariates tend to be somewhat smaller, whilst school resource variables tend to increase in size under less restrictive estimators. The existence of underlying parameter heterogeneity within each of the countries is further supported by the homogeneity test statistics, which remains statistically significant across the table. Again, the WMG estimates should be seen as providing the more meaningful and reliable description of the characteristics of educational production in each country.

These points aside, there remains considerable similarity in the broad direction and magnitude of the parameter estimates across the countries, confirming the validity of the previous analysis. This can be seen from the empirical distributions of the district-level parameter estimates for each country – i.e., the set of parameter estimates taken from the separate regressions deployed in the district-level WMG estimates (columns III, VI and IX). A selection of these are summarised in Table 4, which reports parameter estimates at the 15th, 50th and 85th percentiles, also shown graphically by box-and-whisker plots in Figures A3 and A4. They reveal substantial overlap in the empirical parameter distributions for virtually all variables. For instance, with respect to the child's gender, all three countries indicate a moderate positive average parameter effect. However, the estimated variance of individual parameters is frequently quite different. In Uganda 35% of all children reside in districts where the test score advantage of being female is less than or equal to zero; this compares to 16% in Tanzania and under 10% in Kenya.

Despite these broad similarities, measures of distributional central location are significantly different across the three countries. A few examples suffice. The positive role of parental education is much larger in magnitude in Tanzania than in the other two countries. In both Kenya and Uganda, children whose parents have at most a primary education show no test score advantages compared to children whose parents have no education. In Tanzania, this difference equals around 0.5 predicted year increments. Parents with secondary education tend to confer advantages to their children more systematically, but the magnitude of this effect is twice as large in Tanzania than in either Uganda or Kenya. This can be interpreted in different ways. However, one reading is that equality of educational opportunity is lower in Tanzania; as such, education is unlikely to operate as an effective social leveller. At the same time, learning remains clearly skewed against the socio-economically disadvantaged in all countries. There are also complex and differentiated patterns in the (average) effects of school resources. Amongst these, the positive effect of adding more teachers is significantly larger in Uganda, suggesting that teacher quality may be a particular problem.

Finally, predicted gains associated with age vary between countries in important ways. The estimated partial derivative of the test score with respect to age is a linear function of age and is depicted in Figure A5 (based on the country-specific aggregate WMG estimates). The most notable feature of this is the much flatter slope of the Uganda age curve. Predicted

Table 4: Summary of district-level parameter estimates

	Kenya			Tanzania			Uganda		
Percentile →	15th	50th	85th	15th	50th	85th	15th	50th	85th
Child's age	10.10	13.99	17.42	16.44	18.34	21.03	15.15	17.08	19.00
Child is female	1.24	6.22	11.49	-0.40	5.47	12.30	-5.43	4.21	14.06
Child's birthorder	-11.29	-6.15	-3.02	-8.83	-2.92	2.60	-6.14	-3.65	0.17
Number of children in household	1.07	4.19	7.24	-2.30	4.15	8.53	0.56	3.02	6.20
Household is poor	-16.40	-8.95	-2.68	-19.26	-11.04	-2.45	-16.28	-11.37	-2.84
Household is ultra-poor	-39.16	-16.81	0.71	-39.57	-17.33	-2.64	-35.32	-21.02	1.10
Mother has primary education	-11.89	1.87	13.33	-0.98	10.40	20.80	-9.95	-1.36	8.83
Father has primary education	-11.79	-1.01	10.49	-6.96	8.08	25.10	-9.93	0.76	13.71
Mother has secondary education	-2.61	9.31	23.36	0.18	26.59	49.70	-4.33	11.43	24.52
Father has secondary education	-4.82	6.45	20.86	4.15	25.33	42.88	0.60	12.96	29.16
Teachers per 100 pupils (reported)	0.39	6.92	16.26	0.02	6.17	15.80	-0.12	12.09	26.74
Teacher attendance rate	-0.25	0.14	0.56	-0.13	0.29	0.69	-0.14	0.02	0.24
No. of books for every 100 pupils	-0.29	0.00	0.41	-0.17	0.12	0.43	-0.16	0.06	0.39

Source: author's calculations from the Uwezo 2 data.

Notes: the table summarises the empirical parameter distributions from country-specific WMG estimates of education production functions, using districts as the grouping variable – i.e., districts are the unit over which separate regressions are run; the results correspond to the estimates summarised in columns III, VI and IX of Table B1; district-level sample weights are applied.

gains of Ugandan children in the earliest years of education, holding all other inputs fixed, are around 10% of a standard deviation lower than in the other countries. One explanation for this, aside from somewhat more disadvantaged conditions (see Section 2), may be the more limited coverage of preschool services in Uganda. Statistics in UNESCO (2011) indicate that, in 2009, net enrolment in pre-primary schools is a little under a third of all children in Tanzania and Kenya, versus just 9% in Uganda. To the extent that pre-school education enables children to progress more quickly once they enter school (assuming they do so), the vast majority of Ugandan children lack this school-readiness. In turn, their expected rate of progress would be initially slow, and basic skills would be learnt later on, thereby explaining the flatter nature of the age-curve. Differences between Kenya and Tanzania are less material, but the Tanzanian slope is marginally flatter and located above that of Kenya. This suggests that Kenyan children learn basic cognitive skills somewhat sooner. A corollary implication is that a great deal of what matters for early-school learning is conditioned by experiences during the pre-school period. This encompasses a wide range of factors, including skills acquired from parents and siblings, health, nutrition and experience of a school-like environment. This view is supported by a large existing literature for developed and developing countries (for the African context see Garcia et al., 2008).

4.3 Sub-group differences

The previous subsections found that parameter heterogeneity is significant both between individual countries (on aggregate) and between districts within countries. Continuing to take the country as the appropriate overall unit of analysis, the last step is to investigate how education production varies for particular sub-groups. Motivated by previous findings, the groups are defined by age (i.e., younger and older age cohorts), socio-economic categories, and an aggregate index of school quality. The latter is identified by the fitted score from the first principal component of the schooling inputs used in the regressions, calculated on a country-specific basis. Schools which lie above the median can be thought of as having above-average general conditions – denoted 'higher quality'; schools with below-median conditions are denoted 'poorer quality' . Results of these estimates are reported in Tables 5 to 7, respectively corresponding to Kenya, Tanzania and Uganda. As interest here is on the sub-group specific estimates, spatial heterogeneity can continue to be accommodated by via the WMG estimator. However, due to sample size limitations, the geographic grouping variable is provinces within each country. Column (I) of each table reports the estimates for the full country sample, employing the WMG estimator but now using provinces as the dimension along which parameter slopes and intercepts vary. The remaining columns of each table report the same model for specific sub-samples. Note that for the socio-economic groups ('poor', 'ultra-poor') the respective dummy variables are dropped from the model as they are redundant.

With respect to differences between age cohorts, what matters for learning differs materially between children of different ages. Gender and birth order effects are significantly larger in both absolute and relative terms for younger children. Similarly, disadvantages due to household socio-economic status are larger in younger cohorts. Notably, in the case of Kenya, poor children in the older cohort are expected to perform only slightly below their non-poor counterparts on the Uwezo tests *ceteris paribus*. The effect of school inputs appears reasonably stable across age cohorts. However, an exception is Uganda where school inputs appear to be relatively *more* important for older children. This may well reflect the fact that many younger Ugandan children are insufficiently prepared to learn if/when they enter school (see above), meaning the marginal effects of specific school inputs are small. Nonetheless, in all countries, teacher attendance rates appear to exert a slightly larger positive effect on expected test scores among older children.

Turning to the other subgroups, it suffices to draw attention to a few major points. First, there are important differences in marginal effects between higher and poorer quality schools. A principal finding is that in Kenya and Tanzania, both the teacher-ratio and overall school size matter more for learning in worse-off schools, in the sense of being larger in magnitude. For instance, in higher quality Kenyan schools, addition of an extra teacher per hundred pupils would increase predicted test scores by 0.30 predicted year increments, while in a bad school the effect is more than twice as large at 0.63 predicted year increments. Thus, where school inputs are scarcer, their individual marginal effects tend to be larger. Similarly, doubling the size of worse-off schools (a log increase of one), is associated with a one predicted year increment in test scores.[20] Second, the gender advantage associated with girls is relatively smaller in poorer quality schools and among the ultra-poor households in all countries. Indeed, Ugandan girls in ultra-poor households are expected to perform significantly worse than boys (in Kenya and Tanzania there is no gender effect either way). Notably, these effects cannot be attributed to gender differences in enrolment propensities, which suggests there may be important intra-household resource allocation constraints, such as child labour obligations.

Third, the effect of parental education varies across sub-groups and in different ways across the three countries. Broadly, in Tanzania and Kenya, the beneficial effect of having parents with at

[20]See Section 4.1 for discussion of why this may be the case.

Table 5: Reduced-form educational production functions for Kenya

	(I) Baseline		(II) Ages 6-11		(III) Ages 12-16		(IV) Poor		(V) Ultrapoor		(VI) HQ school		(VII) PQ school	
	β	se	β	se	β	se	β	se	β	se	β	se	β	se
Child's age	32.36	(0.4)	28.66	(0.9)	21.03	(2.8)	31.39	(0.6)	22.56	(1.0)	32.99	(0.5)	30.87	(0.5)
Child's age squared	-1.85	(0.0)	-1.11	(0.2)	-1.13	(0.2)	-1.58	(0.0)	-0.92	(0.1)	-1.97	(0.0)	-1.63	(0.0)
Child is female	5.95	(0.5)	8.13	(0.8)	3.19	(0.6)	6.45	(0.8)	-1.76	(1.3)	7.27	(0.8)	4.39	(0.6)
Child's birthorder	-6.89	(0.5)	-8.49	(0.8)	-4.40	(0.8)	-6.03	(0.8)	-4.07	(1.2)	-7.60	(0.7)	-5.81	(0.6)
Number of children in household	4.48	(0.3)	5.31	(0.7)	4.32	(0.3)	4.48	(0.6)	4.62	(1.0)	4.89	(0.5)	3.69	(0.5)
Mother has primary education	3.90	(1.2)	1.89	(1.5)	6.88	(1.3)	5.68	(1.6)	8.95	(5.5)	1.61	(1.4)	6.81	(1.8)
Father has primary education	0.24	(1.2)	-1.22	(1.6)	2.81	(1.4)	1.17	(1.8)	7.76	(3.4)	-1.21	(1.5)	1.98	(2.0)
Mother has secondary education	14.19	(1.3)	14.88	(1.8)	10.81	(1.5)	17.78	(2.0)	16.96	(22.2)	12.71	(1.6)	19.13	(2.4)
Father has secondary education	8.71	(1.2)	10.91	(1.7)	5.00	(1.4)	10.84	(1.9)	5.84	(6.3)	7.17	(1.6)	9.57	(1.8)
Household is poor	-10.75	(0.8)	-15.29	(1.2)	-3.85	(0.9)					-10.14	(1.0)	-12.28	(1.3)
Household is ultra-poor	-27.89	(2.2)	-32.48	(2.3)	-18.08	(3.2)					-27.71	(3.6)	-24.52	(2.3)
School size (log.)	10.51	(1.0)	10.64	(1.4)	10.30	(1.2)	7.78	(1.7)	27.51	(3.1)	7.27	(1.2)	13.73	(1.7)
Teachers per 100 pupils (reported)	5.25	(0.6)	5.50	(0.9)	4.56	(0.5)	5.08	(0.9)	11.95	(1.4)	3.57	(0.8)	8.72	(0.7)
Teacher attendance rate	0.09	(0.1)	0.04	(0.1)	0.12	(0.0)	0.26	(0.1)	0.49	(0.1)	-0.00	(0.1)	0.26	(0.0)
No. of books for every 100 pupils	0.08	(0.0)	0.09	(0.0)	0.09	(0.0)	0.38	(0.2)	0.21	(0.1)	0.10	(0.0)	0.08	(0.1)
No school-level data available	-13.63	(3.9)	-6.83	(4.4)	-22.94	(5.1)	-20.47	(5.8)	-24.89	(6.7)	1.50	(4.1)	-22.85	(5.6)
Obs.	129,665		76,211		53,454		53,781		18,168		60,249		69,416	
Estimated parameters	264		264		264		248		217		264		231	
R-squared (centered) adj.	0.53		0.47		0.33		0.54		0.55		0.54		0.54	
Hausman stat.	1,783.4		432.6		40,680.4		5,814.9		22.0		123.2		8.4	
Hausman prob.	0.00		0.00		0.00		0.00		0.85		0.00		1.00	

Source: author's calculations from the Uwezo 2 data.

Notes: dependent variable is the standardized combined test score; column headings indicate sub-samples used; 'HQ' indicates higher quality, 'PQ' poorer quality; column (I) is the baseline model estimated on the full Kenyan sample; selected coefficients shown; all models estimated by the WMG estimator, employing provinces as the grouping variable; robust standard errors are given in parentheses, clustered at the PSU-level.

Table 6: Reduced-form educational production functions for Tanzania

	(I) Baseline		(II) Ages 6-11		(III) Ages 12-16		(IV) Poor		(V) Ultrapoor		(VI) HQ school		(VII) PQ school	
	β	se	β	se	β	se	β	se	β	se	β	se	β	se
Child's age	34.37	(0.5)	28.58	(1.3)	37.09	(3.7)	30.88	(0.7)	21.77	(1.5)	36.55	(0.7)	31.78	(0.7)
Child's age squared	-1.58	(0.0)	-0.62	(0.2)	-1.74	(0.2)	-1.17	(0.1)	-0.42	(0.1)	-1.82	(0.1)	-1.31	(0.1)
Child is female	6.26	(0.6)	7.35	(0.7)	4.79	(0.8)	6.26	(0.8)	2.23	(2.0)	7.69	(0.8)	4.83	(0.8)
Child's birthorder	-3.35	(0.5)	-5.58	(0.8)	-2.79	(0.8)	-3.51	(0.7)	-0.97	(1.9)	-3.79	(0.8)	-3.02	(0.7)
Number of children in household	3.52	(0.4)	5.34	(0.8)	3.28	(0.4)	4.16	(0.6)	2.16	(1.5)	3.66	(0.7)	3.29	(0.6)
Mother has primary education	11.81	(1.4)	12.18	(1.9)	11.93	(1.6)	12.02	(1.3)	19.61	(5.0)	13.08	(2.2)	10.22	(1.6)
Father has primary education	10.77	(1.5)	11.07	(1.7)	10.77	(2.0)	11.97	(1.6)	4.04	(2.6)	10.07	(2.2)	10.36	(1.8)
Mother has secondary education	25.26	(2.0)	31.81	(2.7)	18.66	(2.3)	28.94	(4.0)	1.68	(3.0)	26.07	(2.7)	24.35	(2.8)
Father has secondary education	27.09	(1.9)	34.35	(2.3)	18.92	(2.4)	28.90	(2.7)	43.40	(7.7)	24.29	(2.6)	29.27	(2.5)
Household is poor	-12.84	(0.8)	-15.75	(1.0)	-9.26	(1.0)					-13.91	(1.1)	-11.67	(1.1)
Household is ultra-poor	-23.50	(2.3)	-24.62	(2.4)	-22.49	(4.1)					-24.15	(4.2)	-20.90	(2.5)
School size (log.)	15.15	(1.4)	12.76	(1.5)	18.40	(1.6)	17.73	(1.7)	27.32	(3.6)	12.00	(1.7)	17.58	(2.0)
Teachers per 100 pupils (reported)	4.71	(0.4)	4.99	(0.4)	4.49	(0.4)	5.67	(0.5)	12.00	(1.6)	4.37	(0.4)	7.55	(0.7)
Teacher attendance rate	0.16	(0.0)	0.13	(0.0)	0.18	(0.0)	0.17	(0.0)	0.27	(0.1)	0.23	(0.0)	0.12	(0.0)
No. of books for every 100 pupils	0.10	(0.0)	0.11	(0.0)	0.10	(0.0)	0.14	(0.0)	0.22	(0.1)	0.09	(0.0)	0.26	(0.0)
No school-level data available	2.87	(2.9)	9.05	(3.8)	-4.40	(3.3)	3.37	(3.9)	8.41	(6.1)	4.70	(2.8)	-1.96	(3.3)
Obs.	116,232		63,129		53,103		55,666		7,956		57,547		58,685	
Estimated parameters	693		693		693		651		620		693		693	
R-squared (centered) adj.	0.42		0.32		0.24		0.41		0.41		0.43		0.42	
Hausman stat.	1,678.0		315.1		1,096.7		1,040.9		176.9		154.8		103.5	
Hausman prob.	0.00		0.00		0.00		0.00		0.00		0.00		0.00	

Source: author's calculations from the Uwezo 2 data.

Notes: dependent variable is the standardized combined test score; column headings indicate sub-samples used; 'HQ' indicates higher quality, 'PQ' poorer quality; column (I) is the baseline model estimated on the full Tanzanian sample; selected coefficients shown; all models estimated by the WMG estimator, employing provinces as the grouping variable; robust standard errors are given in parentheses, clustered at the PSU-level.

Table 7: Reduced-form educational production functions for Uganda

	(I) Baseline		(II) Ages 6-11		(III) Ages 12-16		(IV) Poor		(V) Ultrapoor		(VI) HQ school		(VII) PQ school	
	β	se	β	se	β	se	β	se	β	se	β	se	β	se
Child's age	22.04	(0.4)	10.81	(0.8)	59.97	(4.1)	17.52	(0.5)	12.13	(1.0)	26.22	(0.6)	17.12	(0.5)
Child's age squared	-0.53	(0.0)	1.53	(0.1)	-3.06	(0.3)	-0.05	(0.0)	0.07	(0.1)	-0.93	(0.1)	-0.05	(0.1)
Child is female	3.18	(0.7)	3.42	(0.8)	3.21	(0.9)	3.44	(0.8)	-3.97	(1.8)	5.33	(1.0)	0.63	(0.7)
Child's birthorder	-3.79	(0.4)	-5.30	(0.6)	-3.60	(0.7)	-3.19	(0.5)	-5.02	(1.4)	-3.29	(0.6)	-4.24	(0.5)
Number of children in household	3.58	(0.4)	4.37	(0.6)	3.82	(0.4)	3.59	(0.5)	4.84	(1.1)	3.16	(0.5)	4.14	(0.5)
Mother has primary education	-1.53	(1.2)	-3.11	(1.3)	1.41	(1.7)	-0.82	(1.2)	2.65	(4.5)	-4.31	(1.8)	2.54	(1.3)
Father has primary education	0.96	(1.5)	-0.52	(1.4)	1.56	(1.9)	1.01	(1.3)	-2.40	(2.4)	-0.79	(2.0)	0.95	(1.4)
Mother has secondary education	12.15	(1.8)	15.99	(2.0)	6.50	(2.3)	7.82	(2.5)	-7.88	(9.4)	11.80	(2.4)	12.15	(2.4)
Father has secondary education	15.04	(1.7)	12.50	(1.8)	16.62	(2.2)	15.31	(1.8)	14.50	(4.6)	13.79	(2.4)	13.80	(1.8)
Household is poor	-10.76	(0.9)	-12.50	(1.0)	-8.47	(1.3)					-12.48	(1.4)	-8.64	(1.1)
Household is ultra-poor	-19.71	(2.1)	-17.97	(2.3)	-21.94	(3.2)					-25.98	(3.0)	-12.41	(2.2)
School size (log.)	14.79	(1.3)	12.65	(1.4)	20.55	(1.8)	12.19	(1.6)	16.01	(3.3)	17.52	(1.9)	14.42	(2.0)
Teachers per 100 pupils (reported)	10.12	(1.1)	9.33	(1.1)	13.25	(1.3)	10.09	(1.2)	8.71	(2.8)	10.13	(1.2)	9.71	(2.0)
Teacher attendance rate	0.02	(0.0)	-0.01	(0.0)	0.07	(0.0)	0.03	(0.0)	0.13	(0.0)	-0.01	(0.0)	0.05	(0.0)
No. of books for every 100 pupils	0.02	(0.0)	-0.02	(0.0)	0.09	(0.0)	0.04	(0.0)	-0.04	(0.1)	-0.03	(0.0)	0.07	(0.0)
No school-level data available	-2.26	(2.5)	1.64	(2.6)	-9.06	(3.3)	-2.15	(3.2)	-9.42	(5.2)	-0.53	(3.1)	-10.14	(4.4)
Obs.	102,270		59,048		43,222		50,481		10,066		48,825		53,445	
Estimated parameters	132		132		132		124		124		132		132	
R-squared (centered) adj.	0.47		0.37		0.28		0.46		0.42		0.48		0.46	
Hausman stat.	57.7		22.5		1,092.2		6,395.8		379.5		29.1		284.9	
Hausman prob.	0.00		0.89		0.00		0.00		0.00		0.61		0.00	

Source: author's calculations from the Uwezo 2 data.

Notes: dependent variable is the standardized combined test score; column headings indicate sub-samples used; 'HQ' indicates higher quality, 'PQ' poorer quality; column (I) is the baseline model estimated on the full Ugandan sample; selected coefficients shown; all models estimated by the WMG estimator, employing provinces as the grouping variable; robust standard errors are given in parentheses, clustered at the PSU-level.

least a primary education is larger in more disadvantaged settings (amongst the ultra-poor and in poorer quality schools); also, in these countries the mother's education is most important relative to the father's. Uganda again presents a distinct picture – whether or not the father has secondary education appears most critical, even in disadvantaged settings. However, this should not be over-interpreted. These results could be an artifact of the crude measure of socio-economic status employed – i.e., parental education could be correlated with omitted income or wealth covariates.

Finally, in a number of instances the estimate on the dummy variable used to identify those PSUs where no school data was collected (primarily due to the absence of any proximate government school is large, negative and significant). Specifically, this result holds for the ultra-poor, the older age cohort and children associated with poorer quality schools in Uganda and Kenya (but not Tanzania). A plausible interpretation is that this is capturing some of the real costs associated with the distance to the nearest primary school. This result does not appear in previous tables, implying it is likely to be correlated with district-level (fixed) effects such as population and school density. In other words, taking into account the opportunity costs of time, test scores suffer where access is more constrained.

5 Conclusions

The objective of this study was to characterise educational production functions in East Africa based on a reduced form framework. This responds to the question: 'what matters for learning' in the region? The analysis was based on a unique, large-scale micro-dataset collected by the Uwezo initiative. Encompassing around 350,000 school-aged children across Kenya, Tanzania and Uganda, the data includes literacy and numeracy test scores, household information and local conditions, including school inputs. Importantly, the data was collected at the household- rather than school-level, thereby avoiding bias from sample selection effects. However, as the surveys only provide retrospective data there is no guarantee that bias from omitted variables or measurement error has been entirely excluded. For these reasons, the estimates should not be given a strong causal interpretation.

The analysis proceeded in three steps. First, the regional reduced-form educational production function was estimated. Second, the underlying national production functions were considered. This permitted a comparison of what matters for learning between each country on aggregate, as well as of the empirical distribution of parameter estimates for administrative districts within each country. Finally, the richness of the data permitted an analysis of differences between sub-groups within each country – namely, younger and older cohorts, disadvantaged socio-economic groups, and children matched to higher versus poorer quality schools.

Throughout the analysis, a concern for parameter heterogeneity was explicit. This is motivated by findings from other areas of economics, which challenge the existence of homogeneous production functions, as well as concerns that the assumption of linear functional forms is overly restrictive. The East African region also is highly diverse (economically, socially and culturally), comprising areas with very high primary school enrolment (and average test scores) and areas where large shares of school-aged children have no primary schooling, despite access fees having been abolished. *A priori*, therefore, there is no strong reason to expect that the same inputs to education matter in the same way for all children. This position is substantiated by an appreciation that empirical estimates of education production functions almost invariably ignore important aspects of behaviour. This can come about through omitted variables, as well as exclusion of relevant non-linearities or interaction terms that would permit the marginal effects of inputs to vary across the empirical distribution of covariates. Following the literature, these omissions are likely to give rise to neglected parameter heterogeneity.

Addressing parameter heterogeneity is not a technical sideshow. Section 3.2 outlined conditions under which standard least squares estimators, possibly including group-specific fixed effects, are inconsistent for any meaningful summary parameters such as average marginal effects. In response, a simple framework for addressing this challenge was developed – specifically, a sample-weighted mean-group is likely to be consistent. The framework also provides formal tests for parameter heterogeneity, based on the most flexible version of the Chow test where all slope coefficients are allowed to vary between groups, as well as a global test of the difference between a pooled and weighted-mean group summary estimator.

The findings of the analysis confirm the suspicion that parameter heterogeneity is material, both within and between countries. This is particularly so with respect to school inputs, where use of OLS or fixed effects estimators leads to a substantial downward bias relative to weighted-mean group estimates. The interpretation is that parameter heterogeneity is most significant for these variables. Nonetheless, broad regularities are apparent. Family background and child-level variables are crucial for learning, explaining around one third of the variation in test scores alone or nearly two thirds (30/50) of all variation that can be attributed to the explanatory variables. Children in ultra-poor households whose parents have no education can be expected to achieve learning outcomes that are lower by up to four years of predicted age-related progression (versus a child from a non-poor household with parents educated to secondary level). This means that educational (dis)advantages are strong and persistent across generations. In contrast, school inputs have a relatively moderate effect in boosting learning on average. That said, smaller class sizes are associated with reasonably large positive marginal effects on learning. Also, schools with more pupils are associated with better performance, likely due to the specific challenges associated with small (remote) schools such as dealing with teacher absence.

With respect to differences between countries, a stark finding is that children are expected to improve on the test scores at different speeds with age *ceteris paribus*. Specifically, the age-curve of children in Uganda is much flatter than that of Kenya and Tanzania, implying the average Ugandan child is expected to acquire basic cognitive skills more slowly and, thus, later on. This may be due to the lower prevalence of pre-schooling in Uganda. However, it could also reflect unobserved factors, such as aspects of school quality that are correlated with age. The results also point to the larger magnitude of parental education effects in Tanzania, especially the effect of parents having only primary education. In turn, this suggests that Tanzanian schools provide less equal opportunities.

Sub-group differences were found to be material. In particular, there is evidence that school inputs have larger positive effects on test scores for children from disadvantaged families and with access to poorer quality schools. At the same time, the gender advantage associated with female children tends to disappear amongst disadvantaged families as well as in poor quality schools. In fact, Ugandan girls from ultra-poor households perform significantly worse on average than their male counterparts. There is also suggestive evidence that distance to the nearest primary school (i.e., due to living in a remote area) is associated with lower learning outcomes.

There are two main implications of the analysis. The first is that the assumption of constant linear marginal effects due to individual explanatory variables is restrictive and potentially misleading at the aggregate level. Marginal effects appear to vary systematically along the empirical distribution of covariates, captured here in terms of parameter heterogeneity within and between countries at the geographic and sub-group levels. In turn, this would suggest that the apparent lack of uniformity of results about the effectiveness of school inputs in previous studies is likely to reflect the underlying complexity and essential variation in how education is produced. Second, the reported diversity in what matters for learning implies that generalised educational interventions may be quite ineffective in specific contexts. Policies to promote learning should take into account local nuances and information. This underlines the need for decentralised decision-making and empowerment of local actors, particularly parents (e.g., see Duflo et al., 2012).

References

Alkire, S. and Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95(7):476–487.

Ammermüller, A., Heijke, H. and Wößmann, L. (2005). Schooling quality in Eastern Europe: Educational production during transition. *Economics of Education Review*, 24(5):579–599.

Angrist, J. (1998). Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants. *Econometrica*, 66(2):249–288.

Ashenfelter, O. and Krueger, A. (1994). Estimates of the Economic Return to Schooling from a New Sample of Twins. *The American Economic Review*, 84(5):1157–1173.

Aturupane, H., Glewwe, P. and Wisniewski, S. (2011). The impact of school quality, socioeconomic factors, and child health on students' academic performance: evidence from Sri Lankan primary schools. *Education Economics*, iFirst Article:1–37.

Bird, K. (2007). The intergenerational transmission of poverty: An overview. Chronic Poverty Research Centre Working Paper 99, Overseas Development Institute.

Bird, K., Higgins, K. and McKay, A. (2010). Conflict, education and the intergenerational transmission of poverty in Northern Uganda. *Journal of International Development*, 22(8):1183–1196.

Bound, J., Brown, C. and Mathiowetz, N. (2001). Measurement error in survey data. *Handbook of Econometrics*, 5:3705–3843.

Bound, J., Jaeger, D. and Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.

Breusch, T. and Pagan, A. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5):1287–1294.

Case, A., Paxson, C. and Ableidinger, J. (2004). Orphans in Africa: parental death, poverty, and school enrollment. *Demography*, 41(3):483–508.

Chow, G. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28(3):591–605.

De Grauwe, A., Lugaz, C., Baldé, D., Diakhaté, C., Dougnon, D., Moustapha, M. and Odushina, D. (2005). Does decentralization lead to school improvement? Findings and lessons from research in West-Africa. *Journal of Education for International Development*, 1(1):1–15.

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy*. Baltimore: Johns Hopkins University Press.

Duflo, E., Dupas, P. and Kremer, M. (2012). School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools. Working Paper 17939, National Bureau of Economic Research. URL www.nber.org/papers/w17939.

Eccles, J. (1999). The development of children ages 6 to 14. *The Future of Children*, 9(2):30–44.

Fehrler, S., Michaelowa, K. and Wechtler, A. (2009). The Effectiveness of Inputs in Primary Education: Insights from Recent Student Surveys for sub-Saharan Africa. *The Journal of Development Studies*, 45(9):1545–1578.

Figlio, D. (1999). Functional form and the estimated effects of school resources. *Economics of Education Review*, 18(2):241–252.

Fuchs, T. and Wößmann, L. (2007). What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics*, 32(2):433–464.

Garcia, M., Pence, A. and Evans, J. (Eds.) (2008). *Africa's future, Africa's challenge: early childhood care and development in Sub-Saharan Africa*. World Bank Publications.

Gibbons, C., Serrato, J. and Urbancic, M. (2011). Broken or Fixed Effects? Unpublished manuscript, University of California, Berkeley.

Glewwe, P. (2002). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature*, 40(2):436–482.

Glewwe, P., Hanushek, E., Humpage, S. and Ravina, R. (2011). School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010. Working Paper 17554, National Bureau of Economic Research.

Glewwe, P. and Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, 2:945–1017.

Glewwe, P., Kremer, M., Moulin, S. and Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya. *Journal of Development Economics*, 74(1):251–268.

Glick, P., Randrianarisoa, J. and Sahn, D. (2011). Family background, school characteristics, and children's cognitive achievement in Madagascar. *Education Economics*, 19(4):363–396.

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

Hanushek, E. and Wößmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3):607–668.

Hanushek, E.A. and Wößmann, L. (2011). The Economics of International Differences in Educational Achievement. In E.A. Hanushek, S. Machin and L. Wößmann (Eds.), *Handbooks in Economics*, volume 3, pp. 89–200. The Netherlands: North-Holland.

Haque, N.U., Sharma, S. and Pesaran, M.H. (1999). Neglected Heterogeneity and Dynamics in Cross-Country Savings Regressions. IMF Working Papers 99/128, International Monetary Fund.

Harper, C., Marcus, R. and Moore, K. (2003). Enduring poverty and the conditions of childhood: lifecourse and intergenerational poverty transmissions. *World Development*, 31(3):535–554.

Jimenez, E. and Sawada, Y. (1999). Do community-managed schools work? An evaluation of El Salvador's EDUCO program. *The World Bank Economic Review*, 13(3):415–441.

Juhl, T. and Lugovskyy, O. (2010). A test for slope heterogeneity in fixed effects models. Unpublished manuscript, Department of Economics, University of Kansas.

Kravdal, O. (2002). Education and fertility in sub-Saharan Africa: individual and community effects. *Demography*, 39(2):233–250.

Lee, K., Pesaran, M.H. and Smith, R. (1997). Growth and Convergence in Multi-country Empirical Stochastic Solow Model. *Journal of Applied Econometrics*, 12(4):357–92.

Michaelowa, K. (2001). Primary education quality in francophone Sub-Saharan Africa: Determinants of learning achievement and efficiency considerations. *World Development*, 29(10):1699–1716.

Mugo, J. (2007). Education and mitigation of child labour in Kenya. *IDS Policy Brief*, 6(2).

Nishimura, M., Yamano, T. and Sasaoka, Y. (2008). Impacts of the universal primary education policy on educational attainment and private costs in rural Uganda. *International Journal of Educational Development*, 28(2):161–175.

Nubé, M. and Sonneveld, B. (2005). The geographical distribution of underweight children in Africa. *Bulletin of the World Health Organization*, 83(10):764–770.

Okuni, A. (2003). Quantity-quality trade-offs after UPE: prospects and challenges of universal access in Uganda. *Norrag News*, 32:32–37. URL www.norrag.org/issues/download/NN/32/en.

Pesaran, M.H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1):79–113.

Pesaran, H.M. and Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, 142(1):50–93.

Psacharopoulos, G. and Patrinos, H. (2004). Returns to investment in education: a further update. *Education economics*, 12(2):111–134.

Ramsey, J. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2):350–371.

Schultz, T. (1999). Health and schooling investments in Africa. *The Journal of Economic Perspectives*, 13(3):67–88.

Sianesi, B. and Reenen, J. (2003). The returns to education: macroeconomics. *Journal of Economic Surveys*, 17(2):157–200.

Tavakol, M. and Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2:53–55.

Temple, J. (2002). Growth regressions and what the textbooks don't tell you. *Bulletin of Economic research*, 52(3):181–205.

Todd, P. and Wolpin, K. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485):F3–F33.

UNESCO (2011). *Global Education Digest 2011: Comparing Education Statistics Across The World*. Quebec: UNESCO Institute for Statistics.

Uwezo (2012). Are our Children Learning? Literacy and Numeracy across East Africa. Policy report, Uwezo. URL www.uwezo.net/wp-content/uploads/2012/08/RO_2012_UwezoEastAfricaREport.pdf.

Wooldridge, J. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, 87(2):385–390.

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. *Oxford Bulletin of Economics and Statistics*, 65(2):117–170.

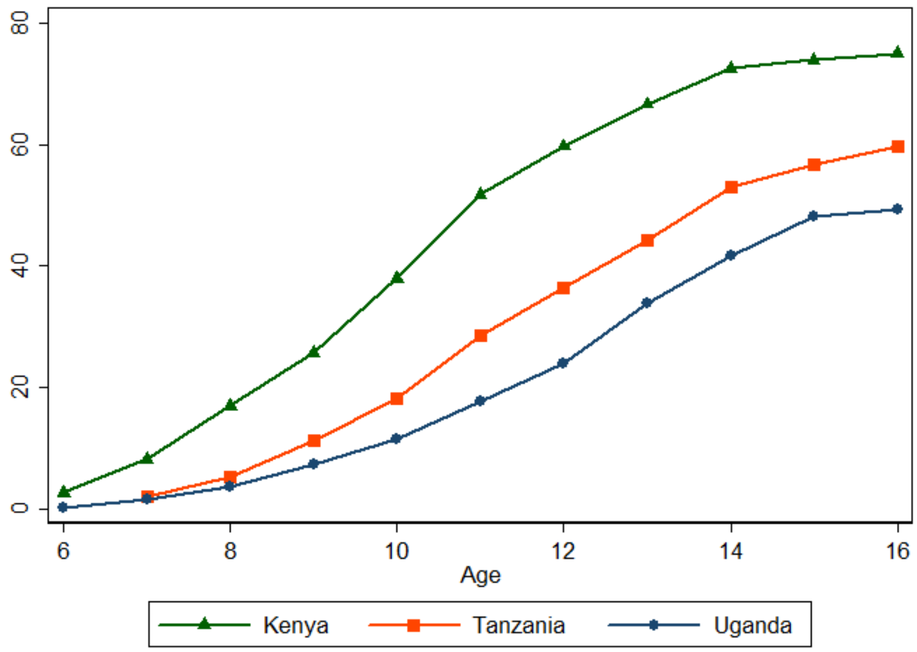
—— (2005). Educational production in East Asia: The impact of family background and schooling policies on student performance. *German Economic Review*, 6(3):331–353.

Zietz, J. (2001). Heteroskedasticity and neglected parameter heterogeneity. *Oxford Bulletin of Economics and Statistics*, 63(2):263–273.

—— (2006). Detecting neglected parameter heterogeneity with Chow tests. *Applied Economics Letters*, 13(6):369–374.

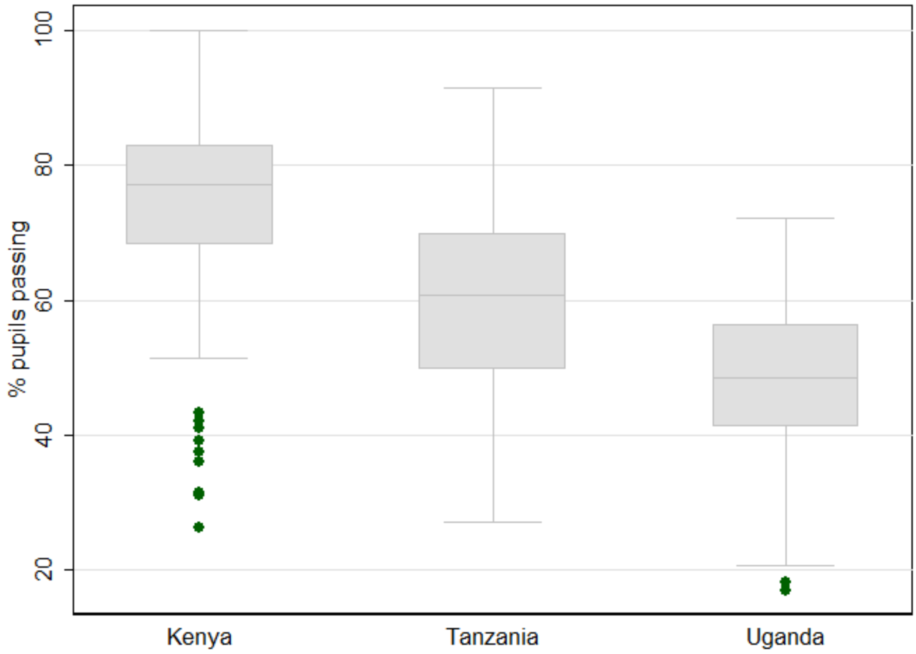
Appendix A: Additional Figures

Figure A1: Share of children achieving maximum mark (pass) on combined tests, by age and country



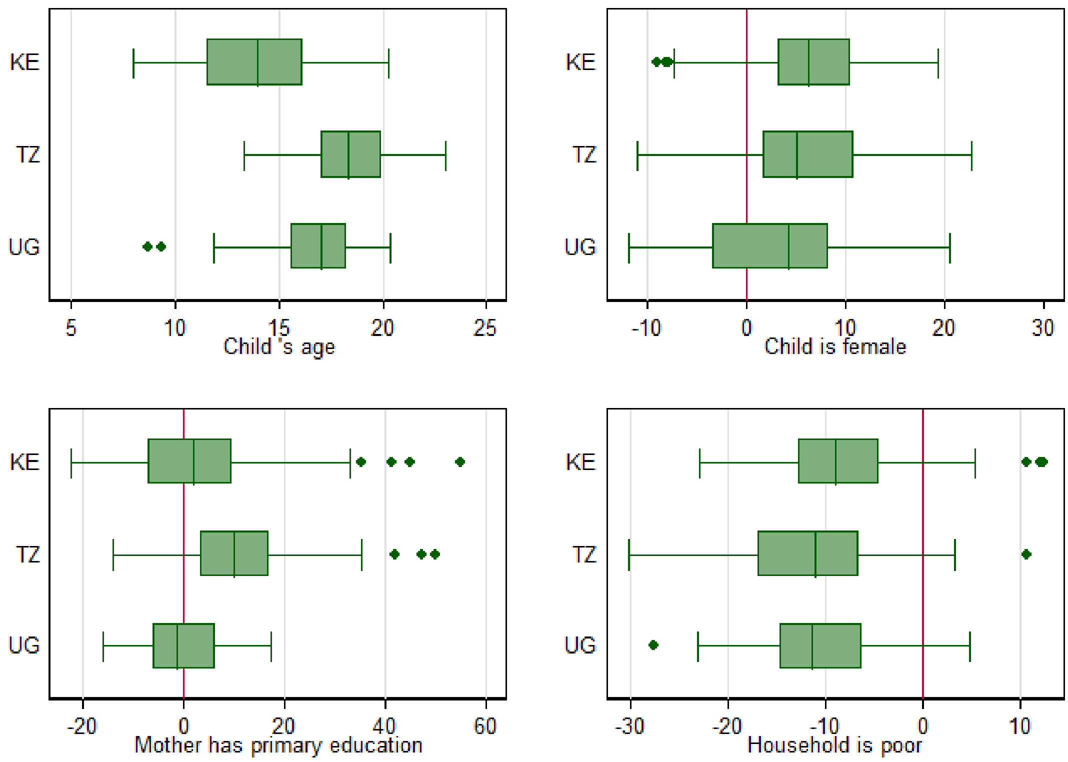
Source: authors' calculations from the Uwezo 2 data.

Figure A2: Box plot of district-level average pass rates on combined tests, by country



Source: authors' calculations from the Uwezo 2 data.
Notes: only children aged 10 and over included.

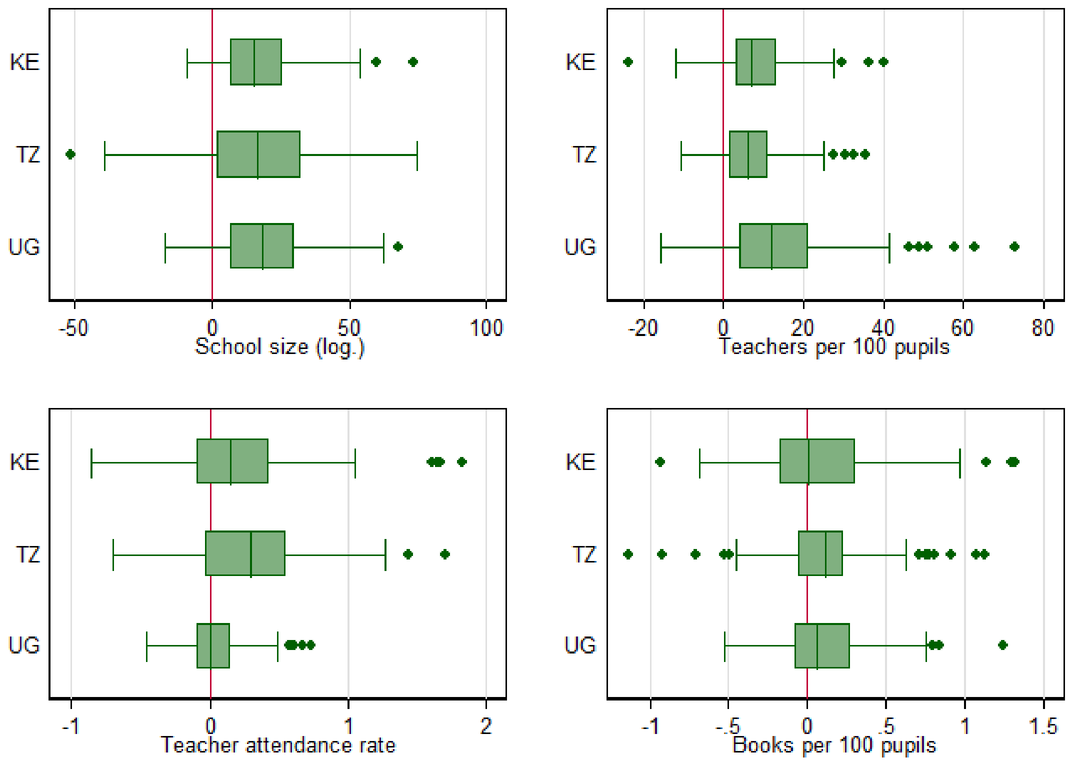
Figure A3: Summary of district-level parameter estimates,selected household variables



Source: authors' calculations from the Uwezo 2 data.

Notes: KE = Kenya, TZ = Tanzania, UG = Uganda; estimates based on district-level regressions corresponding to the WMG estimates of Tables 5-7.

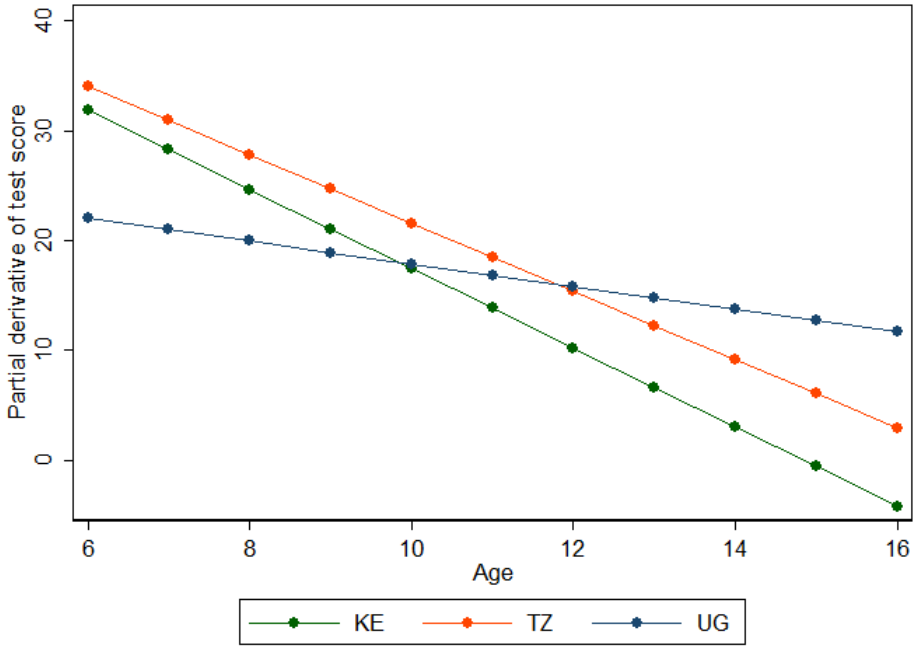
Figure A4: Summary of district-level parameter estimates,selected school inputs



Source: authors' calculations from the Uwezo 2 data.

Notes: KE = Kenya, TZ = Tanzania, UG = Uganda; estimates based on district-level regressions corresponding to the WMG estimates of Tables 5-7.

Figure A5: Marginal effect of age on predicted test score



Source: authors' calculations from the Uwezo 2 data.

Notes: estimates based on sample-weighted parameter averages of the WMG estimates as reported in Tables 5-7.

Appendix B: Additional Tables

Table B1: Estimates of reduced-form national educational production functions

Location →	Kenya			Tanzania			Uganda		
Estimator →	OLS (I)	FE (II)	WG (III)	OLS (IV)	FE (V)	WG (VI)	OLS (VII)	FE (VIII)	WG (IX)
Child's age	32.31***	32.07***	31.87***	35.11***	34.76***	34.03***	22.00***	21.99***	22.00***
	(0.3)	(0.3)	(0.3)	(0.5)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)
Child's age squared	-1.84***	-1.81***	-1.80***	-1.65***	-1.62***	-1.55***	-0.52***	-0.52***	-0.51***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Child's birthorder	-7.13***	-7.10***	-6.47***	-3.55***	-3.74***	-3.33***	-3.76***	-3.79***	-3.46***
	(0.4)	(0.4)	(0.5)	(0.5)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)
Child is female	5.53***	5.51***	6.26***	6.36***	6.36***	5.99***	3.14***	3.19***	3.48***
	(0.5)	(0.5)	(0.5)	(0.6)	(0.6)	(0.5)	(0.7)	(0.7)	(0.6)
Father is not in household	6.00*	7.42**	7.11**	15.53***	13.51***	8.74**	8.14**	8.66**	6.67*
	(2.6)	(2.4)	(2.2)	(2.9)	(2.8)	(2.7)	(2.7)	(2.7)	(3.3)
Mother is not in household	22.28***	20.59***	14.39***	25.08***	21.86***	15.29***	13.58***	12.71***	8.43**
	(3.5)	(3.5)	(2.5)	(3.6)	(3.6)	(3.2)	(3.3)	(3.2)	(2.8)
Age of mother	0.45***	0.42***	0.40***	0.25***	0.24***	0.17***	0.36***	0.35***	0.31***
	(0.1)	(0.1)	(0.0)	(0.1)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)
Mother has primary education	3.21**	2.36*	2.49*	12.35***	11.20***	11.41***	-2.49*	-0.90	-0.61
	(1.2)	(1.2)	(1.0)	(1.2)	(1.2)	(1.2)	(1.1)	(1.1)	(1.4)
Father has primary education	1.70	2.22*	-0.01	13.52***	12.16***	8.57***	0.78	1.53	2.52
	(1.2)	(1.1)	(1.0)	(1.4)	(1.3)	(1.3)	(1.4)	(1.4)	(1.7)
Mother has secondary education	14.46***	12.30***	10.73***	26.70***	23.56***	27.59***	12.70***	11.99***	10.79***
	(1.3)	(1.3)	(1.3)	(2.0)	(2.0)	(2.0)	(1.9)	(1.8)	(1.7)
Father has secondary education	10.68***	10.85***	7.15***	29.87***	26.72***	24.32***	15.19***	14.89***	14.77***
	(1.3)	(1.2)	(1.1)	(1.9)	(1.8)	(1.8)	(1.7)	(1.7)	(1.8)
No. children in household	4.25***	4.71***	4.10***	2.89***	3.78***	3.63***	3.17***	3.67***	3.35***

Table B1: Estimates of reduced-form national educational production functions

Location →	Kenya			Tanzania			Uganda		
Estimator →	OLS (I)	FE (II)	WG (III)	OLS (IV)	FE (V)	WG (VI)	OLS (VII)	FE (VIII)	WG (IX)
	(0.3)	(0.3)	(0.3)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)	(0.3)
No. other household members	-0.88***	-0.52**	-0.53***	-0.93***	-0.54***	-0.43**	-0.59***	-0.36**	-0.58**
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.2)	(0.1)	(0.2)
Household is poor	-12.56***	-10.58***	-8.89***	-16.17***	-13.78***	-11.23***	-12.56***	-10.50***	-10.91***
	(0.8)	(0.8)	(0.7)	(0.9)	(0.8)	(0.8)	(1.0)	(0.9)	(1.2)
Household is ultra-poor	-25.62***	-20.95***	-20.11***	-25.76***	-22.98***	-20.14***	-22.01***	-18.08***	-21.34***
	(1.9)	(1.7)	(1.7)	(1.9)	(1.8)	(2.3)	(1.8)	(1.7)	(2.2)
Household is aware of Uwezo	-0.76	-1.24	-1.26*	2.17	2.61	4.78***	-1.00	0.48	0.90
	(0.7)	(0.7)	(0.5)	(1.7)	(1.6)	(1.4)	(1.3)	(1.3)	(1.3)
School size (log.)	9.30***	15.02***	16.64***	6.74***	12.16***	17.28***	10.12***	13.91***	17.91***
	(1.2)	(1.2)	(1.1)	(1.3)	(1.4)	(1.5)	(1.3)	(1.5)	(1.6)
Teachers per 100 pupils (reported)	4.52***	3.96***	8.27***	0.49*	0.63**	7.59***	11.42***	8.54***	13.91***
	(1.0)	(0.7)	(0.7)	(0.2)	(0.2)	(0.5)	(1.1)	(1.0)	(1.5)
Teacher attendance rate	0.23***	0.17***	0.10	0.09***	0.09***	0.29***	0.02	0.02	0.04
	(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
No. of books for every 100 pupils	0.12***	0.09***	0.08**	0.11***	0.08***	0.10***	0.03	0.04	0.12**
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
School provides subsistence to pupils	-0.23	0.43	0.23	10.22***	9.03***	0.53	3.38**	1.89	5.30**
	(1.2)	(1.2)	(1.3)	(1.3)	(1.6)	(1.8)	(1.3)	(1.5)	(1.8)
School has access to clean water	7.53***	4.20***	2.23	10.82***	4.33**	-3.22	5.08***	3.86**	2.80*
	(1.2)	(1.1)	(1.2)	(1.6)	(1.7)	(2.0)	(1.3)	(1.2)	(1.4)
Head teacher present	6.17***	6.88***	6.46***	9.90***	10.89***	4.97***	9.85***	10.87***	10.45***
	(1.3)	(1.2)	(1.3)	(1.3)	(1.3)	(1.3)	(1.4)	(1.3)	(1.4)
No school-level data available	-11.09*	-7.24	-0.47	3.25	-3.96	1.34	-1.66	-0.17	0.28
	(5.0)	(4.7)	(1.5)	(4.5)	(4.6)	(0.8)	(2.8)	(2.6)	(2.7)
No. of primary schools in village	0.16	0.36	1.74***	2.41***	1.95***	1.07*	0.94	0.47	1.19*
	(0.6)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)	(0.6)	(0.5)	(0.5)

Table B1: Estimates of reduced-form national educational production functions

Location →	Kenya			Tanzania			Uganda		
Estimator →	OLS (I)	FE (II)	WG (III)	OLS (IV)	FE (V)	WG (VI)	OLS (VII)	FE (VIII)	WG (IX)
No. of secondary schools in village	2.52**	1.42	2.65**	1.17	0.72	0.78	3.18**	2.68**	1.47
	(0.9)	(0.8)	(0.9)	(0.9)	(0.9)	(0.9)	(1.2)	(1.0)	(1.0)
Village has access to electricity	5.16***	3.23**	3.35**	7.05***	4.57***	4.48***	6.41***	2.87*	2.41*
	(1.0)	(1.0)	(1.1)	(1.3)	(1.4)	(1.3)	(1.3)	(1.2)	(1.1)
Grouping variable	None	Districts	Districts	None	Districts	Districts	None	Districts	Districts
No. groups	0	122	122	0	119	119	0	79	79
Obs.	129,665	129,665	129,665	116,232	116,232	116,232	102,270	102,270	102,270
Estimated parameters	33	155	4,026	33	152	3,927	33	112	2,607
R-squared (centered) adj.	0.510	0.490	0.520	0.410	0.390	0.440	0.470	0.450	0.490
RESET stat.	785.6	237.9	80.1	814.9	479.6	52.0	1,897.4	1,078.3	161.4
RESET prob.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Chow stat.			3.4			3.9			4.3
Chow prob.			0.00			0.00			0.00
Hausman stat.			82.8			1,145.9			169.9
Hausman prob.			0.00			0.00			0.00

significance: * 5% ** 1% *** 0.1%.

Source: author's calculations from the Uwezo 2 data.

Notes: dependent variable is the standardized combined test score; estimator is indicated by the column headings; standard errors (given in parentheses) are robust and adjust for clustering at the PSU-level; specification / parameter heterogeneity tests are as described in the text, the null hypothesis being that slope parameters are the same across the groups (as indicated); Hausman statistics compare the FE and WMG estimates, based on the same grouping variables; selected coefficients omitted.