RESEARCH ON IMPROVING SYSTEMS OF EDUCATION

WORKING PAPER
February 2019

Learning Equity Requires More than Equality: Learning Goals and Achievement Gaps between the Rich and the Poor in Five Developing Countries

Maryam Akmal and Lant Pritchett

RISE-WP-19/028

Funded by:

Learning Equity Requires More than Equality: Learning Goals and Achievement Gaps between the Rich and the Poor in Five Developing Countries

Maryam Akmal* Lant Pritchett†

February 27, 2019‡

The Sustainable Development Goals (SDGs) for education include the goal that "all youth...achieve literacy and numeracy" (Target 4.6). Achieving some absolute standard of learning for *all* children is a key element of global equity in education. Using the Annual Status of Education Report (ASER) data from India and Pakistan, and Uwezo data from Kenya, Tanzania, and Uganda that test all children of given ages, whether in school or not, on simple measures of learning in math, reading (local language), and English, we quantify the role of achieving *equality* between the richest 20% and the poorest 40% in terms of grade attainment and learning achievement toward accomplishing the global *equity* goal of universal numeracy and literacy for all children. First, excluding Kenya, equalizing grade attainment between children from rich and poor households would only close between 8% (India) and 25% (Pakistan) of the gap to universal numeracy, and between 8% (Uganda) and 28% (Pakistan) of the gap to universal literacy. Second, children from the poorest 40% of households tend to have lower performance in literacy and numeracy at each grade. If such children had the learning profiles of children from rich households, we would close between 16% (Pakistan and Uganda) and 34% (India) of the gap to universal numeracy, and between 13% (Uganda) and 44% (India) of the gap to universal literacy. This shows that the "hidden exclusion" (WDR, 2018) of lower learning at the same grade levels—a gap that emerges in the earliest grades—is a substantial and often larger part of the equity gap compared to the more widely documented gaps in enrollment and grade attainment. Third, even with complete equality in grade attainment and learning achievement, children from poor households would be far from the equity goal of universal numeracy and literacy, as even children from the richest 20% of households are far from universal mastery of basic reading and math by ages 12-13. Achieving universal literacy and numeracy to accomplish even a minimal standard of global absolute equity will require more than just closing the rich-poor learning gap, it will take progress in learning for all.

Keywords: Learning Assessments, Education Quality, Human Capital, Equity, Wealth

JEL Classification Numbers: I25, J24, O15, O53, D31

*Center for Global Development. Akmal: makmal@cgdev.org

†Blavatnik School of Government, University of Oxford. Pritchett: lant_pritchett@harvard.edu

‡The views expressed are those of the authors and should not be attributed to the board of directors or funders of the Center for Global Development. All errors are ours.

Contents

A self-ordained professor's tongue too serious to fool

Spouted out that liberty is just equality in school

"Equality," I spoke the word as if a wedding vow

Ah, but I was so much older then, I'm younger than that now

Bob Dylan, My Back Pages

1 Introduction

The enrollment drive of the past few decades has been enormously successful at getting children to school. In many countries, the beneficiaries of enrollment efforts include traditionally marginalized groups such as the poor, girls, and rural dwellers. In India, the fraction of girls aged 12-13 years from the poorest 40% of households enrolled in school increased from 32.0% to 86.5% from 1992/1993 to 2015/2016. In Uganda, there was a similar increase from 55.2% in 1995 to 87.4% in 2011.[1] However, millions of children remain out of school and it is important to get children to school by overcoming inequities in enrollment and grade attainment across household income, parental characteristics, gender, ethnicity, geographical remoteness, and disability. But merely enrolling children in schools is not sufficient to achieve SDG Goal 4: "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all," or SDG Target 4.6: "By 2030, ensure that all youth...achieve literacy and numeracy." Tragically, many of the world's youth who lack the basic skills of literacy and numeracy can be found where one would least expect to find them—inside the classroom. Whelan (2014) calculated that 96% of children around the world receive some schooling, but only 37% achieve basic learning by the end of primary school. The remaining 63% of children go through an extended period of schooling only to remain illiterate and innumerate. Spaull and Taylor (2015), using data from Southern and East Africa Consortium for Monitoring Educational Quality (SACMEQ), show that 53% of Ugandan children were innumerate at age 12 but only 4% had never enrolled in school, 14% had enrolled in Grade

[1] Data downloaded from World Bank website on July 9, 2018.

1 but dropped out before age 12, and 33% of all children aged 12 had completed Grade 6 but were innumerate. The exclusion of most children from achieving competence in literacy and numeracy is an "equity crisis" on a global scale (Crouch and Rolleston, 2017).

A key question is how much a drive for equality in outcomes across groups within a country helps the disadvantaged in that country move towards a global standard of equity based on a minimal acceptable absolute level of learning achievement. One of the best documented facts about schooling around the world is that grade attainment is lower for children from poorer households.[2] Research by Das (2018) using longitudinal survey data from Punjab, Pakistan shows that children from households with lower socio-economic status (SES) drop out at higher rates than children from households with higher SES. Also well-documented is the fact that assessed cognitive skills at any given grade/age are lower for children from households with lower SES.[3] In regression analyses of the learning of children in the same grade, a measure of household SES is often the single biggest factor explaining learning differences (Hanushek and Woessman, 2011). This association between a child's learning achievement and household SES can be causally mediated by many factors, such as poor nutrition (Alderman and Bundy, 2011), parental education (Dubow et al., 2009) and attention (Davis-Kean, 2005), and stress factors (Lupien et al., 2000). Analysis by Paxson and Schady (2005) in Ecuador shows that children from wealthier households and more educated parents have higher test scores. This association grows stronger as children grow older, implying that there is an increasing gap in test scores between children from rich and poor households with age. Furthermore, positive sorting between households and schools can further aggravate inequalities: richer households are able to select better schools (Anand et al., 2018). Equality of opportunity in terms of both access and learning is clearly a major concern for education systems.

[2]See, among many other sources, Filmer and Pritchett (2001) and the online cross-nationally comparable data for over 50 countries on enrollment and grade attainment by household wealth on the World Bank website.

[3]See, among many other sources, the household SES differences reported in PISA (2015) and WDR (2018).

Yet assessments across the developing world show that absolute learning levels are low across the board, for both children from rich and poor households. When two states of India, Himachal Pradesh and Tamil Nadu, participated in PISA (2009), the 95$^{\text{th}}$ percentile score on mathematics for the two states was only 463, which implies that the "elite" (95$^{\text{th}}$ percentile) were well behind the constructed OECD norm of 500. According to ASER (2015), by age 12, half of the children from rich households can do basic division, which is 20 percentage points more than the figure for children from poor households, but still 50 percentage points away from universal mastery of a basic arithmetic operation. If even children from richer households have low learning levels, then raising the learning of disadvantaged children to that of the privileged may still leave them well short of absolute learning levels needed for global equity. Hence, equity will require more than equalization of the poor to the level of the rich within each country.

We use ASER data for India and Pakistan, and Uwezo data for Kenya, Tanzania, and Uganda as this data has four key features: (1) all children aged 5 to 16 are in the sampling frame, not just those enrolled in school or those in a given grade,[4] [5] (2) all children are given a similar assessment, (3) the child's highest grade of enrollment is reported, and (4) there is data on assets that can be used to construct a proxy for the wealth of each child's household. With these four features, we can construct the grade attainment profile for the cohort of children aged 12 or 13 (we combine the two ages to ensure large sample sizes for all countries) dis-aggregated by wealth. We can also construct a descriptive grade-based learning profile by wealth to show, for instance, what fraction of children in Grade 7 can read a second grade story from the top 20% of households versus bottom 40% of households.

We can then quantify how much learning changes under various counter-factual scenarios:

- How much would the likelihood that a child from a poor household is literate (defined as the ability to read a second grade story) or numerate (defined as the ability to do

[4]ASER India covers all rural districts, while ASER Pakistan covers all rural and some urban districts. Uwezo covers all districts, rural and urban, in Kenya, Tanzania, and Uganda.

[5]Uwezo tends to test children between the ages of 6 and 16.

simple division) change if they had the same grade attainment as a child from a rich household while keeping their existing learning profile?

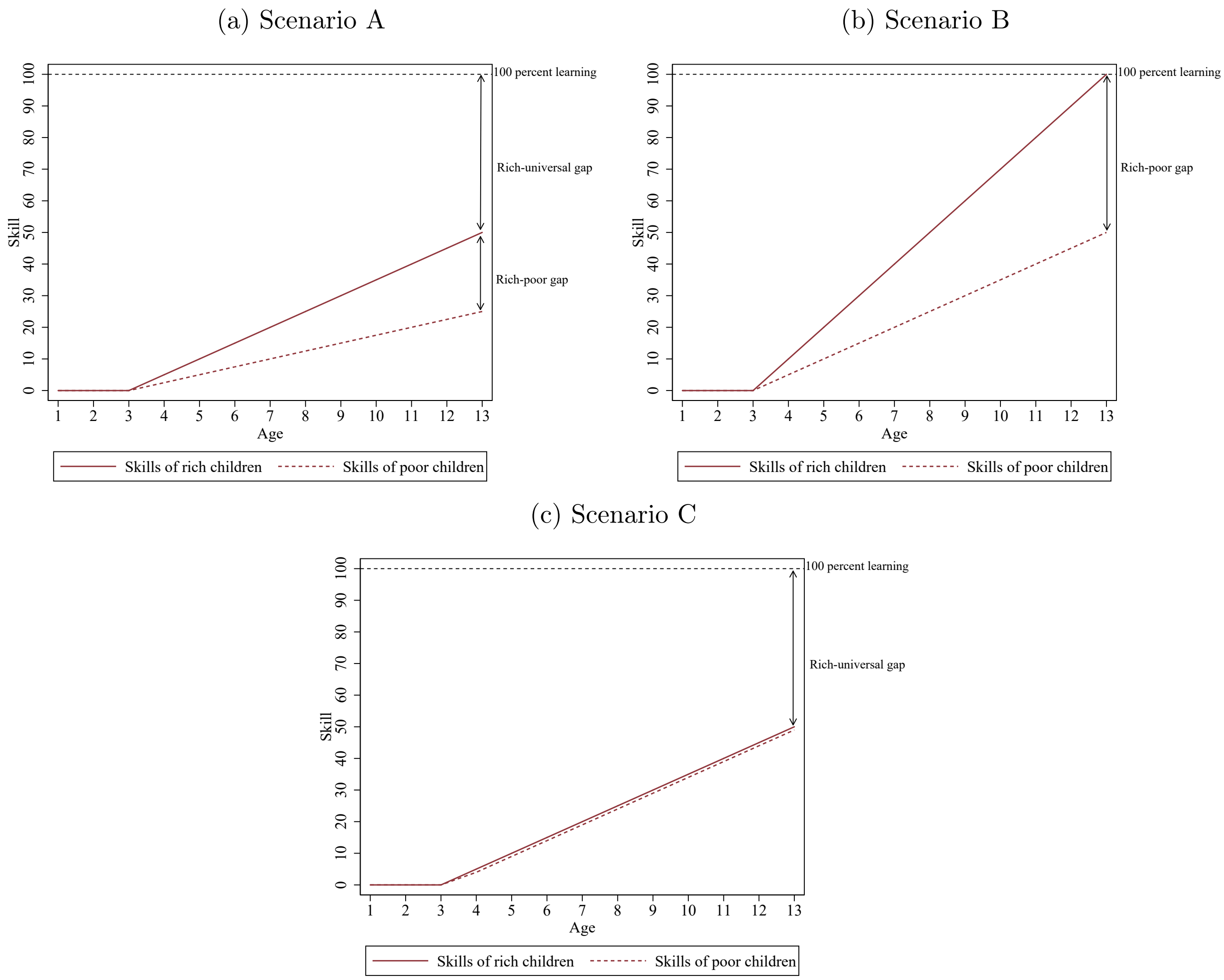
- How much would the likelihood that a child from a poor household is literate or numerate change if they had the learning profile of a child from a rich household while keeping their existing grade attainment profile?
- If children from poor households had exactly the same learning *and* grade attainment as children from rich households, how far would they be from achieving universal literacy or numeracy—how much of the gap to the global equity goal of universal literacy and numeracy will be erased if there was complete within-country equality across the two asset groups?

1.1 Conceptual Framework

The gap between current learning levels and the goal of 100% literacy and numeracy can be decomposed into the rich-poor gap and the rich-universal gap. The rich-poor gap illustrates how far behind children from poor households are in terms of mastery of basic literacy and numeracy compared to children from rich households of similar age, while the rich-universal gap illustrates how far children from rich households themselves are from the goal of 100% literacy and numeracy. Together, the rich-poor gap and the rich-universal gap represent the total difference between current learning levels of children from poor households and a global equity goal of universal literacy/numeracy. While bridging the rich-poor learning gap may bring us closer to achieving universal literacy and numeracy in some countries, in others it will still leave a long way to go.

Figure 1 illustrates the decomposition of the total learning gap into the rich-poor gap and the rich-universal gap. Scenario A in Figure 1 depicts countries where there is a substantial learning gap between children from rich and poor households at age 13, but children from rich households are also far from attaining universal literacy and numeracy. In this case, bridging the rich-poor learning gap helps make the overall learning gap smaller but still

Figure 1: Conceptual Framework



takes us nowhere close to the goal of all children having mastery over basic literacy and numeracy. Scenario B in Figure 1 depicts a case where all children from rich households are literate and numerate by age 13 but the children from poor households still lag far behind. In this case, closing the rich-poor gap significantly helps in bringing us closer to universal literacy/numeracy. Lastly, Scenario C in Figure 1 depicts a case where children from rich and poor households have the same level of learning, implying no wealth gap in learning, but both are far from achieving universal literacy and numeracy. In this case, where children from rich and poor households have similar levels of learning, both wealth groups need to move closer to the goal of universal learning achievement.

2 Data and Methodology

ASER (meaning "impact"), an annual household-based survey of basic learning of children, has been carried out in India and Pakistan. Uwezo (meaning "capability"), an ASER-like survey in Africa, has been carried out in Kenya, Uganda, and Tanzania. We use data for all five countries for all available years to construct a unified ASER-Uwezo data set. The sampling frame for these data includes all children, both enrolled and non-enrolled, between the ages of 5 and 16. The data contain a learning measure, current (or highest) grade, and asset indicators for the household. Most internationally comparable data on learning outcomes (e.g., TIMSS, PIRLS, LLECE, SACMEQ, PASEC, etc.) sample schools. Hence, they only assess enrolled children in a particular grade (or age group) only. With the ASER-Uwezo data, we can estimate learning of an entire age cohort.

2.1 Sample

Our data set contains grade attainment information for approximately 5.7 million children spanning five countries using all available ASER and Uwezo survey years. Table 1 shows the distribution of the highest grade attained for children aged 5-16 years.

Table 1: Percentage of Children with the Relevant Grade as the Highest Grade Attained

Country	0	1	2	3	4	5	6	7	8
India	1.76	11.73	10.94	10.86	10.40	11.31	9.56	8.94	8.89
Kenya	16.42	12.10	11.53	10.96	11.10	9.73	9.30	8.45	6.46
Pakistan	25.02	10.89	11.51	10.46	8.95	9.55	6.10	5.07	5.08
Tanzania	12.27	13.95	13.13	12.23	11.61	10.51	10.59	9.02	2.99
Uganda	9.90	18.90	14.63	14.17	14.00	11.38	8.53	4.99	1.64

Of the total children, math test results are available for 5.2 million children, local language reading results are available for 4.8 million children, and English reading results are available for 3.7 million children. Table 2 shows the distribution of children who took at least one test across the five countries.

Table 2: Number of Survey Years and Total Assessed Children

Country	Years	Number of Children Tested in One or More Subjects
India	2009-2014	2.9 million
Kenya	2009, 2011-2015	0.7 million
Pakistan	2012-2015	0.8 million
Tanzania	2010-2015	0.5 million
Uganda	2010-2015	0.4 million

2.2 Learning Measure

The ASER style assessment is meant to be extremely simple to implement and therefore is also very crude. For reading, there is a single card (in the child's preferred language) that contains letters, words, a short sentence (Grade 1 level), and a short paragraph (Grade 2 level).[6] Each child's performance is categorically coded by the highest level they are comfortable doing: "nothing" is level 1, "recognize letters" is level 2, "read words" is level 3, "read sentence" is level 4, and "read Grade 2 paragraph" is level 5.[7] In this paper we just use the binary indicator for "level 5." Similarly, the math assessment is a single card with one-digit numbers, two-digit numbers, subtraction problems of two-digit numbers (requiring "carry"), and division problems of dividing a one-digit number into a three-digit number with a remainder (e.g., 824/6, 517/4).[8] Again this is categorically coded and we use just

[6] ASER data for India for years 2010, 2011, and 2013 do not contain information about reading in English. However, the "local" language test was sometimes administered in English. We classify such instances as actually testing literacy in English. Uwezo data for Uganda for years 2010 and 2011 do not contain any information about testing literacy in the local language.

[7] ASER data for India and Pakistan codes literacy in English slightly differently from literacy in local language: "nothing" is level 1, "recognize capital letters" is level 2, "recognize small letters" is level 3, "read words" is level 4, and "read sentences" is level 5. Uwezo data from Kenya, Tanzania, and Uganda codes literacy in local language and English the same way.

[8] Uwezo tests dividing a one-digit number into a two-digit number in Kenya and Uganda.

"level 5" as our definition of numeracy.[9]

A forthcoming paper by Patel and Sandefur (2018) creates links between various assessments such as TIMSS and PIRLS (which are normed to a mean of 500 and standard deviation of 100 for a reference group of children) and PASEC, LLECE, and ASER by asking a sample of children in Bihar to sit a test that includes items from the various assessments using the Non-Equivalent Groups with Anchor Test (NEAT) approach (Davier et al., 2004). The authors found that it was difficult to estimate precisely a concordance of ASER to international assessments as the top-coded category of ASER is near the bottom of these assessments. According to the preliminary results, the distribution of TIMSS scores of children with an ASER level of 5 on math would have a mean of 382 and a standard deviation of 75, which is large since 5 is the top-coded level. Hence, all levels of performance above the threshold of "do division" are included in this category. Similarly, an ASER level 5 on reading is equivalent by NEAT to a mean on PIRLS of 378 with a standard deviation of 62. This means that children that are just reaching level 5 in ASER are near the bottom of the distribution of measured performance among children in the OECD. For reference, the score cut-off for minimum level of proficiency on math for TIMSS is 475,[10] well above the TIMSS-equated score of 382 for ASER level 5. It is evident that the "top-coded" level of ASER is a low bar of functional literacy and numeracy and *understates* the true gap to mastery of minimum proficiency in math and reading.

[9]Uwezo tests are mostly coded on a 1-5 scale. However, for certain years and subjects in Kenya, Tanzania, and Uganda, Uwezo tests are coded on a 1-7 or a 1-9 scale. In order to facilitate comparisons across countries and different years, we ensure that the highest order skill measured denotes literacy and/or numeracy. For numeracy, the highest order skill tends to be division, and for literacy, it tends to be the ability to read a simple story. More details about how we re-calibrate those tests to align with the standard 1-5 scale used for ASER can be found in Table A.1 in the appendix. While streamlining the ASER and Uwezo scores does not allow perfect comparability across countries (for example, in Tanzania, multiplication rather than division tends to be highest skill measured on math tests), the conversions ensure that the highest order skill measured by each test gets assigned a 5. Hence, we classify as literate or numerate any individual who is in the top-coded category for that country.

[10]See TIMSS benchmarks on the IEA website.

2.3 Construction of Wealth Index

We create an asset index as a proxy for household wealth by using principal components analysis (PCA) on asset ownership data at the household level, as done in Filmer and Pritchett (2001). The combined data set includes information about 17 asset indicators, but different countries have varying levels of information available about each asset. We run PCA separately for each country, using only those asset indicators that have fewer than 11% missing values. For example, Table 3 shows the scoring factors from PCA for the five asset variables used for India. The wealth ranks produce sharp separation in India: almost everyone in the top 20% wealth group has access to electricity and mobile phones compared to less than half of the population in the bottom 40%. Similarly, almost everyone in the top 20% wealth group has access to a TV and a toilet compared to less than 3% and 11% of the households in the bottom 40%. The PCA tables for all the remaining countries can be found in the appendix in Section 5.[11]

Table 3: Scoring Factors and Summary Statistics for Variables Entering the Computation of the First Principal Component: India

Asset	Scoring Factors	Mean	SD	Scoring Factors X SD	Mean Poorest 40%	Mean Middle 40%	Mean Top 20%
Electricity Available	0.45	0.74	0.44	0.20	0.44	0.96	1.00
Mobile Available	0.40	0.68	0.47	0.19	0.43	0.82	1.00
Solid House	0.40	0.35	0.48	0.19	0.09	0.36	0.95
TV Available	0.53	0.50	0.50	0.26	0.03	0.79	1.00
Toilet Available	0.45	0.42	0.49	0.22	0.11	0.51	1.00

Note: Each variable takes the value 1 if true, 0 otherwise. Scoring factor is the "weight" assigned to each variable in the linear combination of the variables that constitute the first principal component.

[11] The data for Uganda (2011) contained a few observations with different assets for children in the same household. This problem affected a minor number of observations for asset variables denoting access to electricity, TV, mobile, radio, bicycle, bike, and water. To resolve this data anomaly, if one of the children in the same household is assigned as having a certain asset, we assume all the remaining children in the household also have access to that particular asset. This problem was present for fewer than 4% of the observations for each affected asset variable in the Uganda (2011) sample.

We use the household asset index score to give a *unique* rank to each household.[12] We then use the unique rank values to divide the data into deciles. Next, we assign households to a wealth group based on the value of their decile rank, for example, the top two deciles comprise the top 20%, the middle four comprise the middle 40%, and the bottom four comprise the bottom 40%. As done in Filmer and Pritchett (2001), we refer to the bottom 40% as "poor," the middle 40% as "middle," and the top 20% as "rich."[13] Finally, we assign a wealth status to each child by using their respective household's wealth status. As argued in Filmer and Pritchett (2001), this wealth index is more stable than current consumption expenditure or income measures of household SES. Hence, it is more reliably predictive of longer term phenomena like child education. Table 4 shows the distribution of children, including those who take the test and those who do not, across the different wealth ranks.

Table 4: Percentage of All Sampled Children by Household Wealth Category

Country	Bottom 40% of Households	Middle 40% of Households	Top 20% of Households
India	44.25	38.64	17.11
Kenya	42.52	40.53	16.94
Pakistan	41.53	39.66	18.80
Tanzania	39.59	40.96	19.44
Uganda	38.37	41.32	20.31

[12] While assigning unique values is potentially problematic because households with the same asset index may be assigned different rank values, unique values allow creating wealth groups such that the top 20% wealth group has approximately 20% of the population, the middle 40% has approximately 40% of the population, etc. Particularly when there are few (e.g., only five) assets, there will be a large number of ties for household asset index score. This makes it impossible to create such proportional cut-offs without assigning unique values. This is innocuous for our analysis as (1) since the households that were tied were indistinguishable, any assignment of households into the the two adjacent categories (e.g., bottom 40 versus middle 40) would have produced the same expected value of results, and (2) since we are only comparing poorest 40% and richest 20%, there were no ties across those categories.

[13] By the "rich," we deliberately do not exclusively refer to the "elite" (for example, the top 0.1% or 1%) in these countries who are a statistical minority. We are concerned with the schooling and learning opportunities for the top 20%, which encompasses a statistically larger group of wealthiest households in these countries. However, we do conduct sensitivity tests in Section D of the appendix and find that changing the cut-off for the "rich" from the top 20% to 15% or 10% or 5% does not substantially alter the percent literate or numerate among the "rich" at each grade level.

3 Learning Trajectories of the Rich and the Poor by Age

Using the ASER-Uwezo data that samples all children of a given age, we can examine *learning trajectories* by age and household wealth—the likelihood that children of a given age and SES are literate or numerate. In the next section we decompose these wealth gaps in literacy and numeracy by age into grade attainment (how much behind "grade for age" are children from poor households) and into a learning profile by grade (how likely are children from poor households in each grade to be less literate or numerate). The final section computes the counter-factual calculations of how much of the gap from universal attainment of literacy or numeracy for children from poor households can be erased if they achieve the same grade progression and learning profile as children from rich households.

Most national or international assessments of learning outcomes are school-based samples testing children in a given grade or children of a certain age. It is impossible to use them to create aged-based learning trajectories or grade-based learning profiles (except for the small number of grades covered by sampling in-school children by age). ASER and Uwezo data allow the creation of learning profiles showing proficiency at each age as well as at each grade level (Pritchett and Sandefur, 2017).

3.1 Learning Trajectories by Age and Wealth: Math

Figures 2 shows the fraction of children in each wealth group who are numerate, that is, able to solve a simple division problem. Not surprisingly, as division is a relatively late curricular concept, the learning trajectory gaps only begin to emerge after age 7, as essentially none of the children can do division at ages 5 or 6.[14] This shows that we are not measuring some notion of an underlying and relatively fixed child potential or "intelligence" but realized learning, which is an interaction of potential, preparedness through early exposure, effort, and household (maternal, paternal, sibling, etc.) support—all interacting with formal instruction and its quality. However, the gaps in numeracy do emerge, become very large by ages 12

[14] Uwezo tends to test children between the ages of 6 and 16.

and 13, and persist as the children get older.[15]

Table 5 shows the gap in numeracy at age 12 between children from the richest and poorest households. This table demonstrates two key features of inequality across wealth groups within a country but also across countries. First, in all countries, the gap in numeracy at age 12 between children from the richest 20% and poorest 40% of households is equal to or over 20 percentage points, ranging from 20 (Uganda) to 27 (India) percentage points. By age 12, the percentage of Indian children from rich households who can do basic math is roughly twice that of the children from poorer households. This finding is similar to a study by Alcott and Rose (2015) who show that wealth gaps in learning have substantial magnitude.

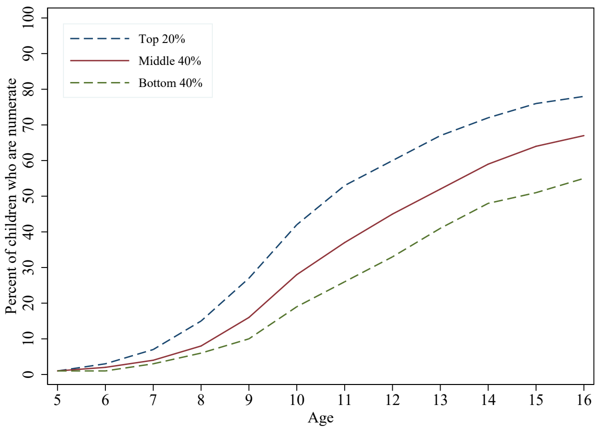
The second feature apparent in Table 5 is that the absolute level of numeracy among the rich is typically far from universal and varies a great deal across countries. In Kenya, 79% of children from rich households can do division, but in India and Pakistan only 60% can do so. In Uganda, the figure is even lower at 53%. This finding implies that if we think global equity requires achieving a minimal level of numeracy for all children, even children from the richest quintile in Pakistan and Uganda, and rural India, are only just past half-way to this equity goal.

It could be argued that ASER and Uwezo top out at a minimal level of content mastery so in principle there could be children who could have scored at a higher level. While ASER is top-coded and reflects a very minimum level of mastery, the fact that even a significant proportion of older children from rich households (see Figure 2) can not master basic numeracy shows that it is unlikely that such children constitute more than a tiny

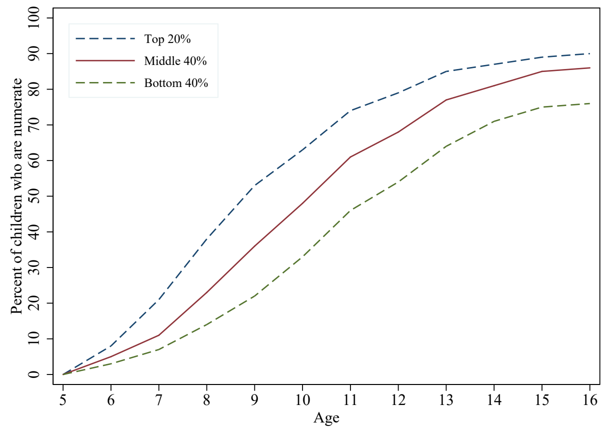
[15]Measuring gaps across groups robustly is extremely technically demanding in any case as even Item Response Theory (IRT) measures produce scores that appear, and are often treated as if they were, cardinal. In fact, IRT scores across students (and therefore, across groups) are only unique up to a monotone transformation (Ho, 2016), and are more properly treated as ordinal, as any set of numbers that preserves the relative rankings of students represents equally well their performance on a set of questions. Our results, which essentially use a single binary-coded question, "Can a child do division?" (and similarly for the measures of local language literacy and English: "Can a child read a Grade 2 paragraph?") are only valid to the extent that we take these particular questions to normatively represent a defensible standard of minimal mastery.

Figure 2: Learning Trajectories by Age and Household Wealth: Math

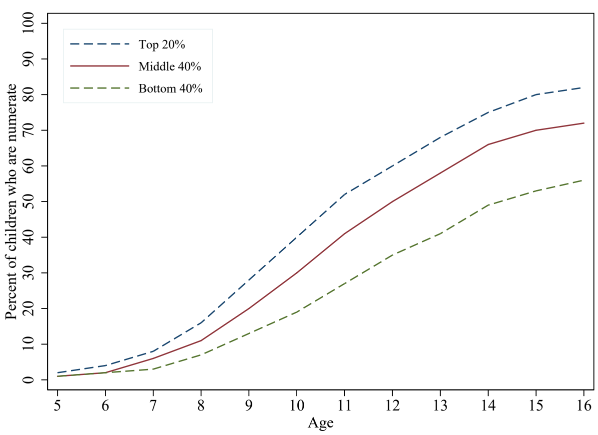
(a) India



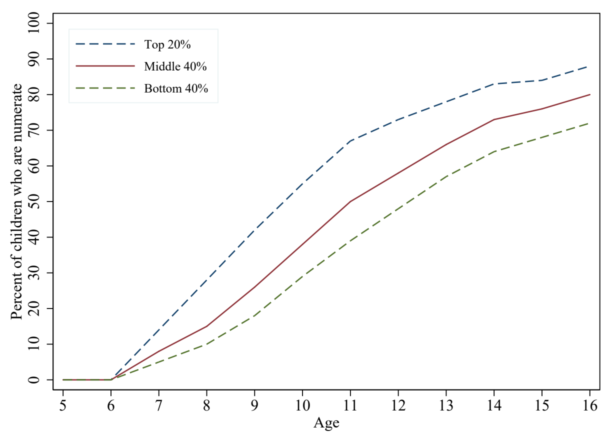
(b) Kenya



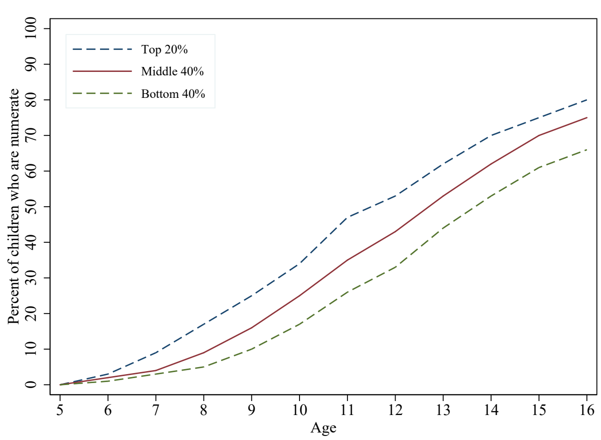
(c) Pakistan



(d) Tanzania



(e) Uganda



percentage at each age.

Table 5: Mastery Gap Between the Richest and Poorest Children at Age 12: Math

Country	Numeracy for Richest 20% (%)	Numeracy for Poorest 40% (%)	Rich-Poor Mastery Gap (% points)	Rich-Universal Mastery Gap (% points)
India	60	33	27	40
Kenya	79	54	25	21
Pakistan	60	35	25	40
Tanzania	73	48	25	27
Uganda	53	33	20	47

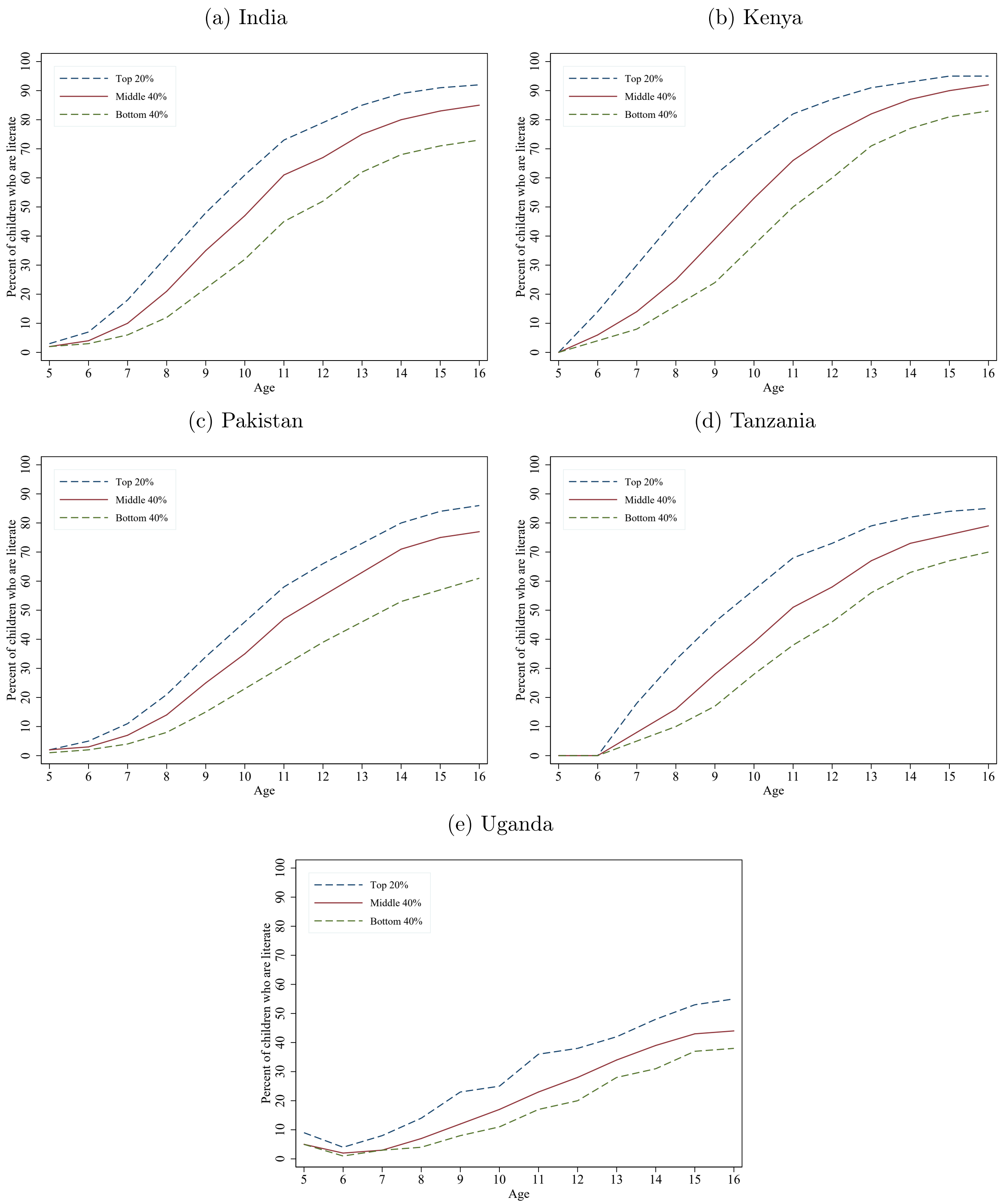
3.2 Learning Trajectories by Age and Wealth: Local Language

Figure 3 shows the learning trajectory by age for literacy, defined as reading a Grade 2 paragraph in the local language, across the five countries. The results are similar to numeracy, with very small differences at very young ages, followed by gaps emerging by ages 7-8, and leading to large gaps across wealth groups by ages 12-13 that persist as the children get older.

Table 6: Mastery Gap Between the Richest and Poorest Children at Age 12: Local Language

Country	Literacy for Richest 20% (%)	Literacy for Poorest 40% (%)	Rich-Poor Mastery Gap (% points)	Rich-Universal Mastery Gap (% points)
India	79	52	27	21
Kenya	87	60	27	13
Pakistan	66	39	27	34
Tanzania	73	46	27	27
Uganda	38	20	18	62

Figure 3: Learning Trajectories by Age and Household Wealth: Local Language

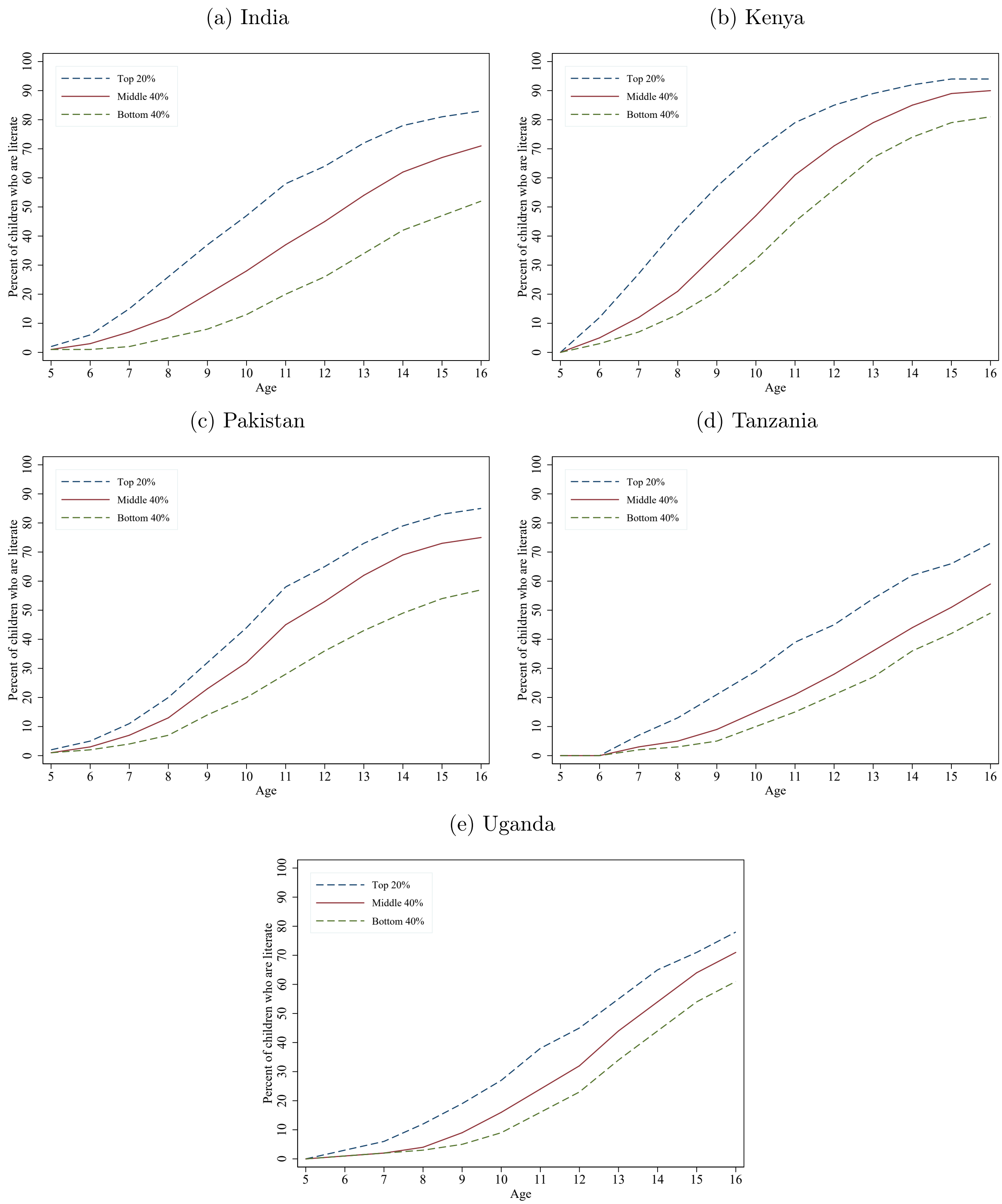


3.3 Learning Trajectory by Age and Wealth: English

While mastery of English is not fundamental like numeracy and literacy, we also show those results as all five countries do assess mastery of English. Figure 4 shows differences in learning emerge by ages 7-8 and grow by ages 12-13. The gap in literacy between children from rich and poor households tends to be above 20 percentage points, except in India where it is almost double that.[16]

[16]Kenya tends to do better on English than Uganda and Tanzania, which could be reflective of the fact that in Kenya English tends to be the medium of instruction as early as Grade 1 (UNICEF, 2016a). In Tanzania, Kiswahili is the primary language of instruction (Bashir et al., 2018). In Uganda, English is the language of instruction from Grade 4 onward (UNICEF, 2016b).

Figure 4: Learning Trajectories by Age and Household Wealth: English



(a) India

(b) Kenya

(c) Pakistan

(d) Tanzania

(e) Uganda

Table 7: Mastery Gap Between the Richest and Poorest Children at Age 12: English

Country	Literacy for Richest 20% (%)	Literacy for Poorest 40% (%)	Rich-Poor Mastery Gap (% points)	Rich-Universal Mastery Gap (% points)
India	64	26	38	36
Kenya	85	56	29	15
Pakistan	65	36	29	35
Tanzania	45	21	24	55
Uganda	45	23	22	55

4 Decomposing the Learning Trajectory by Age into Grade Attainment and Grade-Based Learning Profiles

The learning gap between children from rich and poor households at each age can be decomposed into differences in grade attainment and differences in learning achievement by grade (the *descriptive* learning profile).[17] The grade attainment profile is the fraction of children of a given age in any given group (where a group could denote rich/middle/poor, girl/boy, urban/rural, maternal/paternal education, state/region, etc.) who have completed a particular grade. The *descriptive grade-based learning profile* is the share of children with a particular grade attainment who are literate or numerate. For any age cohort and for any given group, the fraction of the group who are literate is just the grade attainment weighted average of the grade-based learning profile. Therefore, mechanically, a group could have higher literacy because either (1) the group has higher grade attainment—more of the children have completed a higher grade, or (2) the group has a steeper learning profile so that a child from one group is more likely to be literate in any given grade, or (3) both (Pritchett and Sandefur, 2017). Not surprisingly, the two approaches to closing a learning

[17]This is not an exact decomposition, as it potentially leaves an interaction term.

gap encompass: (1) policies that expand the grade attainment of the lagging group (e.g., scholarships, conditional cash transfers, etc.) or policies of automatic promotion that do not address the learning profiles, or (2) policies that address the learning profiles (e.g., teaching at the right level, etc.) and basic ways to meet a learning goal. In short, there are two ways to close a learning gap. One way is to increase grade attainment so that more kids get to higher grades. The second way is to increase learning per grade, or to steepen the learning profile (Pritchett, 2013).

We look at the grade attainment and learning achievement profiles for children aged 12-13. While grade attainment profiles have generally improved over time because of higher enrollments, there is no reason to assume that learning profiles have improved as well. In fact, ASER results from India in 2014 show that learning profiles have worsened. From 2010-2014, the percentage of Grade 5 students who could read a simple story fell from 54% to 48%, and the percentage of Grade 5 students who could do a simple division problem fell from 36% to 26% (ASER, 2014).

Using simple equations, we can decompose the learning trajectory by age into grade attainment and learning achievement per grade. The fraction of a group that is literate or numerate can be calculated using the following equation:

$$\text{Fraction Literate/Numerate}_{group} = \sum_{g=0}^{g=8} \alpha^{g}_{group} * \mathrm{s}^{g}_{group} \tag{1}$$

where α^{g}_{group} is the share of children aged 12-13 from any given group who took the test with grade g as their highest grade attained, and s^{g}_{group} is the share of children aged 12-13 with grade g in the group who are literate/numerate.

The share of children from the poorest 40% of households who are literate/numerate is given by the following equation:

$$\text{Fraction Literate/Numerate}_{poorest\,40} = \sum_{g=0}^{g=8} \alpha^{g}_{poorest\,40} * \mathrm{s}^{g}_{poorest\,40} \tag{2}$$

The share of children from the richest 20% of households who are literate/numerate is given by the following equation:

$$\text{Fraction Literate/Numerate}_{richest\,20} = \sum_{g=0}^{g=8} \alpha^{g}_{richest\,20} * \mathrm{s}^{g}_{richest\,20} \tag{3}$$

Using these simple equations we can calculate various hypothetical scenarios of: (1) equal grade attainment, (2) equal learning achievement, or (3) both. In the "equal grade attainment" scenario, we calculate what the learning levels of children from the poorest 40% of households would be if they had the same grade attainment as children from the richest 20% of households but still had their own existing learning profiles.

$$\text{Fraction Literate/Numerate}^{grade\,attainment\,of\,richest\,20}_{poorest\,40} = \sum_{g=0}^{g=8} \alpha^{g}_{richest\,20} * \mathrm{s}^{g}_{poorest\,40} \tag{4}$$

Alternatively, we can calculate how much higher literacy/numeracy would be for children from the poorest households if they retained their exisiting grade attainment profiles but had the grade-based learning profiles of children from the richest 20% of households:

$$\text{Fraction Literate/Numerate}^{learning\,profile\,of\,richest\,20}_{poorest\,40} = \sum_{g=0}^{g=8} \alpha^{g}_{poorest\,40} * \mathrm{s}^{g}_{richest\,20} \tag{5}$$

The gain in literacy from the improvement in learning profiles of children from poor households is simply the difference between Equation 5 and Equation 2.

$$\Delta Learning = \sum_{g=0}^{g=8} \alpha^{g}_{poorest\,40} * (\mathrm{s}^{g}_{richest\,20} - \mathrm{s}^{g}_{poorest\,40}) \tag{6}$$

Equation 6 shows that the gain in learning from all children from poor households achieving the learning profiles of children from rich households is bigger the larger the share of children from poor households at each grade level, and bigger the gap in literacy between children from rich and poor households. Obviously, if children from the poorest 40% of households had the same grade attainment and the same grade-based learning profile as children from the richest 20% of households, then complete equality in literacy/numeracy between children from rich and poor households would be achieved.

4.1 Grade Attainment Profiles

Figure 5 shows the grade attainment profiles for 12-13 year old children using the highest grade attained for both those still in school and those who have dropped out, dis-aggregated by wealth.[18] If children were to start at age 6 and progress one grade per year, one would expect most 12-13 year old children to be in Grades 7-8. But we know from previous analysis of enrollment and attainment profiles that there is substantial late enrollment, grade repetition, and drop-out. Hence, there is a spread in grade attainment. For the African countries, we see more bunching in the middle (Grades 4-6), which is particularly pronounced for Uganda. However, one can see that children from rich households are more likely to have reached the "age-appropriate" Grades 7-8. For example, in Tanzania, 17% of 12-13 year old children from rich households are in Grade 7 compared to 9% of those from poor households. Note that in India, a substantially higher proportion of 12-13 year old children are in the "age-appropriate" Grades 7-8, almost certainly reflecting policies of automatic promotion. In addition, note that the proportion of "never enrolled" is higher for children from poor households, but above 10% only in Pakistan. Therefore, the grade deficit for children from poor households is mostly late enrollment and lower grade progression rather than the fact these children never enroll in school due to lack of access.

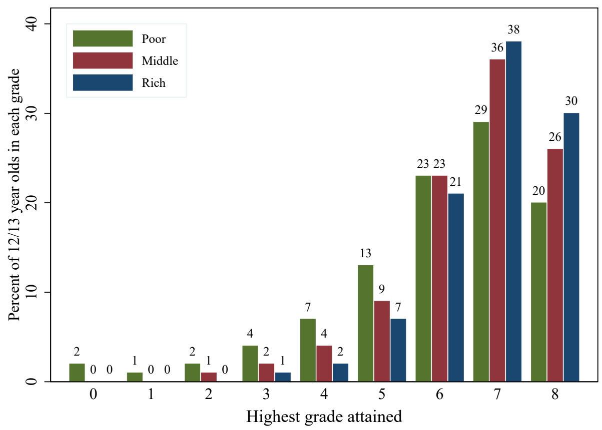
4.2 Descriptive Grade-Based Learning Profiles

The grade-based learning profiles in Figures B.2, B.3 and B.4 in the appendix show that in India, even after 8 years of formal schooling, there is a large, steady, and persistent gap in

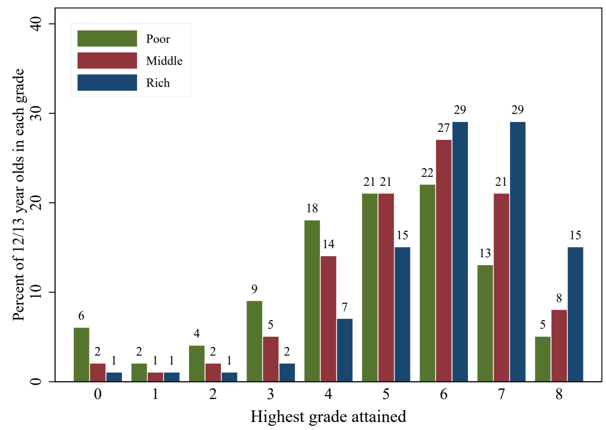
[18] These graphs include only those children who took at least one test.

Figure 5: Grade Attainment by Wealth, Ages 12-13

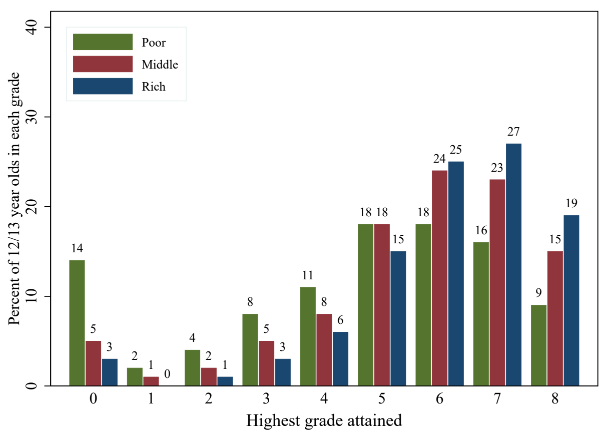
(a) India



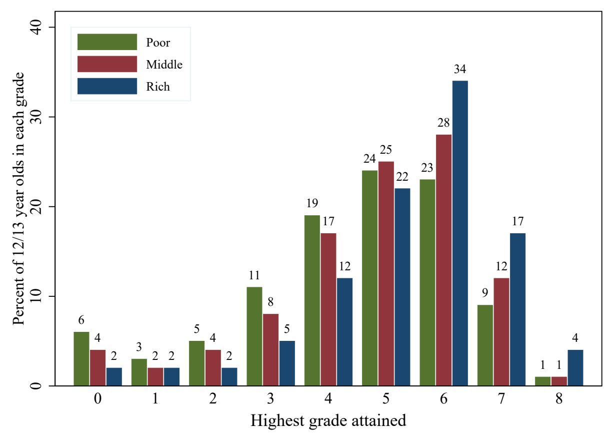
(b) Kenya



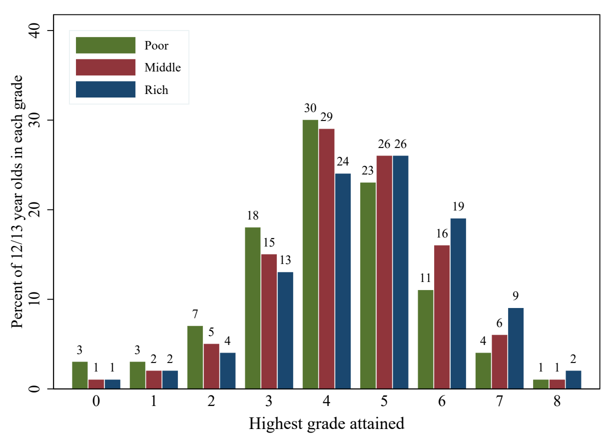
(c) Pakistan



(d) Tanzania



(e) Uganda



basic numeracy and local language literacy between children from rich and poor households.[19] The gap emerges early, as found in a study by Alcott and Rose (2017). In Pakistan, the gap in learning between children from rich and poor households is smaller (compared to India) but also emerges early and remains steady through to Grade 8. In the African countries, we see the gap emerging early, remaining largely steady over the years and then mostly closing by Grade 8. However, as our measure of numeracy is top-coded at a low level, this says nothing about the evolution of the overall gap in terms of a more sophisticated measure of mastery of a broader learning domain called mathematics—children from rich households may be getting further and further ahead on a different measure of mathematics competency. Work by Das (2018) shows that test score gaps that have developed by Grade 3 remain steady over primary school years and then widen dramatically by the time these children reach age 17 due to differential dropouts: children from poor households drop out at higher rates than children from rich households. While low-performing children from rich households may stay in school, even the high-performing children from poor households tend to drop out.

4.3 Counter-factual Calculations

As with any descriptive data, there are limitations to the counter-factual calculations shown earlier in Section 4. First, these calculations assume that the increase in learning from one grade to the next is something like a Local Average Treatment Effect (LATE). For example, the calculations assume that if a child who dropped out in Grade 4 had persisted to Grade 5, their likelihood of learning to read in that year was equal to the average observed gain in literacy from Grade 4 to Grade 5. This assumption is most likely false because of the positive self-selection of students into further grades. Students who drop out in earlier grades are likely to be weaker in terms of cumulative achievement and gain in learning from one year to the next. This positive self-selection of students implies that at least part of the

[19] Please note that the counter-factual calculations use grade-based learning profiles for 12-13 year olds only whereas the graphs in Figures B.2, B.3 and B.4 in the appendix show grade-based learning profiles for children of all ages.

gain in the descriptive learning profile does not reflect causal learning because those with higher cumulative learning persisted in school. Hence, all our counter-factual simulations overstate gains in learning. Since the descriptive learning profile is steeper than the causal learning profile, our estimates are optimistic and expansion of schooling may produce even less literacy than we suggest. Second, our assumption of a constant learning profile implies that the massive expansion of schooling systems over the last few years could not have caused the learning profiles to deteriorate. Again, this *inflates* our estimates for learning gains from higher grade attainment (Pritchett and Sandefur, 2017).

We run various simulations to see how total learning levels would change under different scenarios of grade attainment and learning achievement levels for children from poor households.

4.4 Counter-factual Scenario 1: What if all children from poor households have the grade attainment profiles of children from rich households?

In the first hypothetical scenario, we explore what happens to learning levels if all children from poor households have the grade attainment profiles of children from rich households while keeping their existing learning levels. Such a scenario would still leave more than 40% of Indian, Pakistani, and Ugandan children from poor households innumerate and unable to read a simple English story. In India, such a hypothetical scenario represents a gain of mere 5 and 4 percentage points in numeracy and English literacy respectively—only covering less than 10% of the gap between current learning levels of children from poor households and the goal of universal literacy/numeracy. However, in Kenya, where learning per grade is relatively high, such a scenario would cover close to half the gap from universal literacy in local language and English.

4.5 Counter-factual Scenario 2: What if all children from poor households have the learning achievement profiles of children from rich households?

In the second hypothetical scenario, we test what happens if all children from poor households have the learning profiles of children from rich households while maintaining their current grade attainment levels. A hypothetical scenario where all children from poor households aged 12-13 suddenly have the learning profiles of children from rich households (while keeping their current grade attainment profiles) would still leave more than 40% of the children from poor households innumerate in India, Pakistan, and Uganda. In Tanzania, more than half the children from poor households will still be unable to read a simple English story—with less than one-fourth of the gap from universal literacy being covered. In Uganda, more than half the children from poor households will still be unable to read a simple story in their local language—with only 13% of the gap from universal literacy being covered. This means that for most countries in our data set, a significant proportion of the children will be left illiterate and innumerate even if the learning gap between the rich and the poor was completely closed. However, excluding Kenya and Pakistan, improving learning profiles often covers a larger share of the gap between current learning levels of the poor and the goal of universal learning compared to improving grade attainment profiles.

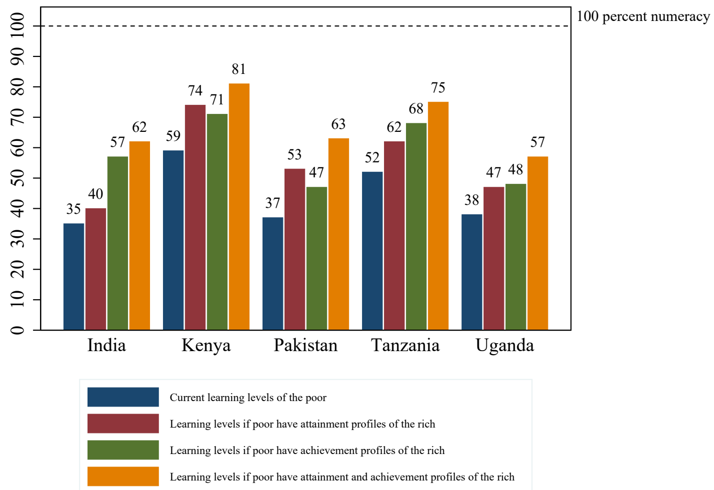
The gains in learning depend on the initial levels of illiteracy and innumeracy among children from poor households. For example, India has high illiteracy and innumeracy among children from poor households. The learning gap between children from the richest and poorest households is also huge: 27 percentage points in math, 26 percentage points in local reading, and 39 percentage points in English. A combination of a large number of illiterate/innumerate children from poor households and a big learning gap between the richest and poorest households leads to a significant jump in literacy/numeracy under the scenario where children from poor households have learning profiles of children from rich households: a jump of 22 percentage points in math, 20 percentage points in local reading, and 34 percentage points in English. Despite these jumps, close to a quarter of the children from poor households remain unable to read a simple sentence. On the other hand, for low illiteracy/innumeracy (among children from poor households) countries such as Kenya, the hypothetical scenario of giving poor children the learning profiles of children from rich households leads to an improvement in learning of 12, 13, and 15 percentage points for math, local reading, and English respectively. For such countries the percentage of children from poor households who are illiterate and innumerate is relatively low, so there isn't much gain to be made.

4.6 Counter-factual Scenario 3: What if all children from poor households have the grade attainment *and* learning achievement profiles of children from rich households?

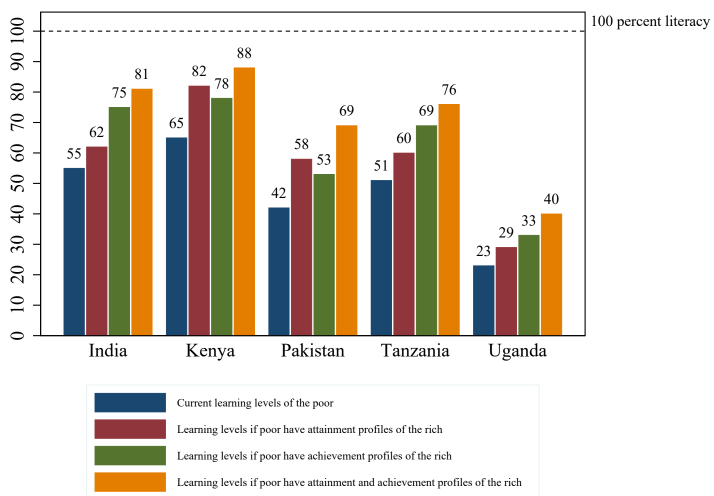
In the third hypothetical scenario, we explore what happens to learning levels if we completely close the gap between children from rich and poor households, that is, all poor children have the grade attainment and learning achievement profiles of children from rich households. For all countries except Kenya, bringing the learning and grade attainment levels of children from poor households to the levels of children form rich households still brings us no where close to the goal of universal mastery of basic literacy and numeracy. A hypothetical scenario where all children from poor households aged 12-13 suddenly have the learning achievement profiles and grade attainment profiles of children from rich households would still leave more than one-third of the children from poor households innumerate in India, Pakistan, and Uganda—with less than half of the gap to universal numeracy being covered. In Tanzania, more than half the children from poor households will still be unable to read a simple English story. In Uganda, more than half of the children from poor households will still be unable to read a simple story in their local language. This means that for most countries in our data set, a significant proportion of the population will still be illiterate and innumerate even if the learning and grade attainment gap between the rich and the poor was completely closed.

Figure 6: Counter-factual Simulations

(a) Math



(b) Local Language



(c) English

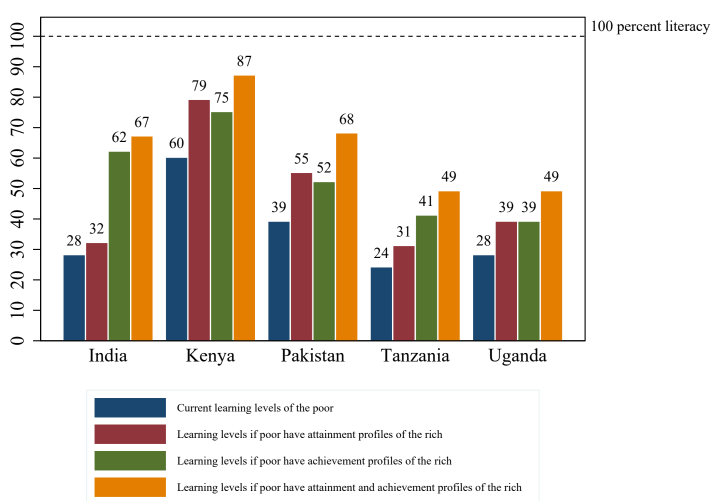


Table 8: Learning Gains from Having Grade Attainment and Learning Profiles of the Rich: Math

Country	Current Learning Levels of Poor (%)	Current Learning Levels of Rich (%)	Learning Gain from Having Same Grade Profile as Rich (% points)	Learning Gain from Having Same Learning Profile as Rich (% points)	Learning Gain from Having Same Grade and Learning Profile as Rich (% points)	Rich-Universal Gap (% points)
India	35	62	5	22	27	38
Kenya	59	81	15	12	22	19
Pakistan	37	63	16	10	26	37
Tanzania	52	75	10	16	23	25
Uganda	38	57	9	10	19	43

Table 9: Learning Gains from Having Grade Attainment and Learning Profiles of the Rich: Local Language

Country	Current Learning Levels of Poor (%)	Current Learning Levels of Rich (%)	Learning Gain from Having Same Grade Profile as Rich (% points)	Learning Gain from Having Same Learning Profile as Rich (% points)	Learning Gain from Having Same Grade and Learning Profile as Rich (% points)	Rich-Universal gap (% points)
India	55	81	7	20	26	19
Kenya	65	88	17	13	23	12
Pakistan	42	69	16	11	27	31
Tanzania	51	76	9	18	25	24
Uganda	23	40	6	10	17	60

Table 10: Learning Gains from Having Grade Attainment and Learning Profiles of the Rich: English

Country	Current Learning Levels of Poor (%)	Current Learning Levels of Rich (%)	Learning Gain from Having Same Grade Profile as Rich (% points)	Learning Gain from Having Same Learning Profile as Rich (% points)	Learning Gain from Having Same Grade and Learning Profile as Rich (% points)	Rich-Universal Gap (% points)
India	28	67	4	34	39	33
Kenya	60	87	19	15	27	13
Pakistan	39	68	16	13	29	32
Tanzania	24	49	7	17	25	51
Uganda	28	49	11	11	21	51

Table 11: Poor-Universal Gap Covered from Having Grade Attainment and Learning Profiles of the Rich: Math

Country	Current Learning Levels of Poor (%)	Poor-Universal Gap (% points)	Gap Covered from Having Same Grade Profile as Rich (%)	Gap Covered from Having Same Learning Profile as Rich (%)	Gap Covered from Having Same Grade and Learning Profile as Rich (%)
India	35	65	8	34	42
Kenya	59	41	37	29	54
Pakistan	37	63	25	16	41
Tanzania	52	48	21	33	48
Uganda	38	62	15	16	31

Table 12: Poor-Universal Gap Covered from Having Grade Attainment and Learning Profiles of the Rich: Local Language

Country	Current Learning Levels of Poor (%)	Poor-Universal Gap (% points)	Gap Covered from Having Same Grade Profile as Rich (%)	Gap Covered from Having Same Learning Profile as Rich (%)	Gap Covered from Having Same Grade and Learning Profile as Rich (%)
India	55	45	16	44	58
Kenya	65	35	49	37	66
Pakistan	42	58	28	19	47
Tanzania	51	49	18	37	51
Uganda	23	77	8	13	22

Table 13: Poor-Universal Gap Covered from Having Grade Attainment and Learning Profiles of the Rich: English

Country	Current Learning Levels of Poor (%)	Poor-Universal Gap (% points)	Gap Covered from Having Same Grade Profile as Rich (%)	Gap Covered from Having Same Learning Profile as Rich (%)	Gap Covered from Having Same Grade and Learning Profile as Rich (%)
India	28	72	6	47	54
Kenya	60	40	48	38	68
Pakistan	39	61	26	21	48
Tanzania	24	76	9	22	33
Uganda	28	72	15	15	29

5 Conclusion

Socio-economic status is a major factor impacting learning outcomes of children. Previous literature has shown that children from poor households are doubly disadvantaged in their learning outcomes relative to children from wealthier households as they both get less schooling and learn less per grade. This paper adds to the literature on equity gaps in learning by showing that these learning gaps between children from rich and poor households arise early. More importantly, this paper decomposes the learning gap into its main components: grade attainment and learning achievement per grade. Through the calculations in this paper, we are able to quantify how much of the learning gap between children from rich and poor households is due to a grade completion disadvantage versus low learning per grade.

The evidence in this paper confirms that learning levels are low across the board, for both children from rich and poor households. While the children from the poorest households are the worst performers, the learning levels for children from rich households are also low relative to any reasonable goals for learning. In order to make significant gains in improving literacy levels, closing the learning achievement and grade attainment gaps of children from rich and poor households is a positive step but not sufficient. For the world to get closer to the goal of universal literacy, all children across all wealth groups will have to experience meaningful gains in learning.

From a policy perspective, the agenda of reaching universal early conceptual mastery of basics is in itself an equity agenda for eliminating hidden exclusions. First, it is important to emphasize that improving equity requires a shift towards emphasis on equity of outcomes rather than equalization of inputs. To that end, authorities need to gather learning data to be able to target inequities. Second, it is necessary to have a focus on early conceptual mastery of basics by all children. Improving equity requires authorities to set minimum learning expectations for all children at each stage of the education cycle (entry, early, middle, late)—expectations that will gradually improve over time. Third, broad-based improvements and targeting marginalized groups can go hand in hand. Countries like Vietnam have demonstrated that equity and high performance (reaching OECD levels on international assessments) are possible, even at low levels of income. Improving equity will require progressive/pro-poor approaches targeting the learning of the marginalized to reduce unjustified inequalities. This will need to go hand in hand with raising learning levels for all.

6 Bibliography

Alcott, B. and Rose, P. (2015). How Can Education Systems Become Equitable by 2030?

Alcott, B. and Rose, P. (2017). Learning in India's Primary Schools: How do Disparities Widen Across the Grades?

Alderman, H. and Bundy, D. (2011). School Feeding Programs and Development: Are We Framing the Question Correctly?

Anand, P., Behrman, J., Dang, H., and Jones, S. (2018). Inequality of Opportunity in Education: Accounting for the Contributions of Sibs, Schools and Sorting across East Africa.

ASER (2014). Annual Status of Education Report (ASER) India. Technical report.

ASER (2015). Annual Status of Education Report (ASER) Pakistan. Technical report.

Bashir, S., Lockheed, M., Ninan, E., and Tan, J. (2018). Facing Forward: Schooling for Learning in Africa. Technical report.

Crouch, L. and Rolleston, C. (2017). Raising the Floor on Learning Levels: Equitable Improvement Starts with the Tail.

Das, J. (2018). Should I Stop Kidding Myself that Learning Leads to Educational Opportunity?

Davier, A., Holland, P., and Thayer, D. (2004). *The Kernel Method of Test Equating*. Springer-Verlag New York.

Davis-Kean, P. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment.

Dubow, E., Boxer, P., and Huesmann, L. (2009). Long-term Effects of Parents' Education on Children's Educational and Occupational Success: Mediation by Family Interactions, Child Aggression, and Teenage Aspirations.

Filmer, D. and Pritchett, L. (2001). Estimating Wealth Effects Without Expenditure Data—Or Tears: An Application To Educational Enrollments In States Of India.

Hanushek, E. and Woessman, L. (2011). The Economics of International Differences in Educational Achievement.

Ho, A. (2016). A Nonparametric Framework for Comparing Trends and Gaps across Tests.

Lupien, S., King, S., Meaney, M., and McEwen, B. (2000). Child's Stress Hormone Levels Correlate With Mother's Socioeconomic Status and Depressive State.

Patel, D. and Sandefur, J. (2018). Comparing Learning Levels Around the World: A NEAT Approach.

Paxson, C. and Schady, N. (2005). Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health, and Parenting.

PISA (2009). PISA 2009 Plus Results: Performance of 15-year olds in Reading, Mathematics and Science for 10 Additional Participants. Technical report.

PISA (2015). Results in Focus. Technical report.

Pritchett, L. (2013). *The Rebirth of Education*. Center for Global Development.

Pritchett, L. and Sandefur, J. (2017). Girls' Schooling and Women's Literacy: Schooling Targets Alone Won't Reach Learning Goals.

Spaull, N. and Taylor, S. (2015). Access to What? Creating a Composite Measure of Educational Quantity and Quality for 11 African Countries. pages 133–59.

UNICEF (2016a). The Impact of Language Policy and Practice on Children's Learning: Evidence from Eastern and Southern Africa. Technical report.

UNICEF (2016b). The Impact of Language Policy and Practice on Children's Learning: Evidence from Eastern and Southern Africa. Technical report.

WDR (2018). World Development Report 2018: Learning to Realize Education's Promise. Technical report.

Whelan, F. (2014). The Learning Challenge: How to Ensure that by 2020 Every Child Learns. Technical report.

Appendix

A Score Conversions

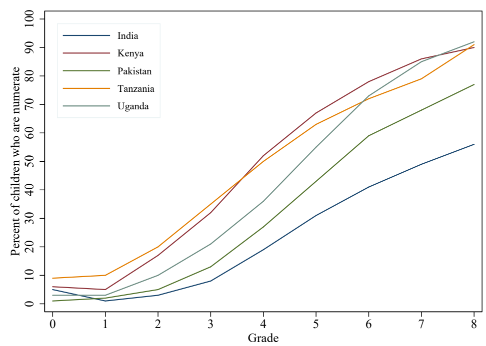
Table A.1: Score Conversions

Country	Year	Uwezo (Original)	ASER (Equivalent)
Kenya	2009, 2015	7 (divide)	5 (divide)
Kenya	2011-2014	8 (divide)	5 (divide)
Tanzania	2015	9 (multiply)	5 (divide)
Tanzania	2010, 2011-2014	7 (multiply)	5 (divide)
Uganda	2010-2015	7 (divide)	5 (divide)

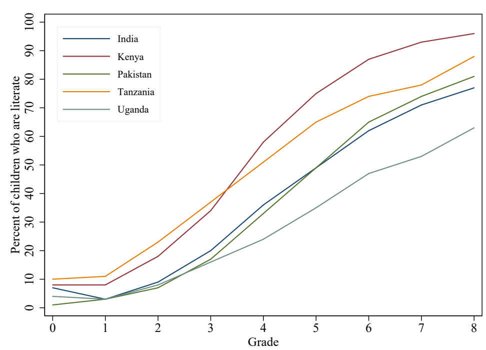
B Grade-Based Learning Profiles for Rich and Poor Children

Figure B.1: Learning Profiles (All Ages)

(a) Math



(b) Local Language



(c) English

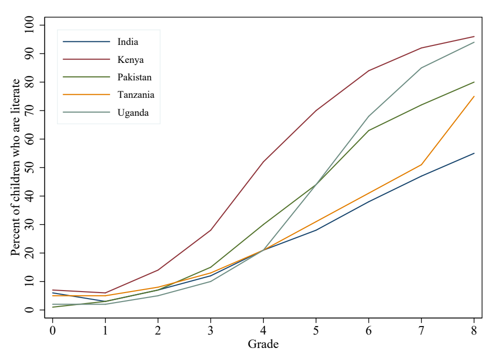
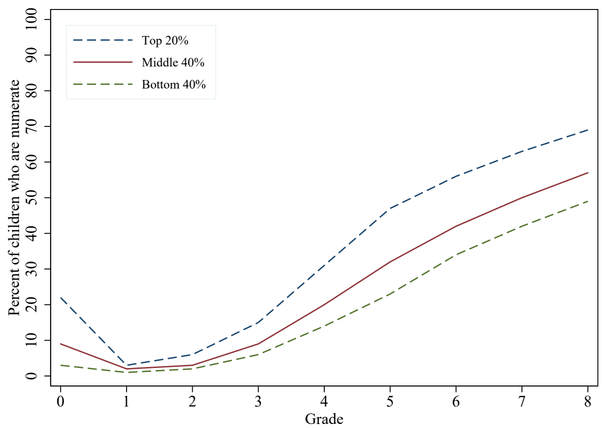
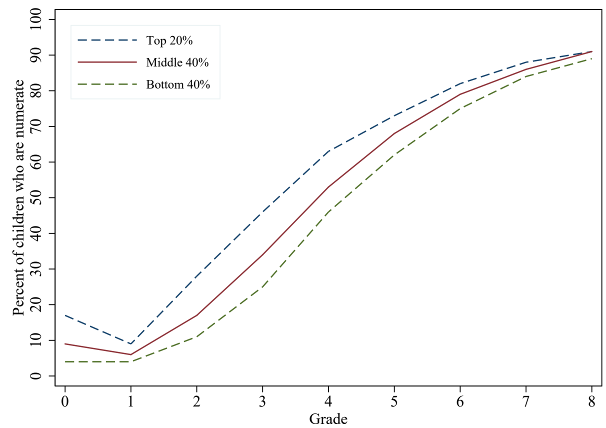


Figure B.2: Learning Profiles by Wealth: Math (All Ages)

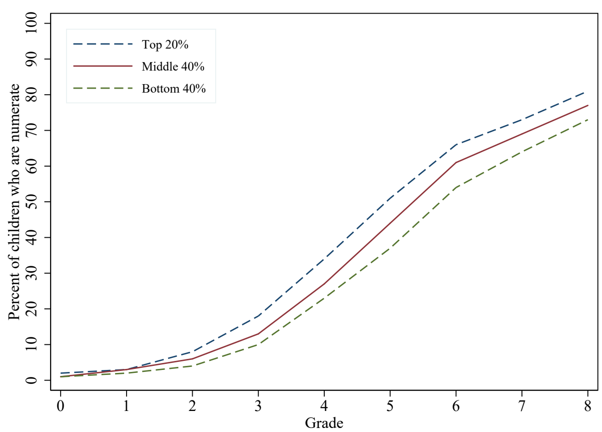
(a) India



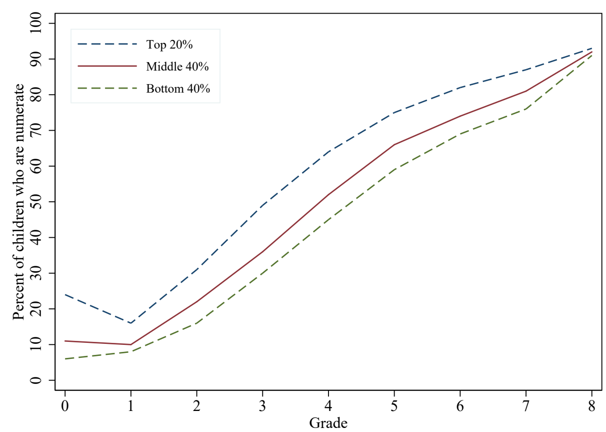
(b) Kenya



(c) Pakistan



(d) Tanzania



(e) Uganda

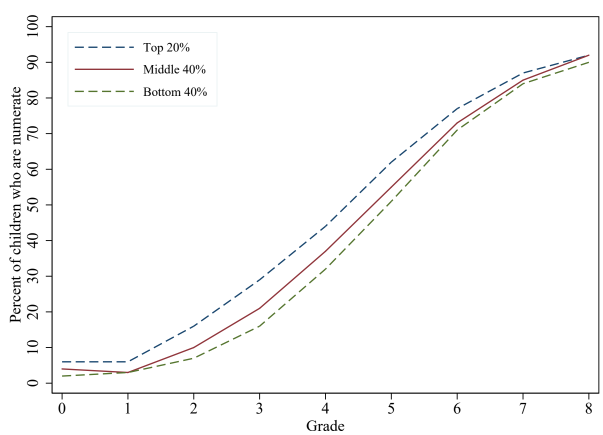
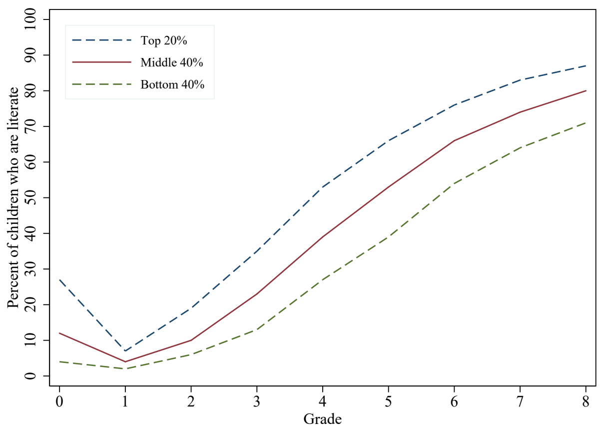
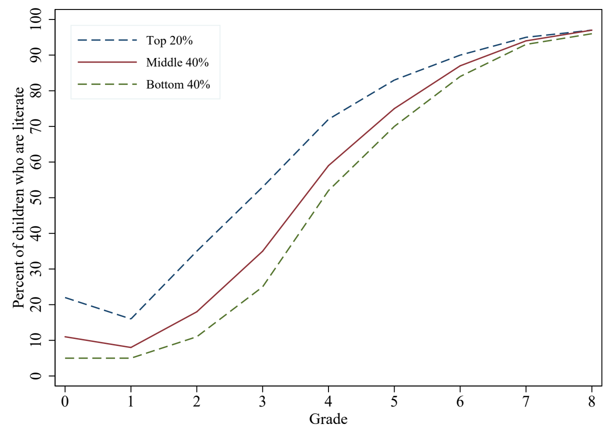


Figure B.3: Learning Profiles by Wealth: Local Language (All Ages)

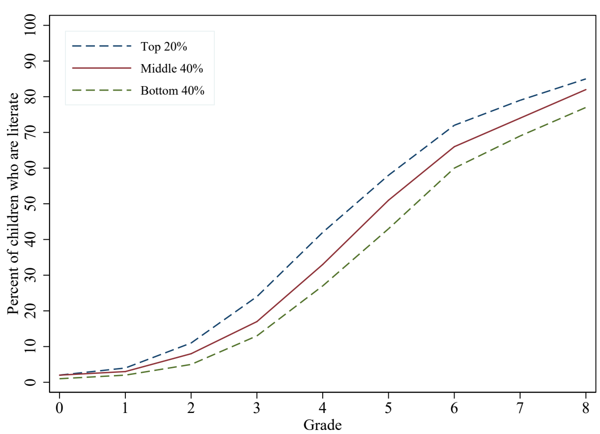
(a) India



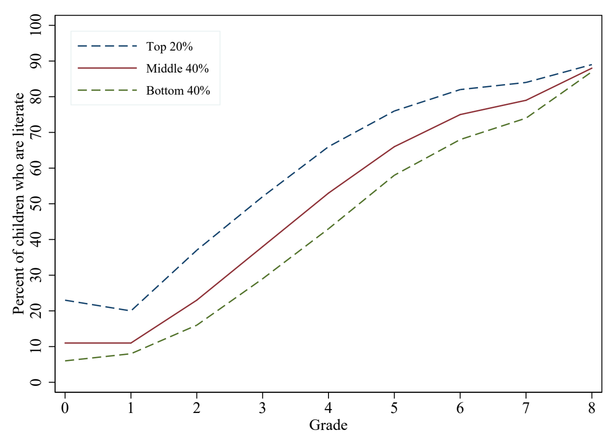
(b) Kenya



(c) Pakistan



(d) Tanzania



(e) Uganda

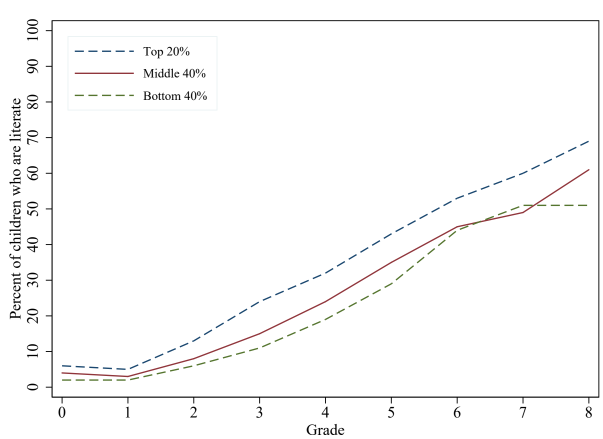
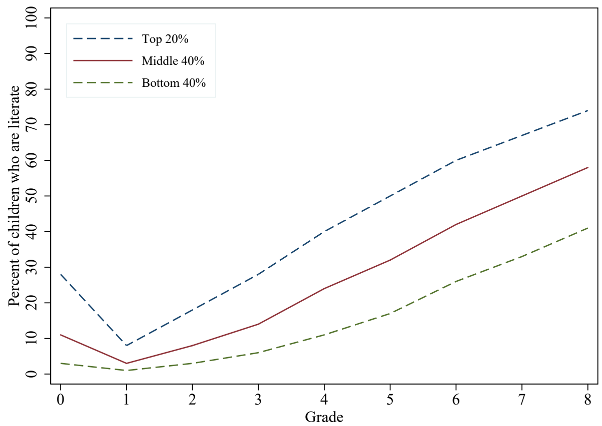
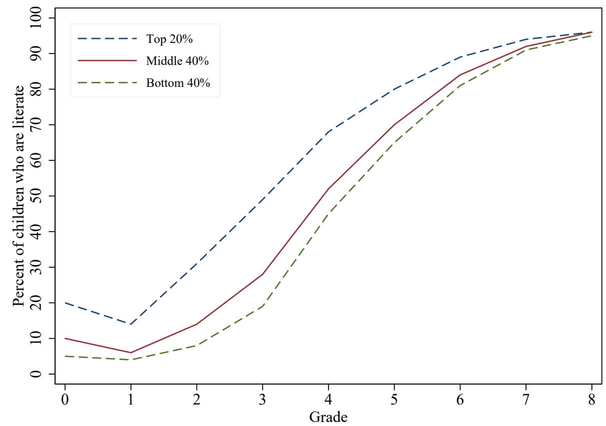


Figure B.4: Learning Profiles by Wealth: English (All Ages)

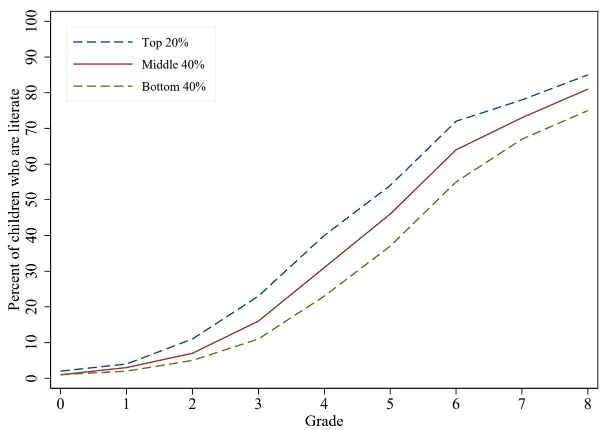
(a) India



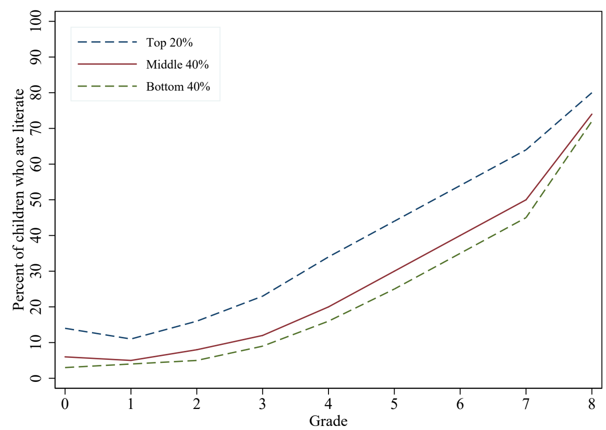
(b) Kenya



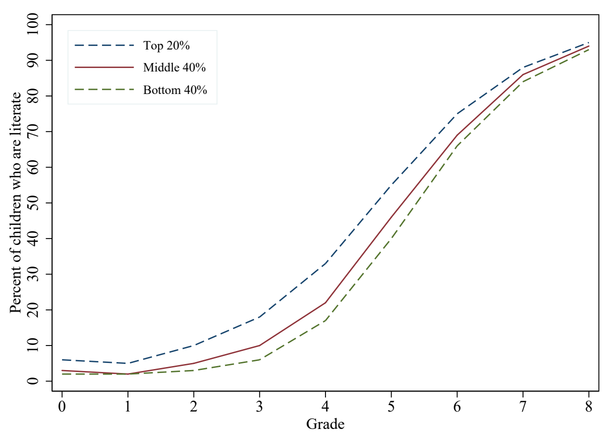
(c) Pakistan



(d) Tanzania



(e) Uganda



C Scoring Factors and Summary Statistics for Variables Entering the Computation of the First Principal Component

Table C.1: Scoring Factors and Summary Statistics for Variables Entering the Computation of the First Principal Component: Kenya

Asset	Scoring Factors	Mean	SD	Scoring Factors X SD	Mean Poorest 40%	Mean Middle 40%	Mean Top 20%
Bicycle Available	0.15	0.26	0.44	0.07	0.09	0.39	0.38
Car Available	0.24	0.03	0.18	0.04	0.00	0.01	0.16
Electricity Available	0.45	0.20	0.40	0.18	0.01	0.16	0.83
House with Wall	0.39	0.37	0.48	0.19	0.05	0.49	0.88
Mobile Available	0.34	0.64	0.48	0.16	0.38	0.84	0.96
Motorbike Available	0.21	0.07	0.25	0.05	0.00	0.07	0.23
Radio Available	0.30	0.66	0.47	0.14	0.41	0.82	0.92
TV Available	0.47	0.19	0.39	0.18	0.00	0.12	0.86
Water Available	0.28	0.41	0.49	0.14	0.22	0.47	0.76

Table C.2: Scoring Factors and Summary Statistics for Variables Entering the Computation of the First Principal Component: Pakistan

Asset	Scoring Factors	Mean	SD	Scoring Factors X SD	Mean Poorest 40%	Mean Middle 40%	Mean Top 20%
Electricity Available	0.51	0.88	0.33	0.17	0.71	1.00	1.00
Mobile Available	0.53	0.81	0.39	0.21	0.54	1.00	1.00
Own House	0.15	0.91	0.28	0.04	0.87	0.92	1.00
Solid House	0.36	0.30	0.46	0.17	0.08	0.22	1.00
TV Available	0.55	0.59	0.49	0.27	0.10	0.91	1.00

Table C.3: Scoring Factors and Summary Statistics for Variables Entering the Computation of the First Principal Component: Tanzania

Asset	Scoring Factors	Mean	SD	Scoring Factors X SD	Mean Poorest 40%	Mean Middle 40%	Mean Top 20%
Bicycle Available	0.08	0.49	0.50	0.04	0.38	0.58	0.53
Car Available	0.27	0.02	0.15	0.04	0.00	0.00	0.12
Electricity Available	0.47	0.24	0.42	0.20	0.01	0.20	0.81
House with Wall	0.36	0.44	0.50	0.18	0.15	0.53	0.85
Mobile Available	0.35	0.57	0.49	0.17	0.24	0.76	0.92
Motorbike Available	0.28	0.09	0.29	0.08	0.00	0.07	0.33
Own Cattle	-0.08	0.30	0.46	-0.04	0.33	0.30	0.24
Own Sheep	-0.12	0.34	0.47	-0.06	0.38	0.35	0.23
Radio Available	0.29	0.63	0.48	0.14	0.32	0.79	0.90
TV Available	0.51	0.15	0.35	0.18	0.00	0.03	0.71

Table C.4: Scoring Factors and Summary Statistics for Variables Entering the Computation of the First Principal Component: Uganda

Asset	Scoring Factors	Mean	SD	Scoring Factors X SD	Mean Poorest 40%	Mean Middle 40%	Mean Top 20%
Bicycle Available	0.20	0.51	0.50	0.10	0.30	0.65	0.62
Electricity Available	0.50	0.11	0.32	0.16	0.00	0.02	0.53
Mobile Available	0.45	0.59	0.49	0.22	0.09	0.89	0.94
Motorbike Available	0.35	0.10	0.30	0.11	0.00	0.02	0.46
Radio Available	0.36	0.66	0.47	0.17	0.39	0.81	0.88
TV Available	0.48	0.07	0.26	0.12	0.00	0.00	0.34
Water Available	0.16	0.30	0.46	0.07	0.21	0.33	0.47

Note: Each variable takes the value 1 if true, 0 otherwise. Scoring factor is the "weight" assigned to each variable in the linear combination of the variables that constitute the first principal component.

D Sensitivity Test

We conduct a sensitivity test to see if the percent literate or numerate at each grade level changes drastically when we change the cut-off defining our top wealth group. It could be argued that the "top 20%" is too broad of a category to pick up true levels of inequality. We do the sensitivity test for India, the country with the largest sample size in our data. The following tables show the percent literate and numerate (for all ages) at each grade level for the top 20%, top 15%, top 10%, and top 5%. The results show that the percent literate or numerate remains roughly constant across the various cut-offs, showing that the top 20% is a reasonable cut-off to define the richest group for the purposes of studying wealth gaps in learning.

Table D.1: Sensitivity Test for Percent Numerate Among the Rich: Math (India)

Wealth	0	1	2	3	4	5	6	7	8
Top 20%	21.56	2.55	5.93	15.32	31.35	46.84	56.30	62.94	69.44
Top 15%	21.67	2.53	5.97	15.78	31.69	47.19	56.75	63.17	69.67
Top 10%	20.85	2.57	6.18	16.07	31.77	47.50	56.56	63.48	69.89
Top 5%	18.65	2.64	6.07	16.04	32.30	47.48	57.06	63.72	69.76

Table D.2: Sensitivity Test for Percent Literate Among the Rich: Local Language (India)

Wealth	0	1	2	3	4	5	6	7	8
Top 20%	27.23	7.24	18.73	34.96	52.61	65.92	76.16	82.53	87.07
Top 15%	26.18	7.43	19.07	35.52	53.16	66.10	76.47	82.54	87.19
Top 10%	25.56	7.58	18.92	35.62	53.09	66.13	76.43	82.44	87.34
Top 5%	23.03	7.59	18.62	35.70	53.15	66.36	77.10	82.72	87.54

Table D.3: Sensitivity Test for Percent Literate Among the Rich: English (India)

Wealth	0	1	2	3	4	5	6	7	8
Top 20%	28.50	7.61	17.96	28.58	39.82	50.12	60.29	67.59	73.95
Top 15%	28.21	7.84	18.24	29.31	40.28	50.39	60.57	67.72	74.03
Top 10%	27.01	7.70	18.00	30.14	40.15	50.76	60.42	67.76	73.54
Top 5%	23.43	7.87	18.32	31.07	41.06	51.00	61.24	67.33	73.12