Preschool Attendance, School Progression, and Cognitive Skills in East Africa[☆]

Jan Bietenbeck[1], Sanna Ericsson[2], Fredrick M. Wamalwa[3]

July 31, 2018

Abstract

We study the effects of preschool attendance on children's school progression and cognitive skills in Kenya and Tanzania. Our analysis uses novel data from large-scale household surveys of children's literacy and numeracy skills, which also collect retrospective information on preschool attendance. Against the backdrop of a large expansion of pre-primary education, our regressions identify the impacts from within-household differences, controlling for a variety of child-specific covariates. In both countries, children who go to preschool tend to enroll in primary school late and thus fall behind in terms of grades completed at early ages. However, once in school, they progress through grades faster and eventually catch up. Importantly, they also show improved learning outcomes in the longer run: at ages 13-16, children who attended preschool score around 0.10 standard deviations higher on standardized tests in both countries and are 3 and 5 percentage points more likely to achieve basic literacy and numeracy in Kenya and Tanzania, respectively.

Keywords: preschool, education, cognitive skills, Sub-Saharan Africa
JEL codes: I21, J24

[☆]We thank Kaveh Majlesi, Alessandro Martinello, Emily Oster, Jorge Pérez Pérez, Andreas Stegmann, Alessandro Tarozzi, Petra Thiemann, and audiences at the University of Groningen, the 2017 Lund-SFI Workshop in Applied Microeconomics, the 2017 Nordic Conference in Development Economics, the autumn 2017 Copenhagen Education Network Workshop, the 2018 CSAE conference, the 2018 GlaD conference, and the 2018 ESPE conference for helpful comments. Corresponding author: Jan Bietenbeck (mail@janbietenbeck.com).

[1]Lund University Department of Economics and IZA.

[2]Lund University Department of Economics.

[3]School of Economics, Faculty of Commerce, University of Cape Town.

1. Introduction

School enrollment in Sub-Saharan Africa has increased substantially over the past two decades. However, many students fall behind the curriculum early on, and grade repetition and early dropout are widespread (UNESCO, 2012). Students also learn remarkably little in school: for example, only one in five third-grade students in East Africa has second-grade literacy and numeracy skills, and less than one third of sixth-grade students in Southern and Eastern Africa can solve a simple subtraction problem (Uwezo, 2015; Bietenbeck, Piopiunik, and Wiederhold, 2018).

One possible reason why students in these countries perform so poorly is that they enter school unprepared. Specifically, to the extent that early and later learning are complementary, a lack of education before starting school reduces children's efficiency in learning once they arrive there (Cunha and Heckman, 2007). High-quality preschool programs which prepare children for school are therefore often seen as a promising way to enhance learning outcomes (e.g. World Bank, 2018). In Sub-Saharan Africa, pre-primary education has been expanding rapidly, with the gross enrollment ratio doubling from 11% to 22% between 2000 and 2014.[1] Whether preschools are actually effective at boosting student outcomes in this region is unclear, however, because rigorous empirical evidence is still scarce.

In this paper, we study the effects of preschool attendance on children's school progression and cognitive skills in Kenya and Tanzania. Our empirical analysis draws on data from Uwezo, which conducts nationally representative household surveys of school-age children's education and their literacy and numeracy skills. The surveys also collect retrospective information on preschool attendance, which we can relate to current outcomes of respondents up to 16 years of age. The main part of our investigation focuses on impacts on the highest grade of school attended and a composite test score, which summarizes a child's performance on the standardized literacy and numeracy assessments. The data contain information on these outcomes for more than half a million children across the two countries, independently of whether they are currently enrolled in school or not.

[1] These figures come from the World Bank's World Development Indicators and can be accessed here: https://data.worldbank.org/indicator/SE.PRE.ENRR.

Our regression framework compares the outcomes of children who did and did not attend preschool. For identification, we rely on within-household differences, thereby controlling for all determinants of outcomes and attendance that vary across families.[2] We argue that the leftover variation between siblings is likely due to changes in the local availability of preschools, which came about because of an expansion of the pre-primary sector during our study period. In support of this claim, we show that even within households, children in later cohorts are much more likely to have attended preschool. To mitigate concerns about endogenous selection from the start, the regressions also control for a variety of predetermined characteristics that still vary between siblings.

The impact of preschool attendance on school progression follows an interesting dynamic pattern. In both Kenya and Tanzania, children often enroll in preschool late and only proceed to primary school once they finished it. At early ages (7-9 years old), these children have therefore completed fewer school grades than their same-aged peers who did not attend preschool. However, we find that once enrolled in primary school, children who attended preschool progress through grades faster and are less likely to drop out. Eventually, they thus catch up with their peers and at ages 13-16 have accumulated the same number of grades of schooling in Kenya and about 0.1 more grades in Tanzania.

In terms of cognitive skills, the estimates for the composite test score show that children who went to preschool outperform their peers in the long run. In Kenya, this effect fades in at early ages and soon stabilizes at a gain of 0.12 standard deviations (SD). In Tanzania, in contrast, an initial gain of 0.11 SD fades out slightly to 0.08 SD later on. Separate regressions reveal that among 13- to 16-year-olds, preschool attendance raises the likelihood of mastering basic, second-grade literacy and numeracy by 3 percentage points in Kenya and by 5 percentage points in Tanzania.

To ensure that these results are not driven by selection, we perform a variety of falsification tests and robustness checks. For example, we

[2]Similar strategies have been used by Currie and Thomas (1995), Garces, Thomas, and Currie (2002), and Deming (2009) to estimate the impacts of the Head Start program in the United States, and by Berlinski, Galiani, and Manacorda (2008) to estimate the effects of preschool in Uruguay.

show that children who attended preschool do not differentially benefit from other educational inputs such as private tutoring, suggesting that our findings are not due to child-specific investments based on unobserved characteristics. Applying the method developed by Oster (2017), we also judge the importance of omitted variable bias more generally by observing the sensitivity of our regression results to the addition of controls. From this approach, selection on unobserved factors would need to be at least four times as large as selection on observed factors to explain away the long-term impacts on the composite test score.

Our paper contributes to a growing literature on the impacts of preschool education on children's outcomes in developing countries, which has focused mostly on Asia and Latin America and which is reviewed in detail in Nores and Barnett (2010) and in Rao et al. (2014). Using retrospective data on preschool enrollment and the same within-household estimator as we do, Berlinski, Galiani, and Manacorda (2008) find that Uruguayan children who attended preschool accumulate 0.8 more years of education by age 15. Applying the same strategy to Egyptian data, Krafft (2015) finds that preschool attendance leads to an additional 0.4 years of schooling among 18-29 year-olds. Behrman, Cheng, and Todd (2004), Berlinski, Galiani, and Gertler (2009), and Brinkman et al. (2017) similarly document positive short-term effects of preschool attendance on children's cognitive skills in Bolivia, Argentina, and Indonesia, respectively. In contrast, a randomized evaluation of a preschool construction program in Cambodia found negative short-term impacts on test scores of targeted children, a result that is partly explained by a shift from underage enrollment in primary school to enrollment in preschool (Bouguen et al., 2018).

To the best of our knowledge, the only other rigorous study of preschool effects in Sub-Saharan Africa is the paper by Martinez, Naudeau, and Pereira (2013), who report on an experimental evaluation of a model preschool program in the Gaza province of Mozambique. The authors find that two years after the start of the program, children were more likely to be enrolled in primary school and had higher cognitive and socio-emotional skills. In contrast to this small-scale evaluation, our study uses nationally representative data on preschool attendance and learning outcomes from two countries. Moreover, unlike most of the previous literature, we are able

to examine the longer-term effects of preschool attendance.

2. Institutional background

2.1. Education and preschools in Kenya

Basic education in Kenya comprises three years of preschool, eight years of primary school, and four years of secondary school. Preschool consists of three distinct grades – baby class (ages 3-4), nursery (ages 4-5), and pre-unit (ages 5-6) –, even though in practice children of different ages are often taught together in the same classroom. Attendance is not compulsory, and children who do not go to preschool typically stay home to help with household chores instead. In the year after they turn six, all children are supposed to enter primary school, which has been free of charge since fees were abolished in 2003. At the end of primary school, students take a national leaving exam (the Kenya Certificate of Primary Education), which largely determines which secondary schools they can enter.

There are two broad types of preschools in Kenya. First, public preschools are run by the government and are usually attached to a primary school. In rural areas, these preschools are often the only available option. Second, private preschools are owned and run by a variety of providers, including non-governmental and faith-based organizations, community-based associations, and private-for-profit agents. They comprise a diverse range of institutions, including: highly unregulated and unregistered non-formal preschools, which are mainly located in informal urban settlements; formal private preschool academies in middle and high-income areas; and a small number of exclusive private preschools catering to very high-income households. All preschools (both public and private) charge tuition fees, which vary widely depending on the type of institution attended, with non-formal private preschools charging the lowest fees. Teacher-student ratios similarly vary, with a recent government publication putting the national mean at 1:28 (Ministry of Education, Science and Technology, 2015).

Unlike in many European countries and the United States, where pre-primary education for younger children typically focuses on play, preschool studies in Kenya are highly academic: students sit at desks and listen to the teacher teach in a classroom-like setting. Curricula, while not standardized,

tend to emphasize the learning of basic numeracy and literacy skills via memorization and recitation. In contrast, only little attention is paid to the development of socio-emotional skills. As children often spend more than 35 hours per week in preschool, most institutions offer a feeding program, which is financed via the tuition fees or via special meal fees (Bidwell, Parry, and Watine, 2013).

We now present some stylized facts to provide context to our results below. First, pre-primary education in Kenya has expanded substantially over the last few decades, leading to high and growing attendance rates: for example, whereas 79% of children born in 1997 went to preschool, 83% of children born in 2004 did so.[3] Second, late enrollment in primary school is common: for example, in 2013, 14% of 7-year-olds were not yet enrolled in primary school, with the vast majority still attending preschool. One reason for this is that enrollment in pre-primary education is often late itself, and parents prefer their children to complete preschool before proceeding to primary school (Bidwell, Parry, and Watine, 2013). Third, students are frequently behind grade for age, either because they enrolled late or because they repeated, with half of all sixth-grade students in 2007 having repeated at least once (Hungi, 2010). Fourth, while enrollment is high compared to other countries in Sub-Saharan Africa, some dropout occurs, with 5% of 13-year-olds reporting not to be enrolled in school in 2013.[4]

2.2. Education and preschools in Tanzania

Basic education in Tanzania consists of two years of preschool, seven years of primary school, four years of lower ('ordinary') secondary school, and two years of upper ('advanced') secondary school. Children can enter non-compulsory pre-primary education at age 5, with those not attending typically staying home to help out with household duties. Children are

[3]All enrollment statistics in this section are based on data from the nationally representative Uwezo surveys, which we describe in detail in the next section. For the preschool figures, we focus on cohorts who were at least 10 years old in our data in order to account for the frequent late enrollment. For dropout rates, we focus on 13-year-olds, who have typically not completed primary school yet in either country. When reporting statistics for individual years, we choose 2013 as this is the latest year in which nation-wide Uwezo surveys were conducted in both Kenya and Tanzania.

[4]For further details on the education system in Kenya, see Tooley, Dixon, and Stanfield (2008), Heyneman and Stern (2014), and Edwards Jr., Klees, and Wildish (2015).

supposed to enter primary school in the year after they turn 7. At the end of grade 7, they take a school leaving exam which regulates access to public secondary schools. Tuition fees for primary education were abolished in 2002, but secondary schools still levied fees during our study period.

Preschools in Tanzania are predominantly public: in 2016, 95% of preschool students were enrolled in a government-run institution (President's Office of the United Republic of Tanzania, 2016). They are often attached to a primary school and charge varying tuition fees. Classrooms are often understaffed, leading to high teacher-student ratios of about 1:50 on average around the year 2008 (World Bank, 2012). Due to these shortages, pre-primary students are sometimes taught together with older, primary school students in multi-grade classrooms. Unlike in Kenya, the Tanzanian government has adopted an official preschool curriculum, which emphasizes the development of both cognitive and socio-emotional skills. In practice, however, preschool teachers often have little or no knowledge of the official curriculum and tend to focus on formal instruction in basic literacy and numeracy (Mligo, 2016). Finally, most preschools offer a feeding program, is financed via the tuition fees or via special meal fees.

Like for Kenya, we now present some stylized facts that provide further background information on the education system in Tanzania. First, there has been a rapid expansion of pre-primary education, with the attendance rate rising from 61% to 69% for the cohorts born in 1995 and 2004, respectively. This rise is spurred by the Tanzanian government, whose expansion strategy has been to attach pre-primary classrooms to existing primary schools. Second, children frequently enroll in primary school late, and some never enroll at all: in 2013, for example, 9% of 8-year-olds reported to be enrolled neither in school nor in preschool. Third, grade repetition rates are lower than in other Sub-Saharan African countries, with 20% of sixth-graders in 2007 having repeated at least once (Hungi, 2010). Fourth, dropout during primary school ages is substantially higher than in Kenya, with 12% of 13-year-olds not enrolled in school in 2013.

3. Data

3.1. The Uwezo surveys

The Uwezo initiative has been conducting large-scale assessments of school-age children's literacy and numeracy skills in Kenya, Tanzania, and Uganda since 2009. The assessments are administered as part of repeated cross-sectional household surveys, which are representative at the district level. An important advantage of this design is that skills are measured also for children who are currently not enrolled in school. The surveys collect information from children aged 6-16 (7-16 in Tanzania) on their current enrollment and highest grade attended as well as on a variety of child and household characteristics. Crucially for our purposes, in recent waves respondents were also asked whether they ever attended preschool.

The literacy and numeracy assessments measure core competencies that children should have achieved after two years of schooling according to the national curriculum. Literacy tests in both English and Swahili assess the following four competencies in order of rising difficulty: (1) recognition of letters, (2) recognition of words, (3) reading a paragraph, and (4) reading a short story. Numeracy tests measure the following six competencies in order of rising difficulty: (1) counting, (2) recognition of numbers, (3) rank ordering of numbers, (4) addition, (5) subtraction, and (6) multiplication. A student's score on each test equals the highest competency level achieved, with a zero indicating that she did not even master the simplest skill assessed. Previous analyses of Uwezo data have shown that a large proportion of students even in higher grades does not master these second-grade competencies (Jones et al., 2014; Uwezo, 2015).

3.2. Variable definitions

The key explanatory variable in our regressions is an indicator for whether a child has attended preschool or not. This variable is based on retrospectively reported information, which has the major advantage that we can estimate longer-term impacts by relating it to current outcomes. A potential concern with retrospective data is that it may be contaminated by recall error; specifically, if such recall error systematically depends on preschool attendance, this could lead to bias in our estimates (Garces, Thomas, and Currie, 2002). We therefore show in a robustness check that

systematic recall error is unlikely to drive our results. For a subsample of children, we also observe for how many years they attended preschool. Below, we use this information to estimate effects at the intensive margin.

Our main analysis focuses on two outcomes: the highest school grade attended and a composite test score. We observe the highest grade attended both for children who are currently enrolled in school and for those who dropped out, with children who are still in preschool coded as having zero grades attended.[5] Because all our regressions include age dummies, this variable is best interpreted as a measure of school progression. We construct the composite test score by first standardizing the literacy (both English and Swahili) and numeracy scores by country, Uwezo survey wave, and age to have mean zero and standard deviation one. In a second step, we then average these scores for each student and standardize the resulting composite test score again.

The control variables include a variety of socio-demographic characteristics, such as age and gender, mother's education, and an index of current household wealth. Moreover, we construct two proxies for early-life economic conditions at the district level from external data sources. The first proxy is the log of average night light density, which has been shown to be a good measure of economic activity (Henderson, Storeygard, and Weil, 2012). The second proxy consists of two separate dummies for positive and negative rainfall shocks, defined as rainfall above the 80th percentile and below the 20th percentile of the long-term district mean, respectively. Rainfall shocks have been used widely as a measure of income shocks in rural economies; see Shah and Steinberg (2017) for a recent example. In order to allow for differential impacts of economic conditions at different ages, we compute our two proxies separately at each age before school entry (ages 0-5 in Kenya and ages 0-6 in Tanzania) for each child. In Online Appendix A, we provide many more details on the construction of these and all other variables used in the empirical analysis.

[5] A previous version of this paper reported results for the highest grade completed rather than the highest grade attended. The main advantage of focusing on the highest grade attended is that it allows us to distinguish between first graders and children who are still in preschool, both of which have zero grades completed.

3.3. Sample selection and descriptive statistics

We use data from all available waves of the Uwezo surveys with information on preschool attendance. These are the 2013 and 2014 waves in Kenya and the four waves conducted between 2011 and 2014 in Tanzania. In Uganda, the only nationally representative Uwezo survey which asked about preschool attendance was conducted in 2013. Unfortunately, this key information is missing for 49% of children in the data for this wave, which led us to exclude Uganda from the analysis (preschool attendance is observed for all children in Kenya and Tanzania). We restrict our attention to children aged 7/8 and above in Kenya/Tanzania because some younger children were still officially of preschool age at the time of the survey. In order to ensure that we focus on comparable siblings in our within-household analysis, we also drop from the sample any children who report never to have enrolled in preschool or school.[6] Our final sample comprises more than 500,000 children with information on preschool attendance and the two main outcomes described above.

Table 1 reports summary statistics for key variables separately for each country. Almost a fifth of children have mothers without any formal education, and more than two thirds live in rural areas. In Kenya, 85% of children attended preschool for an average length of 2.1 years, with the corresponding figures for Tanzania being 62% and 1.3 years. The vast majority of the sample are currently enrolled in education, a statistic that is partly due to our focus on children who ever enrolled in preschool or school. On average, they have attended about four and half grades, but only 58% in Kenya and 43% in Tanzania possess second-grade skills, as measured by an indicator for achieving the highest competency level in the numeracy test and at least one of the two literacy tests. Finally, Appendix Table 1 presents enrollment statistics and outcome means separately by age. As shown there, children tend to be behind grade for age, with a non-negligible share of the younger school-age children still attending preschool. This finding will be important for the interpretation of our results below.

[6]Children of school age who have never enrolled in preschool or school include, for example, children with disabilities. Not surprisingly, if we include these children in our sample, the estimated returns to preschool are substantially higher.

4. Empirical strategy

The main challenge in identifying the causal effects of preschool attendance on later outcomes is that selection into pre-primary education is likely non-random. For example, more educated parents may have a stronger preference for preschool education while also fostering their children's learning in other ways. In this case, a simple regression that does not control for this selection would yield estimates that are biased upward. To address this challenge, we follow a strand of previous literature (Currie and Thomas, 1995; Garces, Thomas, and Currie, 2002; Berlinski, Galiani, and Manacorda, 2008; Deming, 2009) and estimate models with household fixed effects, thus holding constant all determinants of preschool attendance and outcomes that do not vary between siblings:

$$Y_{ij} = \alpha + \beta PRE_{ij} \times \mathbf{AGE}_{ij} + \gamma \mathbf{AGE}_{ij} + \delta \mathbf{X}_{ij} + \eta_j + \varepsilon_{ij}. \qquad (1)$$

Here, i indexes individuals and j indexes households, Y_{ij} is the highest grade attended or the composite test score, PRE_{ij} is the indicator for whether the child has attended preschool, and $\mathbf{AGE}_{ij}$ is a series of age dummies. $\mathbf{X}_{ij}$ is a vector of controls that include interactions between age and cohort effects, dummies for birth order and their interaction with gender, and all the socio-demographic characteristics and early-life economic conditions shown in Table 1. Importantly, we allow the effects of preschool attendance to vary by age by interacting PRE_{ij} with $\mathbf{AGE}_{ij}$; in practice, in order to avoid cluttering, the results tables report estimates from specifications in which PRE_{ij} is interacted with indicators for three age groups instead.

The regression in equation 1 identifies the causal effect of preschool attendance under the assumption that among siblings, selection into preschool is uncorrelated with any other determinants of the outcome. While comparatively weak, this assumption might be violated for several reasons, two of which are particularly salient. First, given that pre-primary education is costly, household income shocks around preschool age may be driving siblings' differential enrollment. Because such income shocks can influence children's educational success also in other ways (e.g. Shah and Steinberg, 2017), this could introduce bias into our estimates. We address this concern by including detailed district-level controls for early-life

economic conditions in our regressions. Second, households with limited resources may choose to invest only in children with the "highest potential."[7] In this case, we would expect children who attended preschool to differentially benefit also from other investments such as private tutoring. However, in a robustness check below, we find no evidence in this regard.

Which factors drive the between-sibling variation in preschool attendance in our data if not income shocks and differential investments based on relative "potential?" We investigate this question in Table 2, which reports results of regressions of the indicator for preschool attendance on the control variables. Columns 1 and 3 show estimates from a specification without household fixed effects for Kenya and Tanzania, respectively. In both countries, children of educated mothers and from wealthier households are more likely to have attended preschool, underlining the importance of controlling for between-family differences. Columns 2 and 4 show that once household fixed effects are included in the regressions, most of the factors that still vary between siblings are no longer predictive of preschool attendance, including the proxies for early-life economic conditions.[8]

The lower part of Table 2 reports the coefficients on the cohort dummies. There is a marked and nearly monotonic trend in both countries, with later cohorts being significantly more likely to have attended preschool. This trend is especially pronounced in Tanzania, where the attendance rate was much lower than in Kenya at baseline (see Section 2) and where it increased by 15 percentage points over the twelve cohorts in our sample. These estimates suggest that the expansion of pre-primary education during our study period led to differences in preschool availability between siblings, which in turn are responsible for differences in attendance. As long as these changes in availability are unrelated to changes in other determinants of

[7]Such reinforcing behavior has been found in several previous studies on developing countries, see Almond and Mazumder (2013). Alternatively, compensatory behavior might lead to negative selection into preschool, biasing our estimates downward.

[8]To avoid cluttering, rather than separate dummies for early-life economic conditions at each age, specifications in Table 2 simply include the number of positive and negative rainfall shocks and the average log night lights before school entry. Results are qualitatively similar if we include the full set of controls instead; in particular, early-life economic conditions appear to be largely orthogonal to preschool attendance. Notably, this is not due to poor measurement, as these variables are highly predictive of children's literacy and numeracy skills (results available upon request).

educational outcomes, this implies that we identify the true causal effects of preschool attendance in the analysis below.[9]

5. Results

We now present our main results. The following subsection reports estimates of the effect of preschool attendance on school progression, and the second subsection shows the corresponding impacts on literacy and numeracy skills. We then present results from regressions that probe for heterogeneity of these effects by children and household characteristics. Finally, in subsection four, we interpret our findings and compare them to the results from the previous literature.

5.1. Preschool attendance and school progression

Table 3 shows estimates of the effect of preschool attendance on the highest grade of school attended. To avoid cluttering, the underlying specifications allow the impact to vary across three age groups – up to 9 years old, 10-12 years old, and 13-16 years old – rather than across specific ages (we discuss effects at specific ages below). Regressions in this and all other tables in the paper are weighted using the sampling weights provided with the Uwezo data, and standard errors are clustered at the district level.[10]

Column 1 reports estimates from a parsimonious specification which only controls for age dummies, their interaction with cohort dummies, and district fixed effects. Due to the frequent late enrollment in pre-primary and subsequently primary education, children who went to preschool initially have accumulated fewer grades than their peers who directly entered primary school. However, these children also progress through grades at a faster pace and eventually overtake their peers: at ages 13-16, they have attended 0.18 and 0.31 more grades in Kenya and Tanzania, respectively.

As discussed above, the results in column 1 are unlikely to reflect the causal impact of preschool attendance due to selection bias. To overcome this problem, columns 2-4 successively add controls for socio-demographic

[9]Ideally, we would like to further investigate this hypothesis using data on preschool openings by district and year. Unfortunately, however, such data do not appear to exist.

[10]Unweighted estimates are qualitatively and quantitatively similar and are reported in Online Appendix Table B.1.

characteristics, early-life economic conditions, and household fixed effects to the regressions. Consistent with the idea of positive selection into pre-primary education, this tends to reduce the coefficients: from our preferred specification in column 4, children who went to preschool are now estimated to accumulate the same number of grades in Kenya and about 0.1 more grades in Tanzania by ages 13-16.[11]

An interesting question is whether children who went to preschool catch up with their peers in terms of grades completed because they skip more or repeat fewer grades while in school, or because they are less likely to drop out of school. To investigate this issue, column 5 of Table 3 presents estimates of the effect of preschool attendance on enrollment based on our preferred specification with household fixed effects. The results indicate that children who went to preschool are indeed more likely to be enrolled, especially at higher ages and in Tanzania. Thus, lower dropout is at least partly underlying the catch-up observed in column 4.[12] Finally, Figure 1 plots estimates from regressions of the highest grade completed and enrollment status in which the effect of preschool attendance is allowed to differ at each age, rather than across age groups. The plots show that the impacts of attendance on these outcomes rise almost monotonically with age, confirming the results from Table 3.

[11]Besides household fixed effects accounting for selection on unobserved factors, there are at least three further potential explanations for why the estimated long-term benefits of preschool attendance in column 4 are smaller than in the previous columns. First, attenuation bias due to measurement error in attendance is aggravated in the between-sibling specification. Second, the inclusion of household fixed effects nets out any positive sibling spillovers. Third, the effects in column 4 are identified only from households with "within" variation, which might differ from those in the full sample. Investigating this last possibility, we found that households with both attending and non-attending children were larger, poorer, and more likely to be located in a rural area. However, when we restricted the sample to these households only, the inclusion of household fixed effects similarly led to a decline in the estimated long-term benefits of preschool. This suggests that identification based on a different sample is not driving the change in coefficients between columns 3 and 4 of Table 3.

[12]Unfortunately, the Uwezo data do not contain information on school starting age and grade repetition, which prevents us from fully disentangling the mechanisms behind this catch-up. In column 5 of Table 3, note that for the younger two age groups, enrollment rates are very high and the predicted probabilities from the linear probability model sometimes exceed 100 percent. Addressing this issue, we confirmed that probit models and a simple comparison of means also suggest that children who attended preschool are more likely to be currently enrolled in school.

5.2. Preschool attendance and literacy and numeracy skills

Table 4 reports estimates of the effect of preschool attendance on children's literacy and numeracy skills. Column 1 shows results from a specification with only basic controls and the composite test score as outcome. In Kenya, children who went to preschool have slightly higher scores than their peers at early ages, and this advantage grows to a sizable 0.1 SD for the two older age groups. In contrast, in Tanzania, children with pre-primary education outperform their peers by 0.26 SD already early on, but this difference decreases to 0.22 SD for the group of 13-16-year-olds. Columns 2-4 successively add control variables and household fixed effects to these regressions. Similar to the pattern found for the highest grade attended, this substantially reduces the size of the estimates for Tanzania, where the impact for the oldest age group is now estimated at 0.08 SD. In contrast, the coefficients for Kenya are relatively stable across specifications, suggesting that there is little selection into preschool based on academic ability.

Column 5 presents results for a specification with the dummy for achieving second-grade literacy and numeracy skills as dependent variable. In both countries, the effect for the youngest age group is close to zero. This should come as no surprise because preschools are unlikely to teach children such advanced skills, and because attendance actually leads to a reduction in grades completed at these ages (see Table 3). Note that this finding implies that the positive impact on test scores for young children in Tanzania in column 4 must be due to pre-primary education boosting their very basic literacy and numeracy skills. Mirroring the long-term gains on the composite test score, the estimates for 13-16-year-olds show a 3.2 (4.7) percentage point increase in the likelihood to achieve second-grade skills in Kenya (Tanzania), which corresponds to a sizable 4 (7) percent over the mean. Finally, Figure 2 shows age-by-age impacts on test scores and the dummy for basic skills which confirm the patterns observed in Table 4.[13]

[13]Online Appendix Figure B.1 shows age-by-age impacts separately for English literacy, Swahili literacy, and numeracy scores. The estimates are very similar to the ones for the composite test score shown in Figure 2.

5.3. Heterogeneity

In Table 5, we explore the heterogeneity of the preschool impacts along several dimensions. Columns 1-3 report estimates from specifications that allow the effects to differ by length of attendance. Note that this information is available only in the 2013 and 2014 waves of the Uwezo survey, such that the sample size for Tanzania in these regressions is reduced by two thirds. Focusing on Kenya, the results reveal that at ages 7-9, children who went to preschool for two or three years have attended fewer school grades than those who went for only one year, likely because they entered primary school later. However, this difference shrinks substantially over time, suggesting that children with longer preschool attendance progress through grades at a faster pace. The impacts on the composite test score similarly tend to be less positive for children with several years of pre-primary education. This could be due to the fact that they are further behind in school, or due to negative selection at the intensive margin, with parents keeping children with lower academic ability in preschool for longer. Moreover, the fact that children of different ages are often taught together in preschool implies that attending for more years does not necessarily mean learning higher-level skills, a reality that might also contribute to the lack of a positive effect of longer attendance on test scores.

Columns 4-5 show that in both countries, the effects of preschool attendance are consistently more positive for girls, even though the differences are not always statistically significant at conventional levels. Furthermore, columns 6-7 reveal that the improvements in literacy and numeracy skills among 13- to 16-year-olds tend to be larger for those with uneducated mothers: in Tanzania, for example, preschool attendance raises the composite test score for these children by 0.14 SD, compared to 0.06 SD for children with mothers who have at least some formal education. Finally, columns 8-9 show that in Tanzania only, longer-term gains in the number of grades attended and test scores are substantially larger in more urban, high-economic-activity areas, as proxied by living in a district with night light density above the 85^{th} percentile of the national distribution.

5.4. Discussion and comparison with previous studies

The results above reveal an interesting dynamic pattern of preschool impacts on children's school progression and their cognitive skills. Due to their frequent late enrollment in primary school, children who attended preschool initially fall behind their same-aged peers in terms of school grades attended. However, they catch up in the long run, partly because they are less likely to drop out of school. Finally, at ages 13-16, children who went to preschool show increased levels of learning as measured by test scores and the indicator for mastering basic, second-grade literacy and numeracy skills. This is an important finding since acquiring such skills is often seen as an essential step towards alleviating poverty.

An interesting question is how exactly preschool attendance leads to these higher levels of learning. One obvious explanation is that with their focus on teaching basic literacy and numeracy, preschools give children a head start that makes it easier for them to follow the primary school curriculum. An implication of this explanation is that these students should have higher skills already at school start. To what extent this is indeed the case is not immediately obvious from our results above, which are based on same-age comparisons that blend the potential effect of a skill boost from preschool attendance with any effect on learning due to later school entry and lower grade attainment. In order to disentangle these two channels, one would ideally want to compare children who differ in terms of their preschool attendance but who started school at the same time. Unfortunately, such a comparison is not feasible here because school starting age is not observed in the Uwezo data.

As an alternative way to separate the test score impacts due to learning in preschool from the impacts due to later school entry, Table 6 reports results from specifications that control for the highest grade attended. If children do indeed get a head start from attending preschool, we would expect the coefficients in these regressions to increase compared to our main results, especially for the youngest children who are furthest behind in school. This turns out to be the case: for example, the effect for 7- to 9-year-olds in Kenya is now estimated to be 0.16 SD, compared to only 0.04 SD when their lower grade attainment is not taken into account. These results suggest that preschools do in fact give children a head start in the

learning of literacy and numeracy skills.

We continue our discussion by comparing our results to the previous literature on preschool effectiveness in developing countries. Focusing first on school progression, our estimates are qualitatively similar to those by Berlinski, Galiani, and Manacorda (2008) for Uruguay. Like us, the authors find that children who go to preschool initially fall behind in terms of grade attainment but that they progress through school faster later on, leading to an increase in educational attainment by 0.8 years of schooling at age 15. For Egypt, Krafft (2015) similarly shows that children with pre-primary education accumulate 0.4 more years of schooling by ages 18-29. While our longer-term estimates for Kenya and Tanzania are substantially smaller, the monotonic trends in Figure 2 suggest that children who attended preschool might eventually acquire more years of schooling also there.

Turning to cognitive skills, Berlinski, Galiani, and Gertler (2009) find that in Argentina, preschool attendance increases third-grade students' math and language test scores by 0.23 SD. This grade-specific impact is best compared to our results from Table 6, which account for children's late entry into primary school and which reveal a somewhat smaller increase of 0.16 SD in the composite test score for the youngest age group in both countries. For Indonesia, Brinkman et al. (2017) show that three years after the establishment of early childhood services, children from poor households improved by 0.20 SD on an index of language and cognitive development, an estimate that is broadly similar to our findings. Behrman, Cheng, and Todd (2004) and Martinez, Naudeau, and Pereira (2013) also document positive effects on measures of child development that include cognitive skills, but which are difficult to compare directly to our outcomes.

Finally, a recent study by Bouguen et al. (2018) shows that in Cambodia, 6-year-old children scored 0.19 SD lower on an index of cognitive development one year after preschools were constructed in their villages. Investigating potential channels, the authors find that preschool construction led to a shift from early enrollment in primary school to enrollment in preschool, suggesting that the decrease in cognitive skills was partly due to the lower emphasis on literacy and numeracy skills in the preschool

curriculum.[14] Interpreted in the light of our results, this shift in enrollment led to a decrease in grades attended early on. While Bouguen et al. (2018) cannot investigate the longer term consequences of this change, our estimates reveal that despite low or negative initial returns, children who attend preschool can catch up in terms of grade attainment and strongly benefit in terms of learning later on. These longer-run estimates on cognitive skills in particular are a key contribution of our paper over the previous literature, which has only been able to study short-run effects.

6. Robustness

We now address several potential concerns regarding the interpretation of our results. First, we argue above that the between-sibling variation in preschool attendance is likely due to changes in availability, which came about because of the expansion of the pre-primary sector during our study period. One might worry, however, that this variation instead reflects child-specific investments that are correlated with unobservables. As an example, households with limited resources may choose to invest only in children with the "highest potential." If this is indeed the case, one would expect that families differentially spend on children who are sent to preschool also in other ways. We test this hypothesis by examining whether children who attended preschool are more likely to benefit from two other costly educational inputs observed in our data: private after-school tutoring and enrollment in private school.[15] Table 7 shows that in regressions of indicators for receiving these inputs, the coefficients on preschool attendance are close to zero and very precisely estimated, suggesting that differential investments based on child unobservables are not driving our results.

[14] Also in our setting, early enrollment in primary school might be part of the counterfactual. However, whereas Bouguen et al. (2018) report that 60 percent of children in control villages attended primary school in the year before they reached the official school starting age, only 26 percent of 5-year-olds in the 1999 Kenyan census and 8 percent of 6-year-olds in the 2002 Tanzanian census did so. Note also that respondents in the Uwezo surveys answer separate questions about current and past preschool attendance and current and past school attendance, allowing us to separate the impacts of preschool attendance from those of early enrollment in primary school.

[15] See Wamalwa and Burns (2017) for an analysis of private school effectiveness in Kenya using the Uwezo survey data.

To judge the importance of selection bias more generally, we next ask how large such bias would need to be in order to explain away our main effects. Our analysis builds on the approach presented in Oster (2017), which relies on comparing the coefficients of interest and the R-squared between regressions with and without control variables to gain insights into the influence of omitted variables. Here, we focus on the calculation of δ, which is the ratio of the impact of unobservables to the impact of observable controls that would drive the coefficient on the treatment variable to zero. As a point of reference, Oster (2017) suggests that effects for which $\delta > 1$ can be considered robust. Applying this method to our case, we contrast estimates of the impact of preschool attendance from a specification with only basic controls (as in column 1 of Tables 3 and 4) with those from our preferred specification with household fixed effects. We restrict our sample to households with variation in preschool attendance for this analysis because for all other households, the fixed effects fully explain preschool attendance, leaving no role for selection on unobservables.[16]

Table 8 reports the results from this exercise. Columns 1 and 3 show estimates from specifications with only basic controls, with regressions underlying columns 2 and 4 adding further controls and household fixed effects. As would be expected, the estimates from our preferred specification are generally very similar, though less precise, in the restricted sample compared to the full sample used in the main analysis. Moving from the basic specification to our preferred specification substantially increases the R-squared and tends to decrease the coefficient estimates, in line with what is observed in Tables 3 and 4. Based on these differences, we report the implied δ for preschool impacts for 13- to 16-year-olds, for whom we find the most positive impacts and where selection on unobservables is thus the most relevant concern. In three out of four specifications, δ is greater than one, implying that selection on unobservables would have to be greater than selection on the observed control variables to drive the preschool impacts to zero. For the case of the composite test score, this value is even above four for both countries, which strongly suggests that omitted variable bias

[16] We thank Emily Oster for this suggestion.

is not driving these results.[17]

Another potential concern is that our results may be affected by recall error in the retrospectively reported preschool variable. Note that such recall error can lead to upward bias only if it systematically varies between siblings who did and did not attend preschool; any general, idiosyncratic recall error will simply drive our estimates towards zero. We investigate this issue by taking advantage of the repeated cross-sectional nature of our data, which lets us follow cohorts over time. Appendix Table 2 shows the fraction of children in each cohort reporting to have attended preschool separately for each Uwezo survey wave.[18] If our data were contaminated by recall error, we would expect these fractions to change over time. This is not the case for the majority of cohorts, though, which suggests that recall error does not bias our estimates.

Finally, one might worry that our results are driven by siblings who are very different in age, and who thus grew up under very distinct circumstances. As an example, the impacts on cognitive skills might vary between age groups only because different cohorts attended preschools of very different quality. To address this issue, Online Appendix Table B.1 presents estimates from regressions in which the sample is restricted to siblings who are born at most five years apart. As can be seen there, the results are qualitatively and quantitatively similar to our main findings.

7. Conclusion

Most children in Sub-Saharan Africa enroll in school nowadays, but they learn remarkably little there. One possible reason is that they en-

[17]There is a strong mechanical relationship between highest grade attended and age, as reflected by the high values of R-squared in column 1 of Table 8. This limits the additional explanatory power any observed controls can have in these regressions, which partly explains the lower values for δ in column 2 compared to column 4 of Table 8.

[18]For this exercise, we focus on a comparable sample of districts that were visited in all waves of the Uwezo surveys. We disregard the 2014 wave in Tanzania because only a small subsample of districts were included in that year's survey. In the raw data, we observe level shifts in preschool attendance rates *for all cohorts* between some of the waves, likely because the question asking about preschool attendance changed. In Appendix Table 2, we therefore report regression-adjusted attendance rates after taking out wave fixed effects. Note that level shifts in preschool attendance do not influence our within-household results, which use variation within survey waves.

ter school unprepared, which makes preschool programs that aim to get children ready for school a promising way to improve learning outcomes. While pre-primary education is becoming increasingly common within the region, to date very little is known about its effectiveness.

In this paper, we provide some of the very first evidence of preschool impacts on learning outcomes in Sub-Saharan African. We use data from large-scale surveys of children's educational attainment and cognitive skills from Kenya and Tanzania, which also collect retrospective information on preschool attendance. Our analysis compares the highest school grade attended as well as achievement on standardized literacy and numeracy tests of siblings who did and did not attend preschool. This strategy allows us to control for any determinants of pre-primary enrollment and outcomes that do not vary within households. We provide evidence that the leftover between-sibling variation in attendance is due to changes in availability, which came about because of a large expansion of preschool education during our study period.

Our results show that preschool education leads to important long-term learning benefits: at ages 13-16, children who went to preschool are three and five percentage points more likely to achieve basic, second-grade literacy and numeracy in Kenya and Tanzania, respectively. These gains materialize relatively late because children who attend preschool tend to enter primary school late and thus fall behind early on. However, the skills learned in preschool give them a head start in school, meaning that they can progress through grades faster and eventually catch up with their peers who did not attend preschool in terms of grades attended. Overall, our analysis shows that increasing access to pre-primary education can be an effective instrument to improve learning outcomes in Sub-Saharan Africa.

References

Almond, D., and B. Mazumder. 2013. "Fetal Origins and Parental Responses." *Annual Review of Economics* 5:37–56.

Behrman, J.R., Y. Cheng, and P.E. Todd. 2004. "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach." *Review of Economics and Statistics* 86:108–132.

Berlinski, S., S. Galiani, and P. Gertler. 2009. "The effect of pre-primary education on primary school performance." *Journal of Public Economics* 93:219–234.

Berlinski, S., S. Galiani, and M. Manacorda. 2008. "Giving children a better start: Preschool attendance and school-age profiles." *Journal of Public Economics* 92:1416–1440.

Bidwell, K., K. Parry, and L. Watine. 2013. *Exploring Early Education Programs in Peri-urban Settings in Africa: Nairobi Report*. Report, Innovations for Poverty Action.

Bietenbeck, J., M. Piopiunik, and S. Wiederhold. 2018. "Africa's Skill Tragedy: Does Teachers' Lack of Knowledge Lead to Low Student Performance?" *Journal of Human Resources* 53:553–578.

Bouguen, A., D. Filmer, K. Macours, and S. Naudeau. 2018. "Preschool and Parental Response in a Second Best World: Evidence from a School Construction Experiment." *Journal of Human Resources* 53:474–512.

Brinkman, S.A., A. Hasan, H. Jung, A. Kinnell, and M. Pradhan. 2017. "The Impact of Expanding Access to Early Childhood Education Services in Rural Indonesia." *Journal of Labor Economics* 35:S305–S335.

Cunha, F., and J. Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97:31–47.

Currie, J., and D. Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review* 85:341–64.

Deming, D. 2009. "Early childhood intervention and life-cycle skill development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1:111–134.

Edwards Jr., D.B., S.J. Klees, and J. Wildish. 2015. "Dynamics of Low-Fee Private Schools in Kenya: Governmental Legitimation, School-Community Dependence, and Resource Uncertainty." *Teachers College Record* forthcoming.

Garces, E., D. Thomas, and J. Currie. 2002. "Longer-Term Effects of Head Start." *American Economic Review* 92:999–1012.

Henderson, J.V., A. Storeygard, and D. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102:944–1028.

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

Hungi, N. 2010. "What are the levels and trends in grade repetition?" SACMEQ Policy Issue Series, No. 5, September 2010.

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

Krafft, C. 2015. "Increasing educational attainment in Egypt: The impact of early childhood care and education." *Economics of Education Review* 46:127–143.

Martinez, S., S. Naudeau, and V. Pereira. 2013. "The Promise of Preschool in Africa: A Randomized Impact Evaluation of Early Childhood Development in Rural Mozambique." Working paper.

Ministry of Education, Science and Technology. 2015. "2014 Basic Education Statistical Booklet." Nairobi, Kenya.

Mligo, I.R. 2016. "Teachers' perceptions and concerns about the implementation of the 2005 preschool curriculum in Tanzania." *Early Years* 36:353–367.

Nores, M., and W.S. Barnett. 2010. "Benefits of early childhood interventions across the world: (Under) Investing in the very young." *Economics of Education Review* 29:271–282.

Oster, E. 2017. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business and Economic Statistics*, pp. 1–18.

President's Office of the United Republic of Tanzania. 2016. "Pre-Primary, Primary and Secondary Education Statistics in Brief." Dar es Salaam: President's Office of the United Republic of Tanzania, Regional Administration and Local Government.

Rao, N., J. Sun, J.M.S. Wong, B. Weekes, P. Ip, S. Shaeffer, M. Young, M. Bray, E. Chen, and D. Lee. 2014. "Early childhood development and cognitive development in developing countries: a rigorous literature review." London: Department for International Development.

Shah, M., and B.M. Steinberg. 2017. "Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital." *Journal of Political Economy* 125:527–561.

Tooley, J., P. Dixon, and J. Stanfield. 2008. "Impact of Free Primary Education in Kenya: A Case Study of Private Schools in Kibera." *Educational Management Administration & Leadership* 36:449–469.

UNESCO. 2012. *Global education digest 2012. Opportunities lost: The impact of grade repetition and early school leaving*. Paris: UNESCO.

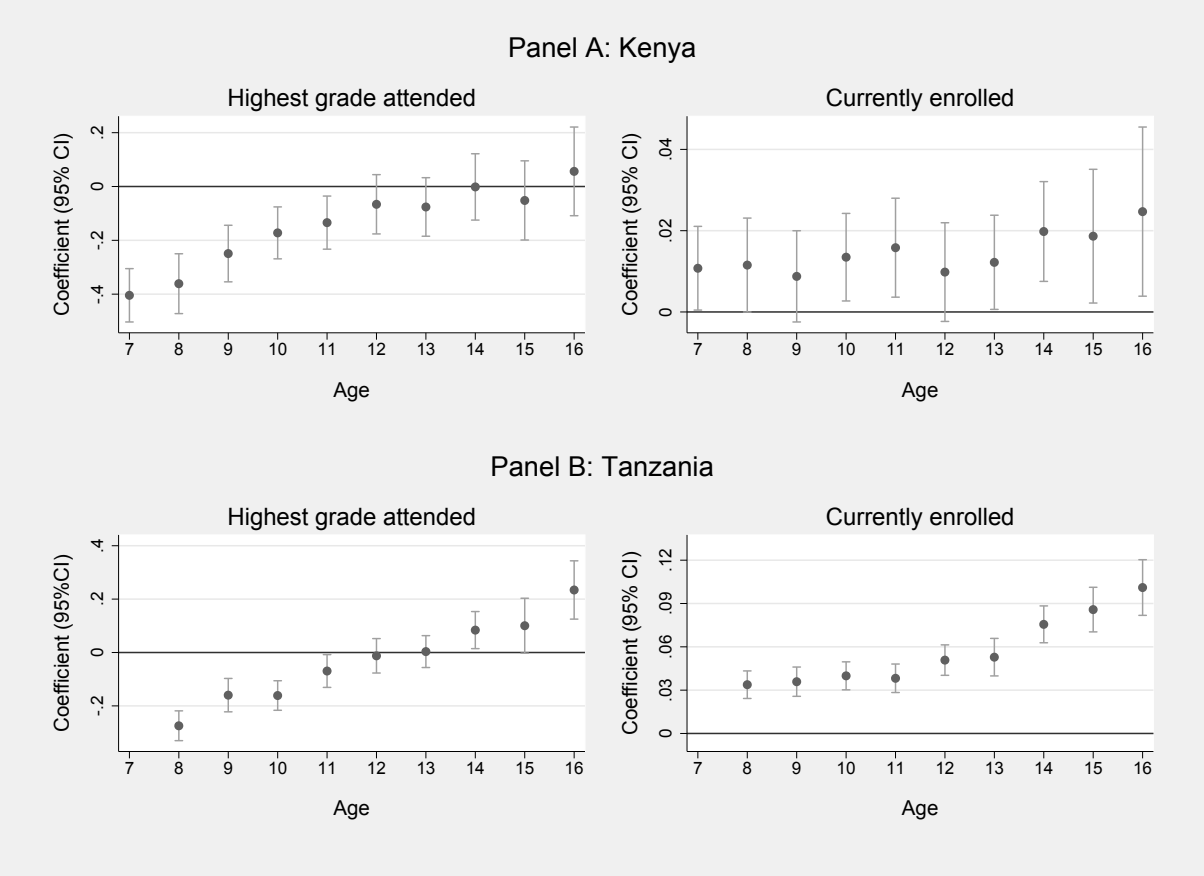
Uwezo. 2015. *Are Our Children Learning? Literacy and Numeracy Across East Africa 2014*. Nairobi: Twaweza East Africa.

Wamalwa, F.M., and J. Burns. 2017. "Private Schools and Student Learning Achievements in Kenya." SALDRU Working Paper No. 202.

World Bank. 2012. *SABER early childhood development country report: Tanzania 2012*. Washington, DC: World Bank.

—. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.

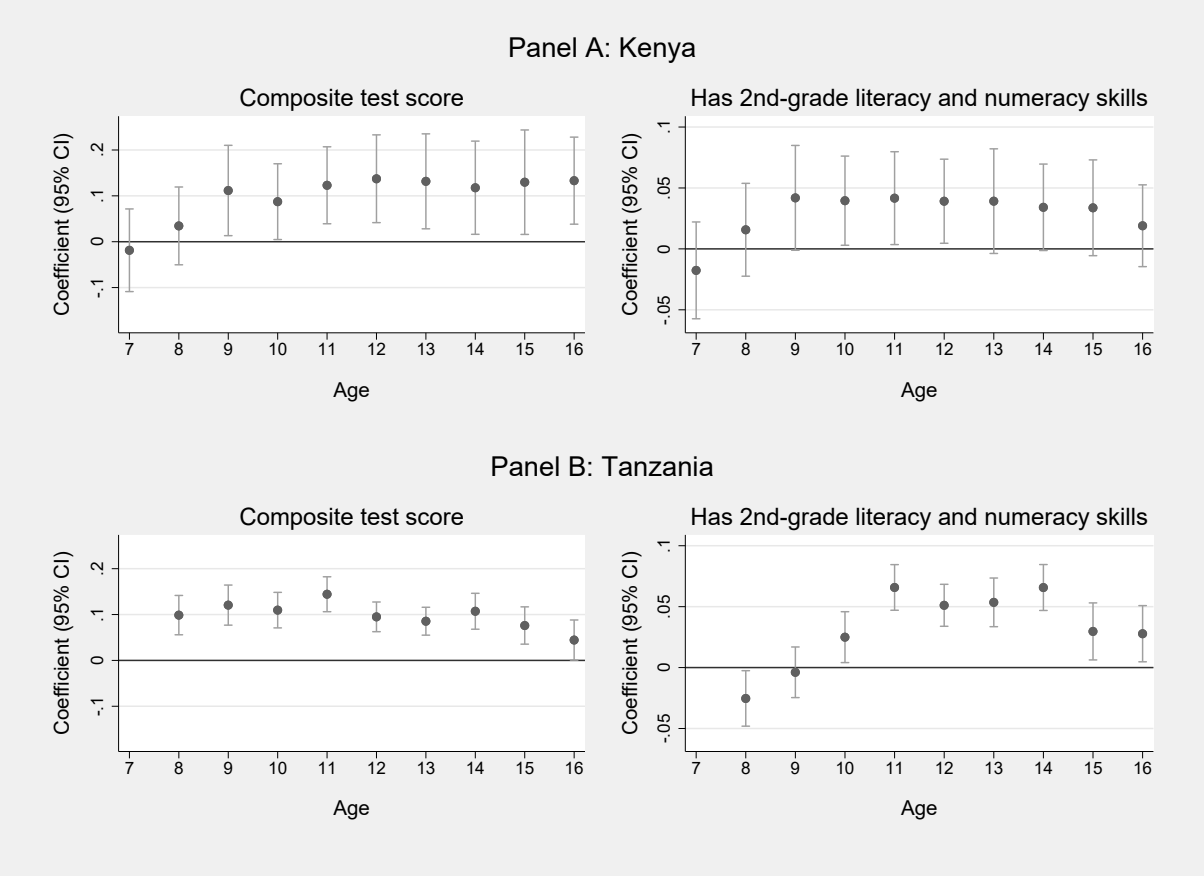
Figure 1
Preschool attendance and school progression, by age



Notes: The figure plots coefficient estimates and 95% confidence intervals from regressions of the highest grade attended and an indicator for current enrollment on preschool attendance. The indicator for preschool attendance is interacted with age dummies, and the figure shows the estimated effect of preschool attendance separately for each age. Specifications are otherwise equal to the household fixed effects regressions reported in columns 4 and 5 of Table 3.

Figure 2

Preschool attendance and literacy and numeracy skills, by age



Notes: The figure plots coefficient estimates and 95% confidence intervals from regressions of the composite test score and an indicator for having second-grade literacy and numeracy skills on preschool attendance. The indicator for preschool attendance is interacted with age dummies, and the figure shows the estimated effect of preschool attendance separately for each age. Specifications are otherwise equal to the household fixed effects regressions reported in columns 4 and 5 of Table 4.

Table 1
Summary statistics

	Kenya	Tanzania
Socio-demographic characteristics		
Age	11.08	11.72
	(2.77)	(2.46)
Female	0.49	0.50
	(0.50)	(0.50)
Mother's education		
None	0.17	0.19
	(0.37)	(0.39)
Some primary or more	0.83	0.81
	(0.37)	(0.39)
No. of children in household	3.09	2.47
	(1.55)	(1.26)
Current household wealth (index)	0.00	0.00
	(1.00)	(1.00)
Rural location	0.67	0.78
	(0.47)	(0.41)
Early-life economic conditions		
No. of negative rainfall shocks	1.44	1.87
	(0.85)	(0.87)
No. of positive rainfall shocks	1.10	1.27
	(0.79)	(0.72)
Log night light density	-1.44	-2.39
	(2.32)	(2.29)
Preschool attendance		
Attended preschool	0.85	0.62
	(0.36)	(0.48)
Years of preschool |attended	2.11	1.32
	(1.08)	(0.75)
Outcomes		
Highest grade attended	4.62	4.42
	(2.53)	(2.28)
Currently enrolled	0.99	0.94
	(0.10)	(0.24)
Composite test score	0.00	0.00
	(1.00)	(1.00)
Has 2nd-grade lit./num. skills	0.58	0.43
	(0.49)	(0.49)
Observations	223,339	293,757
% with within-household variation	3.37	10.61

Notes: The table reports means and standard deviations (in parentheses) of key variables separately for children in Kenya and Tanzania. In regressions, early-life economic conditions are proxied by district-level indicators for negative and positive rainfall shocks and district-level log night lights at each age between 0 and 5 in Kenya (0 and 6 in Tanzania); for conciseness, this table shows totals across all of these ages. Years of preschool are observed only in the 2013 and 2014 waves of the Uwezo survey (N=223,339 in Kenya and N=111,043 in Tanzania). Currently enrolled is an indicator for being currently enrolled in either preschool or school. Has 2nd-grad lit./num. skills is an indicator for achieving the highest competency level in the numeracy test and at least one of the two literacy tests. The final row reports the fraction of children living in households in which at least one child went to preschool and at least one child did not. Further details on the construction of all variables are provided in Online Appendix A.

Table 2
Predicting preschool attendance

	Kenya		Tanzania	
	(1)	(2)	(3)	(4)
Female	0.003	-0.001	0.008**	0.003
	(0.003)	(0.002)	(0.003)	(0.003)
Firstborn	0.005	0.000	-0.002	0.002
	(0.004)	(0.003)	(0.005)	(0.004)
Female × firstborn	-0.001	0.001	-0.004	0.000
	(0.004)	(0.003)	(0.005)	(0.005)
Mother ≥ some primary edu.	0.016**		0.097***	
	(0.007)		(0.006)	
No. of children in household	-0.002		-0.012***	
	(0.002)		(0.002)	
Household wealth index	0.006**		0.052***	
	(0.003)		(0.004)	
Rural location	-0.000*		-0.016	
	(0.000)		(0.022)	
No. of negative rainfall shocks	-0.003**	-0.001	-0.007*	-0.000
	(0.001)	(0.001)	(0.004)	(0.004)
No. of positive rainfall shocks	-0.004	-0.000	0.006	0.007*
	(0.002)	(0.001)	(0.004)	(0.004)
Log night light density	0.009	0.002	0.016	0.011
	(0.006)	(0.005)	(0.016)	(0.012)
Cohort				
1996			0.023***	0.021
			(0.008)	(0.014)
1997			0.030***	0.035***
			(0.009)	(0.011)
1998	0.005	0.002	0.047***	0.050***
	(0.008)	(0.004)	(0.009)	(0.013)
1999	0.012*	0.003	0.062***	0.069***
	(0.007)	(0.004)	(0.010)	(0.014)
2000	0.011	0.006	0.088***	0.084***
	(0.007)	(0.004)	(0.011)	(0.015)
2001	0.019***	0.006	0.090***	0.086***
	(0.007)	(0.005)	(0.011)	(0.016)
2002	0.022***	0.008	0.098***	0.101***
	(0.007)	(0.005)	(0.011)	(0.015)
2003	0.026***	0.004	0.110***	0.106***
	(0.007)	(0.005)	(0.011)	(0.016)
2004	0.023**	0.007	0.113***	0.109***
	(0.009)	(0.005)	(0.012)	(0.017)
2005	0.033***	0.007	0.099***	0.099***
	(0.009)	(0.006)	(0.012)	(0.017)
2006	0.040***	0.014**	0.147***	0.149***
	(0.007)	(0.006)	(0.016)	(0.023)
2007	0.032***	0.013		
	(0.009)	(0.009)		
Household fixed effects	No	Yes	No	Yes
Observations	223,339	223,339	293,757	293,757

Notes: The table reports estimates from regressions of an indicator for preschool attendance on the variables listed in rows and Uwezo wave dummies. Standard errors in parentheses are clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 3
Preschool attendance and school progression

	Highest grade attended				Currently enrolled
	(1)	(2)	(3)	(4)	(5)
	Panel A: Kenya				
Attended preschool					
7-9 years old	-0.234***	-0.252***	-0.242***	-0.336***	0.011*
	(0.040)	(0.042)	(0.042)	(0.045)	(0.005)
10-12 years old	0.017	-0.006	-0.006	-0.123***	0.013**
	(0.024)	(0.023)	(0.023)	(0.044)	(0.006)
13-16 years old	0.178***	0.161***	0.153***	-0.023	0.018***
	(0.034)	(0.034)	(0.033)	(0.055)	(0.006)
Observations	218,728	218,728	218,728	218,728	218,728
	Panel B: Tanzania				
Attended preschool					
8-9 years old	-0.071***	-0.127***	-0.115***	-0.212***	0.036***
	(0.018)	(0.018)	(0.018)	(0.025)	(0.005)
10-12 years old	0.081***	0.028	0.031*	-0.074***	0.044***
	(0.020)	(0.018)	(0.019)	(0.025)	(0.004)
13-16 years old	0.313***	0.258***	0.243***	0.096***	0.076***
	(0.026)	(0.026)	(0.024)	(0.031)	(0.006)
Observations	284,396	284,396	284,396	284,396	284,396
	Controls included in panels A and B				
Age × cohort effects	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	No	No
Socio-demographics	No	Yes	Yes	Yes	Yes
Early-life conditions	No	No	Yes	Yes	Yes
Household fixed effects	No	No	No	Yes	Yes

Notes: The table reports estimates from regressions of the highest grade attended and enrollment status on a dummy for preschool attendance and control variables as indicated in the lower panel. The dummy for preschool attendance is interacted with three age-group dummies (7/8-9, 10-12, and 13-16 years), and the table reports the estimated effect of preschool attendance separately for each group. Socio-demographic controls include the variables shown in Table 1, dummies for birth order and their interactions with gender, and sibling age span. Controls for early-life economic conditions include district-level indicators for negative and positive rainfall shocks and district-level log night lights at each age between 0 and 5 (0 and 6 in Tanzania), all of which are interacted with a dummy for rural location. Specifications in columns 1-3 additionally control for Uwezo wave dummies. See Appendix Table 1 for outcome means by age. Standard errors in parentheses are clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 4
Preschool attendance and literacy and numeracy skills

	Composite test score				2^{nd}-grade lit./num.
	(1)	(2)	(3)	(4)	(5)
	Panel A: Kenya				
Attended preschool					
7-9 years old	0.036*	0.015	0.012	0.042	0.013
	(0.022)	(0.022)	(0.021)	(0.040)	(0.016)
10-12 years old	0.097***	0.078***	0.077***	0.114***	0.040**
	(0.022)	(0.021)	(0.021)	(0.041)	(0.015)
13-16 years old	0.104***	0.095***	0.098***	0.125***	0.032*
	(0.021)	(0.021)	(0.021)	(0.044)	(0.017)
Observations	218,134	218,134	218,134	218,134	218,134
	Panel B: Tanzania				
Attended preschool					
8-9 years old	0.256***	0.204***	0.195***	0.108***	-0.016*
	(0.017)	(0.016)	(0.015)	(0.018)	(0.009)
10-12 years old	0.263***	0.212***	0.209***	0.113***	0.046***
	(0.014)	(0.013)	(0.013)	(0.014)	(0.007)
13-16 years old	0.217***	0.163***	0.169***	0.081***	0.047***
	(0.016)	(0.015)	(0.014)	(0.014)	(0.008)
Observations	288,084	288,084	288,084	288,084	288,084
	Controls included in panels A and B				
Age × cohort effects	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	No	No
Socio-demographics	No	Yes	Yes	Yes	Yes
Early-life conditions	No	No	Yes	Yes	Yes
Household fixed effects	No	No	No	Yes	Yes

Notes: The table reports estimates from regressions of the composite test score and the indicator for achieving second-grade literacy and numeracy skills on a dummy for preschool attendance and control variables as indicated in the lower panel. The dummy for preschool attendance is interacted with three age-group dummies (7/8-9, 10-12, and 13-16 years), and the table reports the estimated effect of preschool attendance separately for each group. Socio-demographic controls include the variables shown in Table 1, dummies for birth order and their interactions with gender, and sibling age span. Controls for early-life economic conditions include district-level indicators for negative and positive rainfall shocks and district-level log night lights at each age between 0 and 5 (0 and 6 in Tanzania), all of which are interacted with a dummy for rural location. Specifications in columns 1-3 additionally control for Uwezo wave dummies. See Appendix Table 1 for outcome means by age. Standard errors in parentheses are clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 5
Heterogeneity

	by number of years attended			by gender		by mother's education		by night light density	
	main effect (1)	× 2 years (2)	× 3 years (3)	main effect (4)	× female (5)	main effect (6)	× some educ. (7)	main effect (8)	× > 85% (9)
Panel A: Kenya									
1. Highest grade attended									
7-9 years old	-0.211***	-0.163***	-0.292***	-0.341***	0.013	-0.270***	-0.024	-0.336***	0.053
	(0.050)	(0.036)	(0.041)	(0.049)	(0.046)	(0.073)	(0.096)	(0.044)	(0.184)
10-12 years old	-0.043	-0.141***	-0.192***	-0.164***	0.083**	-0.094	-0.033	-0.127***	0.009
	(0.048)	(0.030)	(0.044)	(0.047)	(0.036)	(0.070)	(0.086)	(0.037)	(0.211)
13-16 years old	-0.031	-0.050	-0.032	-0.071	0.100**	-0.024	-0.046	-0.054	0.093
	(0.058)	(0.031)	(0.053)	(0.066)	(0.049)	(0.073)	(0.092)	(0.047)	(0.225)
Observations		218,728		218,728		218,728		218,728	
2. Composite test score									
7-9 years old	0.045	-0.051*	0.029	0.034	0.019	0.076	-0.041	0.022	0.104
	(0.041)	(0.027)	(0.034)	(0.046)	(0.032)	(0.072)	(0.091)	(0.043)	(0.098)
10-12 years old	0.134***	-0.049**	-0.017	0.086**	0.055**	0.143*	-0.069	0.116**	-0.013
	(0.043)	(0.021)	(0.030)	(0.041)	(0.024)	(0.079)	(0.094)	(0.045)	(0.104)
13-16 years old	0.168***	-0.064**	-0.068**	0.100**	0.054	0.235**	-0.148	0.136***	-0.037
	(0.046)	(0.024)	(0.033)	(0.046)	(0.038)	(0.100)	(0.120)	(0.049)	(0.103)
Observations		288,084		288,084		288,084		288,084	
Panel B: Tanzania									
1. Highest grade attended									
7-9 years old	-0.237***	-0.273***		-0.256***	0.087***	-0.255***	0.070	-0.205***	-0.000
	(0.054)	(0.076)		(0.028)	(0.029)	(0.043)	(0.049)	(0.027)	(0.068)
10-12 years old	-0.156**	-0.237***		-0.080***	0.012	-0.127***	0.058	-0.095***	0.102
	(0.060)	(0.064)		(0.028)	(0.029)	(0.046)	(0.047)	(0.026)	(0.072)
13-16 years old	-0.052	-0.125*		0.100***	-0.010	-0.047	0.167***	0.048*	0.310***
	(0.058)	(0.069)		(0.032)	(0.039)	(0.055)	(0.062)	(0.027)	(0.104)
Observations	107,825			284,396		284,396		284,396	
2. Composite test score									
7-9 years old	0.080**	0.077**		0.080***	0.054**	0.080**	0.031	0.096***	0.029
	(0.034)	(0.037)		(0.022)	(0.025)	(0.039)	(0.050)	(0.018)	(0.060)
10-12 years old	0.127***	0.017		0.089***	0.048***	0.140***	-0.039	0.097***	0.090***
	(0.025)	(0.027)		(0.016)	(0.018)	(0.037)	(0.044)	(0.015)	(0.032)
13-16 years old	0.124***	-0.056**		0.069***	0.024	0.144***	-0.082*	0.075***	0.105**
	(0.028)	(0.031)		(0.017)	(0.018)	(0.040)	(0.046)	(0.014)	(0.049)
Observations	107,605			288,084		288,084		288,084	

Notes: The table reports estimates from specifications in which the dummy for preschool attendance is interacted with three age-group dummies as in Tables 3 and 4 and additionally with two dummies for attending preschool for two and three years (columns 1-3), a female dummy (columns 4-5), a dummy for having a mother with at least some formal education (columns 6-7), and a dummy for living in a district with night light density above the 85th national percentile in the year 2000 (columns 8-9). The table reports the main effects for each age group as well as the coefficients on the interactions with these characteristics. Base levels in columns 1/4/6/8: attended for 1 year/male/no formal education/district night lights below the 85th national percentile. Outcome variables are indicated in cursive in the rows above the respective regressions. All regressions include household fixed effects and controls as in column 4 of Tables 3 and 4. Standard errors in parentheses are clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 6

Effects of preschool attendance on composite test scores after controlling for highest grade attended

	Composite test score	
	Baseline	Controlling for highest grade attended
	(1)	(2)
	Panel A: Kenya	
Attended preschool		
7-9 years old	0.042	0.157***
	(0.040)	(0.036)
10-12 years old	0.114***	0.148***
	(0.041)	(0.035)
13-16 years old	0.125***	0.177***
	(0.044)	(0.039)
Observations	218,134	218,134
	Panel B: Tanzania	
Attended preschool		
8-9 years old	0.108***	0.157***
	(0.018)	(0.017)
10-12 years old	0.113***	0.126***
	(0.014)	(0.014)
13-16 years old	0.081***	0.104***
	(0.014)	(0.013)
Observations	288,084	288,084

Notes: Column 1 replicates the estimates shown in column 4 of Table 4. Specifications in column 2 add separate dummies for the number of grades attended to these regressions as controls. See the notes to Table 4 for further details on included control variables. Standard errors in parentheses are clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 7
Preschool attendance and other educational investments

	After-school tutoring (1)	Private school (2)
	Panel A: Kenya	
Attended preschool		
7-9 years old	-0.002	-0.004
	(0.011)	(0.010)
10-12 years old	0.005	-0.014
	(0.011)	(0.011)
13-16 years old	0.020	-0.015
	(0.017)	(0.011)
Observations	223,339	208,424
	Panel B: Tanzania	
Attended preschool		
8-9 years old	-0.000	-0.005
	(0.025)	(0.004)
10-12 years old	0.018	-0.006*
	(0.024)	(0.003)
13-16 years old	0.020	-0.001
	(0.019)	(0.003)
Observations	25,346	264,810

Notes: Columns 1 and 2 report estimates from regressions in which the dependent variables are an indicator for receiving private after-school tutoring and and indicator for currently attending private school, respectively. The specifications are otherwise identical to the one in column 4 of Tables 3 and 4. Means of the dependent variables for the youngest/middle/oldest age group: Kenya – after-school tutoring: 0.27/0.31/0.37; Tanzania – after-school tutoring: 0.22/0.25/0.25; Kenya – private school: 0.24/0.15/0.09; Tanzania – private school: 0.06/0.05/0.08. After-school tutoring is only observed in 2014 in Tanzania. Private school is only observed for students who are currently enrolled in school. Standard errors in parentheses are clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 8
Judging the importance of selection on unobservables

	Highest grade attended		Composite test score	
	(1)	(2)	(3)	(4)
	Panel A: Kenya			
7-9 years	-0.703***	-0.423***	-0.104	0.023
	(0.072)	(0.070)	(0.063)	(0.066)
10-12 years	0.110	-0.090	0.251***	0.165**
	(0.070)	(0.078)	(0.064)	(0.066)
13-16 years	0.144*	-0.020	0.092	0.111
	(0.086)	(0.085)	(0.069)	(0.075)
Observations	7,229	7,229	7,141	7,141
R-squared	0.688	0.869	0.114	0.619
δ (13-16 years)		-0.119		11.317
	Panel B: Tanzania			
8-9 years	-0.115***	-0.147***	0.108***	0.093***
	(0.030)	(0.041)	(0.028)	(0.032)
10-12 years	-0.154***	-0.088**	0.073***	0.111***
	(0.034)	(0.038)	(0.023)	(0.022)
13-16 years	0.110***	0.083**	0.128***	0.120***
	(0.042)	(0.038)	(0.022)	(0.024)
Observations	29,981	29,981	30,427	30,427
R-squared	0.622	0.831	0.072	0.637
δ (13-16 years)		1.150		4.553
	Controls included in panels A and B			
Age × cohort effects	Yes	Yes	Yes	Yes
District fixed effects	Yes	No	Yes	No
Socio-demographics	No	Yes	No	Yes
Early-life conditions	No	Yes	No	Yes
Household fixed effects	No	Yes	No	Yes

Notes: The table reports regression estimates that provide the inputs into the computation of Oster's (2017) δ, that is, the ratio of the impact of unobservables to the impact of observable controls that would drive the coefficient on preschool attendance for 13- to 16-year-olds to zero. For details on the underlying method, see text and Oster (2017). For this analysis, the sample is restricted to households with variation in preschool attendance. The regressions are otherwise identical to those in columns 1 and 4 in Table 3 (columns 1 and 2 in the current table) and columns 1 and 4 in Table 4 (columns 3 and 4 in the current table). To calculate δ, we use the Stata command `-psacalc-`, setting the maximum achievable R-squared ('*Rmax*') to 1.3 times the R-squared in the regression with household fixed effects (and at most 1). * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table 1

Grade progression, enrollment, and literacy and numeracy skills, by age

	Highest grade attended	Currently enrolled	Still in preschool	Raw numeracy score	Raw English lit. score	Raw Swahili lit. score	Has 2nd-grade lit./num.
				Panel A: Kenya			
Age 7	1.49	1.00	0.13	3.33	2.00	2.01	0.14
Age 8	2.27	1.00	0.06	4.12	2.43	2.46	0.26
Age 9	3.00	1.00	0.03	4.70	2.82	2.87	0.40
Age 10	3.80	1.00	0.02	5.05	3.10	3.15	0.51
Age 11	4.66	0.99	0.00	5.40	3.36	3.41	0.65
Age 12	5.39	0.99	0.00	5.56	3.53	3.57	0.73
Age 13	6.17	0.99	0.00	5.70	3.66	3.70	0.80
Age 14	6.83	0.98	0.00	5.75	3.75	3.78	0.84
Age 15	7.42	0.97	0.00	5.78	3.80	3.83	0.88
Age 16	8.15	0.95	0.00	5.81	3.84	3.85	0.90
				Panel B: Tanzania			
Age 8	1.73	0.98	0.04	3.09	1.01	1.64	0.10
Age 9	2.43	0.98	0.02	3.71	1.29	2.06	0.18
Age 10	3.19	0.98	0.01	4.17	1.58	2.41	0.27
Age 11	3.96	0.97	0.01	4.62	1.87	2.75	0.39
Age 12	4.73	0.96	0.00	4.88	2.11	2.97	0.46
Age 13	5.55	0.94	0.00	5.12	2.38	3.20	0.56
Age 14	6.19	0.91	0.00	5.30	2.65	3.35	0.64
Age 15	6.71	0.86	0.00	5.39	2.81	3.44	0.69
Age 16	7.28	0.82	0.00	5.47	2.99	3.50	0.73

Notes: The table reports means of the variables indicated in the column heads across children of the age indicated in rows. Raw numeracy scores range from 0 to 6. Raw English and Swahili literacy scores range from 0 to 4. For definitions of all other variables, see the notes to Table 1 and Online Appendix A. Note that the figures reported here can differ from the enrollment statistics reported in Section 2 of the paper because those statistics are based on the unrestricted sample of children in the Uwezo data.

Appendix Table 2

Reported preschool attendance by cohort and wave

Wave:	2011	2012	2013	2014
	Panel A: Kenya			
Cohort				
1997			0.83	
1998			0.84	0.83
1999			0.84	0.84
2000			0.83	0.84
2001			0.84	0.85
2002			0.85	0.84
2003			0.84	0.85
2004			0.85	0.84
2005			0.85	0.85
2006			0.87	0.85
2007				0.85
	Panel B: Tanzania			
Cohort				
1995	0.57			
1996	0.59	0.58		
1997	0.59	0.58	0.60	
1998	0.61	0.60	0.61	
1999	0.63	0.61	0.62	
2000	0.65	0.64	0.64	
2001	0.66	0.65	0.63	
2002	0.69	0.66	0.65	
2003	0.69	0.67	0.64	
2004		0.68	0.65	
2005			0.64	

Notes: The table shows the fractions of children reporting to have attended preschool by country, cohort, and Uwezo survey wave. The sample is restricted to districts that were visited in all waves of the Uwezo survey. We disregard the 2014 wave in Tanzania because only a subsample of districts were sampled (45 districts versus more than 120 districts in the three previous waves). In the raw data, we observe level shifts in preschool attendance rates for all cohorts between some of the waves, likely because the question asking about preschool attendance changed. The table therefore shows regression-adjusted attendance rates after taking out wave fixed effects.

ONLINE APPENDIX TO:

Preschool Attendance, School Progression, and Cognitive Skills in East Africa

Jan Bietenbeck

Sanna Ericsson

Fredrick M. Wamalwa

Online Appendix A: Data Appendix

A1. The Uwezo surveys: sampling and test design

Uwezo, which means 'capability' in Kiswahili, is a non-governmental organization that aims to improve competencies in literacy and numeracy among school-aged children in East Africa. Since 2009, Uwezo has conducted annual assessments of the basic literacy and numeracy skills of children in Kenya, Tanzania, and Uganda. The assessments are administered as part of repeated cross-sectional household surveys, which also collect information on a variety of child and household characteristics and education outcomes. Households are selected in a two-stage sampling design: first, in each census district of each country, 30 enumeration areas (which typically correspond to one or several villages) are sampled with probability proportional to size; then, 20 households in each of these enumeration areas are randomly selected to participate in the survey.[1] The resulting sample is representative at both the national and the district level. Weights which reflect this sampling design and which implement a number of ex-post corrections are provided with the data; we use these weights throughout our analysis.[2]

In participating households, all children aged 6-16 (7-16 in Tanzania) are assessed on core literacy and numeracy competencies that should be achieved after two years of schooling according to the national curriculum. Two separate literacy tests in English and Swahili measure the following four competencies in order of rising difficulty: (1) recognition of letters, (2) recognition of words, (3) reading a paragraph, and (4) reading a short story. The numeracy test measures the following six competencies in order of

[1] A few districts were excluded in some rounds of the survey due to security concerns. In 2014, Tanzania selected households from a random subsample of districts only.

[2] Unfortunately, the weights included with the Tanzanian data over-emphasize the importance of observations in 2014. Specifically, as reported in the previous footnote, only a random subsample of districts was surveyed in 2014, and this wave correspondingly includes less than a third of the observations compared to any previous wave. Nevertheless, the weights in 2014 add up to about 125% of the weights in all previous waves. We attempt to correct for this irregularity by re-scaling the 2014 weights at the district level, using the relative importance of each district in the 2013 wave as a scaling factor. Our results are however robust to using the original weights, not using any weights at all, or dropping the 2014 wave for Tanzania altogether.

rising difficulty: (1) counting, (2) recognition of numbers, (3) rank ordering of numbers, (4) addition, (5) subtraction, and (6) multiplication.[3] For each assessment, there are several test booklets in order to prevent children within the same household from copying each other's answers. A child's score on each test equals the highest competency level achieved, with a zero indicating that she did not even master the simplest skill assessed.

A2. Variable definitions

Household identifier. The data contain a household identifier, which we use to construct household-level variables such as number of children and wealth. Because polygamy is common in some communities, a few households contain children from different mothers. For each child, we observe his/her mother's age and education, which we use to construct a unique mother identifier. Our within-household specifications are based on this more conservative mother identifier rather than the household identifier, even though in practice this makes little difference.

Socio-demographic characteristics. We define a child's cohort as Uwezo survey wave minus age. Mother's education is recorded differently between countries and survey waves; we make this variable comparable by collapsing it into two categories: no education and at least some primary education. To construct the index of current household wealth, we follow Schady et al. (2015) and aggregate the following dwelling characteristics and assets using the first principal component: wall materials, source of lighting, tv, radio, computer (only Kenya), mobile phone, bicycle, motorbike, and motor vehicle. We compute this index separately for each country and normalize it to have mean zero and standard deviation one.

The rural indicator describes the location of the enumeration area. For Kenya, this variable is not included with the publicly available data, but we were able to obtain it directly from Uwezo. For Tanzania, the variable is included in the publicly available data for the 2011 and 2012 survey waves;

[3] In Kenya, children who master multiplication are also assessed on their division skills. We ignore this seventh, higher competency here in order to ensure comparability of test scores across Kenya and Tanzania.

as we were not able to obtain the variable for the 2013 and 2014 waves, it is missing for children observed in these years.[4]

Early-life economic conditions. We construct two proxies for district-level economic conditions using external satellite data on night lights and rainfall. For Kenya, district definitions in the Uwezo data are based on the 2009 census. For Tanzania, we create a crosswalk which maps districts in the Uwezo data to districts in the 2002 census. We use GIS census district boundary files from IPUMS International to compute summary statistics for our two proxies for each district and year.[5]

We obtain the night lights data from the Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS).[6] The data provide yearly measurements of average light density at a fine geographical level, with light density ranging from 0 to 63. For a detailed description of these data, we refer to Henderson, Storeygard, and Weil (2012). Our rainfall measures are derived from the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) data.[7] These data provide annual measures of precipitation since 1981; for details, see Funk et al. (2015).

From the satellite data, we construct a variable measuring average log night lights and indicators for positive and negative rainfall shocks at each age before school entry (ages 0-5 in Kenya and ages 0-6 in Tanzania). In line with recent literature (e.g. Shah and Steinberg, 2017), we define rainfall shocks as precipitation above the 80th percentile and below the 20th percentile of the long-term district mean.

Preschool attendance. Recent waves of the Uwezo survey ask respon-

[4]As usual in survey data, there are some missing values also in other control variables. In order not to unnecessarily reduce sample size, we impute missing values at the sample mean and include separate dummies for missing values on each control variable in all of our regressions.

[5]The district boundary files for the Kenyan 2009 census and the Tanzanian 2002 census are available here: https://international.ipums.org/international/gis_yrspecific_2nd.shtml.

[6]We use the Average Visible, Stable Lights, and Cloud Free Coverages series, which is available here: https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html.

[7]We use the CHIRPS-2.0 global annual yearly data series, which is available here: ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/global_annual/tifs/.

dents whether they ever attended preschool and whether they are currently still enrolled in preschool. From the answers to these two question, we construct our key explanatory variable as an indicator which takes value 1 if a child ever attended preschool and 0 otherwise. In the 2013 and 2014 waves, we moreover have information on length of attendance in years. As a few respondents indicate lengths of attendance far beyond the usual, we winsorize this variable at 3 years in Kenya and 2 years in Tanzania (i.e. at the maximum "normal" length according to the national education system).

Outcome variables. Our first main outcome is the highest grade of school attended. Children who are currently enrolled in preschool are coded as having zero grades attended. Children who are currently in school report the grade they are attending. Children who have dropped out of school report the grade during which they dropped out; for them, the highest grade attended equals the dropout grade. We winsorize the resulting variable such that children can be ahead at most two grades; for example, a 10 year-old child can have attended at most grade six in Kenya and grade five in Tanzania.

The second main outcome variable is the composite test score. We construct this score by first standardizing the English, Swahili, and numeracy scores by country, Uwezo survey wave, age, and test booklet to have mean zero and standard deviation one. In a second step, we then average these standardized scores for each student and normalize the resulting composite again to obtain the score used in the regressions.

In auxiliary regressions, we also use a number of further outcomes. These include an indicator for current enrollment, which takes value 1 if the child reports to be currently enrolled in preschool or school and 0 otherwise. We also construct an indicator for achieving second-grade literacy and numeracy, which takes value 1 if a child achieves the highest competency level in the numeracy test and at least one of the two literacy tests and 0 otherwise.

A3. Sample selection

We use data from all available waves of the Uwezo surveys with information on preschool attendance. These are the 2013 and 2014 waves in

Kenya and the four waves conducted between 2011 and 2014 in Tanzania. We decided to drop Uganda from the analysis because the only survey with national scope and which collected information on preschool attendance there was fielded in 2013, and information on preschool attendance is missing for 49% of respondents in the corresponding data.

We restrict our attention to children aged 7 and above (8 and above) in Kenya (Tanzania) because some younger children were still of preschool age at the time of the survey. In order to ensure that we focus on comparable siblings in our within-household analysis, we also drop from the sample any children who report never to have enrolled in preschool or school. Our final sample comprises more than half a million children with information on preschool attendance and at least one of the two main outcomes described above. Note that because a few children are observed with only one of these outcomes, observation numbers in regression tables vary slightly.[8]

References

Funk, C., P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, and J. Michaelsen. 2015. "The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes." *Scientific Data* 2:150066.

Henderson, J.V., A. Storeygard, and D. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102:944–1028.

Schady, N., J.R. Behrman, M. Caridad Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, K. Macours, D. Marshall, C. Paxson, and R. Vakis. 2015. "Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries." *Journal of Human Resources* 50:446–463.

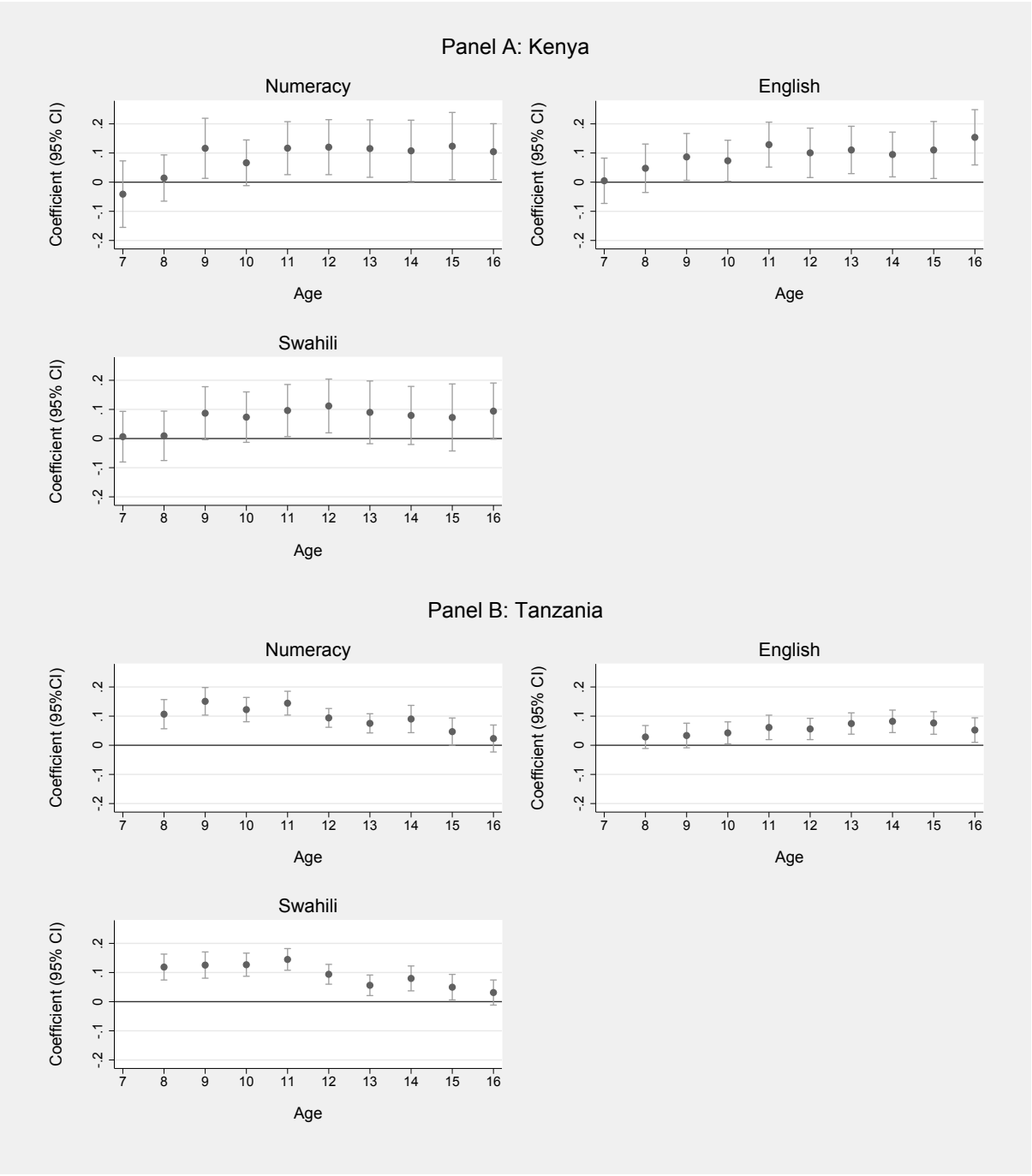
Shah, M., and B.M. Steinberg. 2017. "Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital." *Journal of Political Economy* 125:527–561.

[8] All results are robust to focusing on a slightly smaller sample of children observed with both outcomes.

Online Appendix B: Figures and Tables

Online Appendix Figure B.1

Preschool attendance and numeracy, English, and Swahili skills, by age



Notes: The figure plots coefficient estimates and 95% confidence intervals from regressions of numeracy, English, and Swahili scores on preschool attendance. Scores are standardized by country and age to have mean zero and standard deviation one. The indicator for preschool attendance is interacted with age dummies, and the figure shows the estimated effect of preschool attendance separately for each age. Specifications are otherwise equal to the household fixed effects regressions reported in column 4 of Table 4.

Online Appendix Table B.1

Further robustness checks

	Only siblings born ≤5 years apart		No sampling weights	
	Highest grade	Composite score	Highest grade	Composite score
	(1)	(2)	(3)	(4)
	Panel A: Kenya			
Attended preschool				
7-9 years old	-0.257***	0.067	-0.377***	0.033
	(0.064)	(0.052)	(0.038)	(0.034)
10-12 years old	-0.083	0.136***	-0.146***	0.124***
	(0.068)	(0.051)	(0.034)	(0.035)
13-16 years old	-0.038	0.123**	-0.026	0.148***
	(0.076)	(0.054)	(0.042)	(0.039)
Observations	158,132	158,408	218,728	218,134
	Panel B: Tanzania			
Attended preschool				
8-9 years old	-0.207***	0.101***	-0.207***	0.103***
	(0.027)	(0.021)	(0.024)	(0.014)
10-12 years old	-0.081**	0.120***	-0.074***	0.112***
	(0.031)	(0.016)	(0.022)	(0.012)
13-16 years old	0.058*	0.077***	0.053**	0.058***
	(0.032)	(0.018)	(0.024)	(0.013)
Observations	235,967	238,852	284,396	288,084

Notes: In columns 1 and 2, the sample is restricted to families with children born at most 5 years apart. Columns 3 and 4 report estimates from regressions that do not use the sampling weights provided with the data. Standard errors in parentheses are clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.