Policy Research Working Paper 9020

A Socioeconomic Disaggregation of the World Bank Human Capital Index

Ritika D'Souza
Roberta Gatti
Aart Kraay

WORLD BANK GROUP
Human Development Global Practice
&
Development Research Group
September 2019

Policy Research Working Paper 9020

Abstract

This paper documents inequality in health and education outcomes by constructing an index of human capital disaggregated by quintiles of socioeconomic status (SES) for a sample 51 mostly low- and middle-income countries. The index measures the expected future human capital of children born today, following the methodology of the World Bank Human Capital Index that was launched in October 2018. Within-country disparities in human capital outcomes across SES quintiles are large, accounting for roughly one-third of the total variation. On average, human capital outcomes increase with income at roughly the same rate across socio-economic groups within countries as they do across countries.

This paper is a product of the Human Development Global Practice and the Development Research Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at rdsouza@worldbank.org, rgatti@worldbank.org, and akraay@worldbank.org.

A Socioeconomic Disaggregation of the World Bank Human Capital Index

Ritika D'Souza Roberta Gatti Aart Kraay

Keywords: Human capital, health, education, inequality
JEL Classification Codes: I1, I2, O1, O4

This is a technical background paper for the forthcoming World Bank Human Development Practice Group Chief Economist study on geographical and socioeconomic disaggregation of the World Bank's Human Capital Index. The authors are Research Analyst, Office of the Human Development Chief Economist (D'Souza), Human Development Chief Economist (Gatti) and Director, Development Research Group (Kraay) at the World Bank, and can be contacted at rdsouza@worldbank.org, rgatti@worldbank.org, akraay@worldbank.org. We are grateful to Noam Angrist, Deon Filmer, Sven Neelsen, Adam Wagstaff, Harry Patrinos, and Syedah Aroob Iqbal for providing the data required for this paper, and to Noam Angrist, Deon Filmer, Luca Flabbi, Margaret Grosh, Syedah Aroob Iqbal, Leonardo Lucchetti, Aakash Mohpal, Adam Wagstaff, and David Weil for helpful comments. The views expressed here are the authors', and do not reflect those of the World Bank, its Executive Directors, or the countries they represent.

1. Introduction

The World Bank Human Capital Index (HCI) was launched in October 2018 as part of the Human Capital Project (HCP), a flagship World Bank initiative to accelerate progress in human capital outcomes around the world. The HCI measures the human capital that a child born today can expect to attain by her 18th birthday, given the risks of poor health and poor education prevailing in her country.[1] The HCI brings together indicators of health (child survival rates, stunting rates, and adult survival rates) and indicators of the quantity and quality of schooling (expected years of school, and international test scores). Using estimates of the economic returns to education and health, the components are combined into an index measuring the expected productivity as a future worker of a child born today, relative to the benchmark of complete education and full health. The index ranges from zero to one, and an HCI value of x implies that a child born today will only be $x \times 100$ percent as productive as a future worker as she would be if she enjoyed complete education and full health. By benchmarking shortfalls in future worker productivity due to current gaps in human capital outcomes among the young across countries, the HCI has underscored the urgency of accelerating progress in human capital outcomes and has been instrumental to elevating World Bank policy engagement around investments in human capital.[2]

The HCI covers 157 countries at the national level and is calculated using national averages of its component data. While the comparison of such national average human capital outcomes across countries is important, it masks significant differences across groups within a country, particularly between richer and poorer households. In this paper, we shed light on these inequalities by developing a disaggregation of the HCI by socioeconomic status (SES). Socioeconomic inequality in the distribution of human capital can reflect the presence of financial and access barriers to investing in human capital and

[1] The HCI was introduced in World Bank (2018a,b), and the methodology of the HCI is detailed in Kraay (2019).
[2] The global HCI, as well as the HCI disaggregated by socioeconomic status developed in this paper, both measure the expected future human capital of children born today under the assumption that they experience currently-prevailing risks of poor health and poor education faced by children aged 0-17. For terminological convenience we refer to this as "human capital" or "human capital outcomes" for short. However, we emphasize that this is a measure of the expected future *increment* to the stock of human capital of a country (defined as the productive capacity attributable to health and education of the entire workforce), and not the stock of human capital itself.

can itself affect the human capital of the next generation.[3] Measuring these inequalities is a first step towards targeting interventions to build human capital to the most disadvantaged households.[4]

A limitation of the global HCI is that several of its component data sources cannot readily be disaggregated by SES. Here we instead rely on comparable cross-country data from Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS) to measure child survival rates, enrollment rates, and stunting rates disaggregated by quintiles of SES. We also draw on the student-level harmonized test score data underlying the HCI to obtain test scores disaggregated by quintiles of SES. We then combine these into an SES-disaggregated human capital index (SES-HCI) using the methodology of the HCI. The resulting SES-HCI dataset consists of 88 country-year observations covering 51 primarily low- and middle-income countries. Half of these countries have more than one data point, thus enabling an analysis of within-country over-time disparities in human capital outcomes across SES quintiles.

Our headline finding is that gaps in human capital outcomes across SES quintiles within countries are large. Pooling all SES quintiles in all countries, roughly one-third of the total variation in human capital outcomes is due to variation across SES quintiles within countries. We also document how human capital outcomes vary with proxies for the average incomes of each SES quintile. We find a striking pattern: human capital outcomes increase with income across countries at roughly the same rate as they do within countries across SES quintiles. This finding is surprising because it indicates that the sharing of income-related human capital risks is on average no better within countries (where in principle social protection programs might mitigate these risks) than it is between countries at different income levels.

We also examine patterns in rich-poor gaps in human capital across countries and over time. Across countries, we find that gaps between the top and bottom quintile in child survival and expected years of school are narrower in richer countries than in poorer countries, reflecting a tendency for outcomes in the poorest quintiles to increase more steeply with country-level average income across countries. In contrast, rich-poor gaps in test scores tend to be higher in richer countries with higher test scores, although as discussed in more detail below, this to some extent is a consequence of the methodology used to harmonize test scores. Within countries over time, we find a weak tendency for rich-poor gaps in the overall SES-HCI and its components to decline as average human capital outcomes

[3] Flabbi and Gatti (2018) review the literature of household investment in human capital. Rossi (2019) reviews the macro literature linking human capital with income and growth.

[4] Disaggregating the HCI along different margins - by geographic areas, by gender, or by ethnic groups – can also prove useful for policy making and targeting of interventions.

improve. We also find some evidence that improvements in country-level averages of the SES-HCI and its components are slower in faster-growing countries. However, these latter findings must be taken with a grain of salt, since they are based on a sample of just 24 countries where we have comparable data at two points in time on all of the SES-HCI components.

Although the SES-HCI uses the same methodology as the HCI, it differs in several key respects. *First*, the SES-HCI uses household survey-based measures of school attendance to measure the quantity of schooling, while the HCI primarily uses administrative data on enrollment rates. As is well-known, these two measures of school participation can differ considerably.[5] While we do not offer new insights into the sources of these discrepancies, we document their extent in this particular context. *Second*, due to data limitations, the SES-HCI measures expected years of school between ages 6 and 17, while the HCI relies on administrative data on pre-primary through upper-secondary enrollment, covering the 4 to 17 age range. *Third*, the household survey data we rely on does not provide estimates of adult mortality, and so we cannot calculate adult survival rates by SES quintile. This means that the health component of the SES-HCI is based only on stunting rates, unlike the HCI which uses stunting rates and adult survival rates for all of the countries covered in the SES-HCI. *Fourth*, there are a number of minor discrepancies between the SES-HCI and HCI data on child survival, stunting, and test scores, that we detail in Section 3. Taken together, these differences imply that the SES-HCI data at the quintile level, and averaged to the national level, are not fully comparable or consistent with the global HCI, and countries' scores and relative positions can differ between the SES-HCI and the HCI. Accordingly, comparisons between the two should be made cautiously and recognizing these differences which we detail below.

This paper builds on a very large literature that has documented inequalities in health and education outcomes across income levels and/or proxies for socioeconomic status. See for example early contributions such as Wagstaff (2000, 2002) on health outcomes, and Filmer and Pritchett (1998, 2001) on school participation. There also is a large literature documenting variation in test scores across students of differing socioeconomic status.[6] Our modest contributions to this literature are to (a) assemble a large cross-country panel dataset combining health outcomes, school participation outcomes, and test scores disaggregated by socioeconomic status, (b) to assemble these components into a measure of human capital compatible with the World Bank's HCI, and (c) to document patterns across countries and over time in this particular measure of human capital of the next generation, and its components.

[5] See for example Urquiola and Calderon (2006).

[6] See for example the discussion in Garcia et al. (2016).

This paper is organized as follows. Section 2 provides a brief overview of the components and aggregation methodology of the existing national-level HCI, and readers familiar with the HCI methodology can skip this section. Section 3 describes the component data sources of the SES-HCI. Section 4 calculates the SES-HCI by quintiles and documents the differences between the average (across quintiles) of the SES-HCI and the HCI. Section 5 contains our main findings, documenting patterns in inequalities in human capital outcomes within and between countries. Section 6 concludes.

2. A Recap of the HCI

The HCI measures the quantity of human capital that a child born today can expect to attain by age 18, given the risks of poor health and poor education that prevail in the country where she lives.[7] The HCI is designed to highlight how improvements in current health and education outcomes shape the productivity of the next generation of workers, assuming that children born today experience over the next 18 years the educational opportunities and health risks that children in this age range currently face.

The HCI measures key points along the trajectory from birth to adulthood of a child born today. In the poorest countries in the world, there is a significant risk that the child does not even survive to her fifth birthday. Even if she does reach school age, there is a further risk that she does not start school, let alone complete the full cycle of 14 years of school from pre-school to Grade 12 that is the norm in rich countries. The time she does spend in school may translate unevenly into learning, depending on a variety of factors including the quality of teachers and schools she experiences. When she reaches age 18, she carries with her lasting effects of poor health and nutrition in childhood that limit her physical and cognitive abilities as an adult.

The HCI quantifies the key stages in this trajectory and their consequences for the productivity of the next generation of workers, with three components:

Component 1: Survival from birth to school age, measured using under-5 mortality rates.

Component 2: Expected Years of Learning-Adjusted School, combining information on the quantity and quality of education. The quantity of education is measured as the expected number of years of school a

[7] The discussion in this section follows closely World Bank (2018a, Chapter 3), World Bank (2018b), and Kraay (2019).

child can expect to obtain by age 18 given the prevailing pattern of enrollment rates across grades. The quality of education reflects ongoing work at the World Bank to harmonize test scores from major international student achievement testing programs.[8] These are combined into a measure of quality-adjusted school years using the "learning-adjusted years of school" conversion metric proposed in the 2018 World Development Report.[9]

Component 3: Health, which in the absence of a single broadly-accepted, directly-measured, and widely-available metric, is captured by two alternative proxies: (i) adult survival rates, defined as the fraction of 15-year-olds that survive until age 60, and (ii) the rate of stunting for children under age 5. Adult survival rates can be interpreted as a proxy for the range of fatal and non-fatal health outcomes that a child born today would experience as an adult if current conditions prevail into the future. Stunting is broadly accepted as a proxy for the pre-natal, infant and early childhood health environment, and so summarizes the risks to good health that children born today are likely to experience in their early years – with important consequences for health and well-being in adulthood.

These three components are converted into contributions to productivity relative to benchmarks of complete education and full health. Multiplying these contributions to productivity gives the overall HCI. The resulting index ranges between 0 and 1. A country in which a child born today can expect to achieve both full health (no stunting and 100 percent adult survival) and full education potential (14 years of high-quality school by age 18) will score a value of 1 on the index. Therefore, a score of 0.70 signals that the productivity as a future worker for a child born today is 30 percent below what could have been achieved with complete education and full health. Because the theoretical underpinnings of the HCI are in the development accounting literature, the index is linked to real differences in how much income a country can generate in the long run. If a country has a score of 0.50, then the gross domestic product (GDP) per worker could be approximately twice as high if the country reached the benchmark of complete education and full health.[10]

[8] The harmonized test score database used in the HCI is described in Patrinos and Angrist (2018).

[9] For a detailed justification of this methodology, see Filmer, Rogers, Angrist and Sabarwal (2018)

[10] As discussed in further detail on page 23 of Kraay (2019), the economic logic underlying this statement requires abstracting from the (small) contribution of cross-country differences in child survival rates to cross-country differences in the HCI.

3. SES-Disaggregated Data for the SES-HCI

3.1 Overview of Data Sources

To construct the SES-HCI, we require SES-disaggregated data on mortality rates among children under 5, stunting rates among children under 5, school enrollment rates by age, and harmonized test scores. We obtain the first three indicators from Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS). Both are nationally-representative household surveys that collect data on measures of population, health and nutrition. The DHS program has fielded over 400 surveys across 90 countries, while over 300 MICS have been carried out in more than 100 countries. Many (but not all) DHS and MICS datasets include birth histories (used to calculate child mortality rates); school attendance rates by age (used to calculate expected years of school); and anthropometric data (used to calculate stunting rates among children under the age of 5). Crucially, the DHS and MICS also contain information on household characteristics and asset ownership, and there is an accepted methodology for combining these into a wealth index. This wealth index can be used to disaggregate child mortality, school attendance, and stunting by SES quintile.[11]

We obtain this data by drawing on two existing compilations of DHS and MICS data disaggregated by SES quintile. SES-disaggregated school enrollment data by age come from the latest update to the household wealth and educational attainment dataset first described in Filmer and Pritchett (1998). The latest version of their dataset contains 345 DHS and MICS surveys, with enrollment rates for 99 countries over the period 1990-2017.[12] The SES-disaggregated under-5 mortality rates and stunting rates come from the latest edition of the Health Equity and Financial Protection Indicators (HEFPI) database, described in Wagstaff, Eozenou, Neelsen and Smitz (2019). The database assembles information from multiple survey programs, including the DHS and MICS that are used in this analysis, and reports health outcome and health care usage indicators disaggregated by SES quintiles for a large cross-section of countries. Both datasets calculate the SES index in the same way, using principal component analysis to

[11] In the DHS/MICS context, these indexes are usually referred to as “wealth” or “asset” indexes. As discussed below, international testing program databases construct conceptually similar indexes but use terms such as “economic, social and cultural status (ESCS)” – as used in the PISA assessments, for example. For terminological convenience, we refer to all of these measures generically as “socioeconomic status (SES)” indexes.

[12] Filmer, Deon. 2018. “Education Attainment and Enrollment around the World: An International Database.” http://www.worldbank.org/en/research/brief/edattain. Accessed February 2019.

aggregate responses to questions on asset ownership and housing characteristics into a household-level SES index.[13]

The final ingredient in this exercise is harmonized test score data disaggregated by SES. We obtain this data from Abdul-Hamid and Iqbal (2019), who in turn draw on the same database of student-level harmonized test scores used in the 2018 HCI, as described in Patrinos and Angrist (2018). This database harmonizes test scores from seven major international and regional student achievement testing programs into common units, resulting in "harmonized learning outcomes (HLO)" covering over 160 countries since 2000. As discussed in more detail below, Abdul-Hamid and Iqbal (2019) develop proxies for the SES of the households in which each student resides, based on data collected by the testing program on students' home possessions, as well as parental education and occupation. We obtain average harmonized test scores by centiles of the SES index for all tests converted into HLO units, and as discussed below we use these detailed centile distributions to align quintiles of test takers with quintiles of households in the DHS and MICS.

The DHS/MICS are typically carried out approximately every five years in a given country, although the timing and frequency of surveys varies considerably across countries. Similarly, each of the major testing programs adheres to its own cycle and covers different sets of countries. To create the sample of countries and years for this paper, we align data from DHS/MICS surveys with testing datasets in two stages. First, we create a single cross-section of 51 countries, using the most recently-available DHS/ MICS, and testing data available. In this first stage, we require the gap between the DHS/MICS and the testing data to be no more than ten years. Second, for countries with more than one DHS/MICS and more than one test, we create a panel dataset by matching each available DHS/MICS with the nearest available testing program. In this second stage, we limit gaps between the DHS/MICS and the testing data to no more than five years, and we never assign the same testing data to more than one DHS/MICS. That is, we align data to ensure a gap of no more than five years, with no survey or assessment being used more than once.

---

[13] See for example documentation of the methodology for the DHS wealth index in Rutstein and Johnson (2004). Filmer and Pritchett (1998) show that the asset index is strongly correlated with more traditional welfare indicators such as household-size-adjusted consumption expenditures in datasets where both can be constructed. They also show that the asset index works as well as, or better than consumption expenditures as a proxy for long-run household wealth in predicting children's school enrollment. In related, Filmer and Scott (2011) study LSMS data from 11 countries and find weaker correlations of rankings of households based on per capita expenditure versus based on asset indexes, but nevertheless conclude that analyses of inequalities in education and health outcomes tend to be quite robust to the use of either measure.

This approach to aligning DHS/MICS with testing program data gives us a sample of 88 country-year observations covering 51 countries. Appendix Table A1 describes the country coverage of the HCI socio-economic disaggregation sample by region and income group. Low-income and lower-middle-income countries comprise the majority of observations, reflecting the DHS/MICS coverage. Appendix Table A2 list the full set of 88 country-year observations in our dataset and indicates the timing and source of all the component data. In the remainder of this section, we discuss each of the SES-disaggregated components of the SES-HCI in more detail, and briefly summarize differences between SES-HCI data sources and the data sources used in the national-level HCI. Throughout we will frequently refer to **Table 1** which reports summary statistics for the components of the SES-HCI by SES quintiles.

3.2 Probability of Survival to Age 5

The probability of survival to age 5 is calculated as the complement of the under-5 mortality rate. The under-5 mortality rate is the probability of a child born in a specified year dying before reaching the age of 5 if subject to current age-specific mortality rates and is frequently expressed as a rate per 1,000 live births. For SES-HCI, under-5 mortality rates are obtained from the HEFPI database and are calculated using the same life-table synthetic-cohort probability method employed in DHS reports and programmed in the Stata module SYNCMRATES (Masset 2016). Survey-based mortality rates rely on birth histories in the five years prior to the survey[14], and consequently report average rates for this period. Summary statistics on child survival rates by quintile are reported in the first panel of **Table 1**. The top panel of **Figure 1** shows the dispersion in child survival rates across countries and across quintiles, plotting national average child survival as well as child survival rates in the top and bottom SES quintile (on the vertical axis) against log GDP per capita (on the horizontal axis) for the most recent cross-section of 51 countries. Child survival rates range from around 0.85 in the poorest countries to around 0.98 in the richest countries in this sample. The range of outcomes within countries is wide as well, particularly in the poorest countries.

[14] The majority of survey data comes in one of two forms: the full birth history (FBH), whereby women are asked for the date of birth of each of their children, whether the child is still alive, and if not, the age at death; and the summary birth history (SBH), whereby women are asked only about the number of their children ever born and the number that have died (or equivalently the number still alive). FBH data, collected by all Demographic and Health Surveys (DHS) and increasingly also Multiple Indicator Cluster Surveys (MICS), allow the calculation of child mortality indicators for specific time periods in the past. This allows DHS and MICS to publish under-five child mortality estimates for five 5-year periods before the survey, that is, 0 to 4, 5 to 9, 10 to 14, etc. (UN IGME, 2019).

In Burkina Faso for example, child survival averages 0.82 in the lowest quintile and 0.91 in the highest quintile, a range that is nearly three-quarters as large as the cross-country range.

In the bottom panel of **Figure 1** we plot the national average of child survival rates as used in the SES-HCI against those used in the HCI. The data used in the HCI come from annual estimates of child mortality constructed by the UN Interagency Group for Child Mortality Estimates (UN-IGME). UN-IGME compiles all available nationally representative data relevant to the estimation of child mortality - including data from vital registration systems, population censuses, household surveys and sample registration systems – and calculates mortality rates, making adjustments for data quality if needed.[15] In practice, in many low-income countries in our sample where vital registries are weak, DHS/MICS are an important data source for UN-IGME. Overall, therefore, national averages of child survival rates obtained directly from the UN-IGME (as used in the HCI) and those obtained from the HEFPI dataset (as used in the SES-HCI) are highly correlated. One further consideration in comparing the SES-HCI versus the global HCI data is that the global HCI used UN-IGME data for 2017, while the SES-HCI data refers to the year of the underlying DHS/MICS, which frequently is several years older than 2017. Since child survival rates in the UN-IGME data trend upwards for most countries, this means that the SES-HCI data (which comes from earlier years) on average shows slightly lower child survival rates than the global HCI data for the same country. Consistent with this, three countries that are the largest outliers in this graph (Burkina Faso, Burundi and Niger) are also countries where the SES-HCI data are relatively old (2010, 2010, and 2012 respectively).

3.3 Expected Years of School

Expected years of school is defined as the number of years of school that a child can expect to attain over a given age range if she experiences currently-observed enrollment patterns. It is measured as the sum of age-specific enrollment rates over the age range of interest. We implement this calculation using DHS/MICS data on school attendance (not enrollment) by age, disaggregated by SES quintiles, as constructed in the latest version of the database originally presented in Filmer and Pritchett (1998). They rely on DHS/MICS questions which gather information on whether individuals between the ages of 6 and

[15] The UN IGME assesses data quality, recalculates data inputs and makes adjustments if needed by applying standard methods. It then fits a statistical model to these data to generate a smooth trend curve that averages over possibly disparate estimates from the different data sources for a country. Finally, it extrapolates the model to a target year (UN IGME 2019).

24 attended a given grade or level of education at least one day during the academic reference year.[16] For the SES-HCI, we sum the SES-quintile-averaged attendance rates across ages 6-17 to arrive at a measure of expected years of school, with a maximum possible of 12 years. Summary statistics on expected years of school by quintile are reported in the second panel of **Table 1**. In parallel with the analysis for child survival, the top panel of **Figure 2** shows the dispersion in expected years of school across countries and across SES quintiles, plotting national average expected years of school as well as the corresponding averages in the top and bottom SES quintile (on the vertical axis) against log GDP per capita (on the horizontal axis) for the most recent cross-section of 51 countries. Expected years of school range from below six years in countries such as Mali, Burkina Faso and Niger, to around 11 years in the richest countries in this sample. The range of SES quintiles is again widest in the poorest countries. In Niger for example, the top-bottom SES quintile range is nearly 5 years of school, while in middle-income countries such as Egypt the gap is only 1.2 years. However, this pattern is not universal: for example, a very poor country such as Burundi has a smaller rich-poor gap of 2.3 years while much richer Guatemala has a gap of 3 years.

There are significant conceptual and practical differences between the measure of expected years of school based on survey data that we use in the SES-HCI and the corresponding measure used in the global HCI calculations. First, the global HCI calculates expected years of school from ages 4 to 17, for a maximum possible of 14 years, while the SES-HCI for reasons of data availability focuses on the 6-17 age range.[17] Second, the HCI uses administrative data on enrollment rates, adjusted for repetition where possible, by four levels of school (pre-primary, primary, lower secondary, upper secondary) to proxy for age-specific enrollment rates within these brackets. This difference in the type of data introduces a plethora of potential sources of discrepancies between the resulting measure of expected years of school. Conceptually, attendance and enrollment are different since not all children who are enrolled attend school. The DHS/MICS survey data does not provide information on whether a child is repeating the grade of school she is currently attending, while administrative data permits adjustment for repetition. Administrative data may be collected by school officials who have incentives to over-report enrollment in order to obtain greater school financing. These differences are readily apparent in the bottom panel of

[16] The DHS questions are typically:

1) Did [NAME] attend school at any time during [REFERENCE ACADEMIC YEAR] school year?
2) During (this/that) school year, what level and grade is/was) [NAME] attending?

[17] DHS and MICS surveys have collected information on children 5 and younger only in recent years, and the data on pre-primary schooling is still less than fully consistent across a large sample of countries.

Figure 2, which plots the SES-HCI measure of expected years of school (on the vertical axis) against the corresponding HCI measure (on the horizontal axis). The absolute difference between the two measures is less than one year for 23 countries, and less than 1.5 years for 36 countries. While there are many factors contributing to these differences, two key offsetting ones are worth noting in this context. On the one hand, the SES-HCI measure of expected years of school is lower than the corresponding measure in the global HCI because the latter includes up to two years of pre-primary education while the former does not. On the other hand, the household survey-based data in the SES-HCI suggests higher rates of school participation than do administrative enrollment data for lower- and upper-secondary school. In contrast, administrative and household data on primary schooling tend to be better aligned in our sample.

Shedding new light on the relative importance of these (and many other potential) sources of such discrepancies is beyond the scope of this exercise. Rather we simply note that these differences are present and limit the comparability of the expected years of school measure between the SES-HCI and the HCI. Importantly, these discrepancies suggest a significant long-term measurement agenda to better measure school participation, particularly in the poorest countries where administrative data is likely to be the weakest.

3.4 Harmonized Test Scores

In the SES-HCI, we follow the HCI in using data on harmonized learning outcomes (HLO) to adjust expected years of school for quality. We draw on the student-level test score data in the HLO database used in the HCI, as described in Patrinos and Angrist (2018). These authors assemble data from seven international testing programs and develop a methodology to convert these into common units. The HLO is measured in Trends in International Mathematics and Science Study (TIMSS)-equivalent units, where 400 corresponds to minimum proficiency and 625 is advanced attainment. Abdul-Hamid and Iqbal (2019) develop SES indexes for the households in which the children taking the test live, based on data collected by each testing program on students' home possessions as well as the educational and occupational status of the students' parents.[18] We are grateful to these authors for kindly providing to us the distribution of

[18] The student-level SES index available in international testing programs typically is based on some combination of measures of (a) parents' highest level of education, (b) parents' occupation, and (c) a list of household possessions. Abdul-Husein and Iqbal (2019) draw on existing SES indexes provided by Programme for International Student Assessment (PISA), Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), and the 2014 round of Program of Analysis of Education Systems (PASEC)). They then apply similar methodologies to data on student characteristics to calculate student-level SES indexes for Progress in International Reading Literacy Study (PIRLS), Trends in International Maths and Science Study (TIMSS), Latin American Laboratory for Assessment

test scores by SES, in the form of average test scores within each SES centile, separately for all country x test x subject x grade x year groupings represented in their dataset. We then average these across subjects, grades, and tests using the same rules that were used in the HCI to obtain country x year HLO scores by centile of the SES distribution.

To combine the DHS/MICS data with test scores, we make the crucial (and unavoidable) assumption that the ordering of households in the SES index in the test score database is the same as the ordering of households in the SES index in the DHS/MICS data. With this assumption, we can estimate average learning outcomes for children in each SES quintile of the DHS/MICS data. To do so, however, we have to address one further complication, which is that the SES distribution of test score data by construction covers only the households of children who are attending school (since the HLOs are calculated using only school-based tests), while the DHS/MICS data covers all households, including those whose children do not attend school.

We address this complication by "merging" the DHS/MICS data on attendance by age and SES quintile with the distribution of HLO scores by SES centiles. Consider for example a country where the test scores are taken from PISA, which is administered to 15-year-olds. We use enrollment rates for 15-year-olds by SES quintile in DHS/MICS to calculate the fraction of students attending school associated with each SES quintile in the DHS/MICS. For example, if students in the bottom SES quintile are more likely to drop out of school, then students in the bottom quintile might for example represent only 15 percent of test takers even though they account for 20 percent of households. In this case, we assign the average HLO score for the poorest 15 percent of test takers (according to the SES index in the test score data) to the households in the poorest quintile in the DHS/MICS. We apply a similar approach for each quintile to arrive at average HLO scores for each DHS/MICS quintile.

Summary statistics on HLOs by SES quintile are reported in the third panel of **Table 1**. The top panel of **Figure 3** shows the dispersion in HLOs across countries and across quintiles, plotting the overall national average HLO as well as the corresponding averages in the top and bottom quintiles (on the vertical axis) against log GDP per capita (on the horizontal axis) for the most recent cross-section of 51 countries. HLOs vary widely across countries: in the poorest countries such as Niger, Chad or Mali, the

---

of the Quality of Education (LLECE), Early Grade Reading Assessments (EGRA), and pre-2014 rounds of PASEC. They combine this with student-level data on harmonized test scores developed in Patrinos and Angrist (2018) as used in the global HCI to arrive at SES-disaggregated SES scores. We are grateful to the authors for providing distributions of harmonized test scores by centiles of the wealth distribution required for our calculations described below.

average student has an HLO score around 300, well below the TIMSS standard of minimum proficiency of 400. In contrast, in richer countries such as Vietnam, Moldova or Turkey, average test scores fall in the 450-500 range. As with the previous indicators, the range of outcomes within some countries is wide as well. Moreover, in contrast to the SES-HCI components discussed in the previous subsections, it is visually apparent from the top panel of **Figure 3** that dispersion in test scores on average tends to be larger in richer countries in our sample. However, as discussed in more detail in Section 5.3, this pattern is to some extent an artifact of the methodology used to harmonize test scores.

In the bottom panel of **Figure 3** we plot the national average HLO as used in the SES-HCI (on the vertical axis) against the national average HLO as used in the HCI. Since both come from the same underlying student-level test score data, the resulting differences are minimal, and are attributable to two factors. First, for some countries data required to calculate the SES index are missing for some students, and so the SES-HCI overall average HLO is based on a slightly different set of students than the HLO used in the calculation of the global HCI. Second, in some cases it was not possible to obtain the distribution of test scores by SES centile for the test that was used in the global HCI, and so we used a distribution of test scores for another test for the same country where the SES-centile disaggregation was available.[19]

3.5 Fraction of Children Under 5 Not Stunted

In the SES-HCI, the health component is based on one indicator, the fraction of children under 5 who are not stunted, which is the complement of the under-5 stunting rate. We obtain stunting rates disaggregated by SES quintile from the HEFPI database, where they are calculated using 2006 WHO growth standards. Specifically, children are defined as stunted if their height-for-age is more than two standard deviations below the WHO Child Growth Standards median. Child-level stunting prevalence is averaged across the relevant 0-5 age range to arrive at an overall under-5 stunting rate.

Summary statistics on not-stunted rates are reported in the fourth panel of **Table 1**. The top panel of **Figure 4** shows the variation in stunting outcomes within and between countries. In some of the poorest countries in our sample, only half of children fall within the normal height range for their age, while in the richest countries in our sample over 90 percent of children do so. Gaps in stunting rates

[19] This accounts for the three most apparent outliers in the bottom panel of **Figure 3**. For Kyrgyzstan we use the 2009 PISA in the SES-HCI while the global HCI used the 2017 EGRA which resulted in a substantially higher score. Similarly for Malawi and Zambia, we used the 2015 and 2013 EGRAs, respectively, since student-level data for the 2013 SACMEQ used in the global HCI was not available. In both cases, the SACMEQ scores used in the global HCI were considerably higher than the EGRAs.

across SES quintiles are large in some countries: in Guatemala for example only around one-third of children in the bottom SES quintile are not stunted, while over 80 percent of children in the top SES quintile are not stunted.

The stunting data used in the global HCI exercise primarily comes from the UNICEF-WHO-World Bank Joint Malnutrition Estimates (JME) database. Since the DHS and MICS are among the primary data sources in the JME compilation, there are few differences in the stunting rates reported in the HEFPI and the JME databases. This is readily apparent in the bottom panel of **Figure 4**, where the data from the two different sources align almost perfectly along the 45-degree line.

4. Calculating the SES-HCI

The SES-HCI is calculated following the methodology of the global HCI. The HCI measures the expected productivity as a future worker of a child born today, assuming that the child faces the same risks of poor health and incomplete education that current cohorts of children face. It is calculated as the product of three terms:

(1) $$SHCI_q = Survival_q \times School_q \times Health_q$$

where $Survival_q$, $School_q$ and $Health_q$ indicate the three components of the SES-HCI for each quintile q. Each of the components is expressed relative to the benchmark of complete education and health and is defined as follows.

Child survival is measured as the probability of survival to age 5, p, relative to the benchmark of no child mortality over this age range, i.e. relative to the benchark of $p^* = 1$:

(2) $$Survival_q \equiv \frac{p_q}{p^*} = \frac{1 - Under\ 5\ Mortality\ Rate_q}{1}$$

Schooling is measured as the contribution to worker productivity of learning-adjusted expected years of school. Expected years of school is calculated as described above, as the sum of enrollment rates by age between ages 6 and 17. The adjustment for learning follows the HCI methodology, and is described in Filmer, Rogers, Angrist and Sabarwal (2018). It consists of discounting expected years of school by the ratio of observed HLO scores relative to the TIMSS international benchmark for high achievement of 625. For example, this means that a year of schooling obtained by a child who scores 300 is “worth” only half as much as a year of schooling obtained by a child who scores 625. Years of school are converted into contributions to worker productivity using a central estimate of the labour market return to schooling, which is $\phi = 0.08$ or 8 percent per additional year of school. Finally, this is expressed relative to the benchmark level of worker productivity corresponding to a full 12 years of high-quality school resulting in a test score of 625, i.e. a full 12 years of learning-adjusted school years:[20]

$$\text{(3)} \qquad School_q \equiv e^{\phi\left(Expected\ Years\ of\ School_q \times \frac{Harmonized\ Test\ Score_q}{625} - 12\right)}$$

Finally, the health component of the SES-HCI relies on stunting as a proxy for the health risks faced by children. This is linked to future worker productivity through two steps. First, children who face health risks that are manifested in stunting are also likely to be below-normal stature as adults. Second, there is a large literature that measures the effects of childhood health shocks on adult health and earnings, by estimating the effect of adult height on earnings. The combination of these two effects suggests that a 10 percentage point reduction in stunting rates would raise productivity as a future worker by 3.5 percent, i.e. the "return" to reduced stunting is $\gamma_{Stunting} = 0.35$. The benchmark of complete health corresponds to no stunting, or a not-stunted rate of 1. Putting these pieces together implies that the health component of the SES-HCI is:

$$\text{(4)} \qquad Health_q \equiv e^{\gamma_{Stunting} \times (Not\ Stunted\ Rate_q - 1)}$$

Multiplying together the three components results in the overall SES-HCI, which takes on values between zero and one, and represents the shortfall in expected productivity as a future worker of a child born today that is attributable to shortfalls in her expected health and education outcomes relative to the benchmark of complete high-quality education and full health.

Before turning to our main findings on gaps in SES-disaggregated human capital between and within countries, we first briefly summarize the differences between the global HCI and national averages of the SES-HCI. We calculate the national average of the SES-HCI by evaluating the HCI at the national

[20] Note that this differs slightly from the global HLO, where expected years of school is calculated over a 14-year age range, and so the benchmark of complete high-quality education is 14 years, not 12.

averages of its four components.[21] Then in **Figure 5**, we plot the national average SES-HCI (on the vertical axis) against the global HCI (on the horizontal axis). To isolate the first source of difference between the two, the global HCI on the horizontal axis is calculated using only stunting as the proxy for health, as is the case for the SES-HCI. As discussed in the previous section, national averages of child survival, stunting, and test scores used in the SES-HCI are very similar to their counterparts in the global HCI. This means that the differences between the SES-HCI and the global HCI displayed in **Figure 5** are primarily due to differences in expected years of school as calculated from survey data (as in the SES-HCI) as opposed to administrative data (as in the global HCI). These differences are manifested in two different ways in **Figure 5**. First, although the correlation across countries between the SES-HCI and the global HCI is high at 0.93, it is not perfect. This reflects the less-than-perfect correlation between expected years of school based on survey versus administrative data, as shown in the bottom panel of **Figure 2**. Second, recall that the SES-HCI calculates expected years of school over a shorter 12-year age range (age 6-17), while the HCI considers a 14-year age range (age 4-17). This means that the dispersion in expected years of school across countries is smaller in the SES-HCI because it does not capture cross-country differences in pre-primary school participation. Since the overall index values reflects gaps in human capital relative to the benchmark, these gaps also are smaller in the SES-HCI. This in turn means that the values of the SES-HCI are on average larger than in the HCI, as can be seen from the fact that nearly all countries are above the 45-degree line in **Figure 5**.

5. Results

5.1 Differences in Human Capital Between and Within Countries

Figure 6 reports the SES-HCI for the most recent cross-section of countries, plotting the SES-HCI for the top and bottom quintiles as well as the mean (on the vertical axis) against log GDP per capita (on the horizontal axis). Focusing first on the country-level average scores (indicated as solid circles), the SES-HCI ranges from between 0.3 to 0.4 in the lowest-performing countries (such as Niger, Mali and Chad) to around 0.7 in the best-performing countries such as Vietnam and Armenia. For the lowest-scoring countries, shortfalls in health and education imply that productivity as future workers of children born

[21] Note that the human capital index is a convex function of its components. As a result, the SES-HCI evaluated at the national averages of its component data will differ slightly from the average of the SES-HCI across quintiles due to Jensen's inequality.

today will only be 30 to 40 percent of what it could be had these children enjoyed complete education and full health, while in the best performing countries productivity will be 70 percent of what it could have been. As noted above, the SES-HCI covers primarily low and middle-income countries where DHS/MICS data are available. Although not strictly comparable for the reasons described above, we also show the global HCI in **Figure 6** (as light grey points), in order to help to visually situate the countries covered in the SES-HCI in the global distribution. Comparing these two measures, it is clear that the SES-HCI sample of countries represents roughly the poorest two-thirds of countries covered in the global HCI.

The most striking feature of **Figure 6** is the large gaps in human capital within countries, as measured by the range between the SES-HCI in the top versus bottom quintiles (indicated as vertical ranges on the graph). These gaps are apparent across the income spectrum. For example, in Madagascar the SES-HCI ranges from 0.40 in the poorest quintile to 0.58 in the richest quintile, while in much richer Vietnam the gap ranges from 0.58 to 0.85. In both cases, the within-country range between the top and bottom quintile is roughly half the size of the cross-country range between the highest and lowest country-average SES-HCI values in our sample.

How do these differences in human capital between richer and poorer people within a country contribute to overall differences in human capital around the world? To answer this question more systematically, we rely on a simple variance decomposition to quantify the relative importance of between versus within-country variation in the SES-HCI. Let h_{qc} denote the SES-HCI for quintile q in country $c = 1, \dots, C$; h_c denotes the country average of h_{qc}; and h denotes the global average across countries of h_c. Then the total variation (across quintiles q and countries c) of human capital outcomes as measured by the SES-HCI can be written as:

$$\text{(5)} \qquad \frac{1}{5C}\sum_{c=1}^{C}\sum_{q=1}^{5}\left(h_{qc}-h\right)^2 = \frac{1}{C}\sum_{c=1}^{C}\left(\frac{1}{5}\sum_{q=1}^{5}\left(h_{qc}-h_c\right)^2\right) + \frac{1}{C}\sum_{c=1}^{C}(h_c-h)^2$$

The first term on the right-hand side reflects variation across quintiles and is the cross-country average of the within-country variance across quintiles of the SES-HCI. The second term represents the variation across countries in the country-level average SES-HCI. The first column of **Table 2** presents the results of this variance decomposition for the HCI. It shows that 34 percent, or nearly one-third of the variation in human capital consists of differences between rich and poor groups within countries. The remaining columns of **Table 2** unbundle this by performing the same variance decomposition on the components of the HCI. Differences in child survival and test scores across groups within countries account for a relatively smaller share of the overall variation in these outcomes across countries and groups: 21 and 23 percent, respectively. In contrast, within-country rich-poor gaps in expected years of school and stunting account for a considerably larger share of the overall variation in these outcomes, at 31 and 33 percent, respectively.

5.2 Human Capital-Income Gradients Between and Within Countries

In this section we document the relationship between human capital differences and income differences within and between countries. To do so, we require estimates of income differences across the SES quintiles for which the SES-HCI is calculated. Since the SES quintiles are based on wealth or asset indexes in the underlying DHS/MICS and test score data rather than an absolute measure of income, this requires a further assumption and additional data. We assume that the ordering of households according to the SES index in the DHS/MICS is the same as the ordering of households according to their income or consumption expenditures.[22] We draw on the World Bank's PovcalNet database to retrieve the quintile shares in income or consumption (depending on the type of survey available) for the nearest household survey to each of the country-year observations in our SES-HCI dataset. Appendix Table 2 documents the source year for these data. We then combine these quintile shares with real per capita GDP to calculate a proxy for average per capita income within SES quintiles.[23]

Figure 7 shows the within-country gradients between log income and the SES-HCI across quintiles for the most recent cross-section of 51 countries. This gradient is of interest because it can be interpreted as a comparable measure of inequalities in human capital outcomes across income levels within a country. For example, if the slope within a country is zero, it means that human capital outcomes are equalized across income groups, while if the slope is steep it means that human capital outcomes among the rich are much better than among the poor. To make the graphs legible, we divide the countries in our sample

[22] Naturally this assumption is unlikely to be literally true. However, our goal is not to match individual households but rather to measure average incomes within each quintile. As long as most differences in the rank ordering of households by SES versus by income average out within quintiles, the quintile averages of income will be reasonable approximations. See also Filmer and Scott (2011) who document less-than-perfect rank correlations of income and SES indexes across individuals in a number of LSMS surveys, but nevertheless conclude that distributional analysis of health and education outcomes is robust to the use of either measure. Wagstaff and Watanabe (2003) arrive at similar conclusions, noting that analysis of socioeconomic inequalities in child malnutrition are broadly similar whether a measure of household income or an asset index is used.

[23] Specifically, log average income in quintile $q = 1, \ldots, 5$ is calculated as $\ln\left(\frac{s_q}{0.2}\right) + \ln(GDP\ Per\ Capita)$, where s_q is the income or expenditure share of quintile q.

into regional subgroups in the six panels of the graph. In each panel we plot log per capita income (on the horizontal axis) against the SES-HCI by quintiles (on the vertical axis). The five observations corresponding to the five quintiles for each country are connected by lines, and the slope of these lines shows the rate at which human capital increases with income across quintiles within countries. Not surprisingly, these lines are all upward-sloping, reflecting the pattern of higher income and higher human capital across SES quintiles. Interestingly, however, the slopes are generally fairly similar for most countries. To highlight the exceptions to this pattern, we identify the "steepest" and "flattest" country within regional grouping with green and red coloured lines. For example, in the top-left panel, the fairly flat green line for Haiti shows that Haiti has relatively small differences in human capital outcomes across SES quintiles despite quite large income differences across quintiles, while conversely the steep red line for Guatemala highlights its large differences in human capital outcomes across SES quintiles given its income inequality.

How do these within-country relationships between human capital outcomes and income levels compare with the corresponding pattern across countries? To answer this question, we pool the quintile-level data on the SES-HCI and log income per capita for all 51 countries in our cross-section of most recent data, and then estimate a "between" and a "within" regression of the former on the latter. The "between" slope captures the relationship between country-level average income and average SES-HCI across countries, while the "within" slope captures the average slope of this relationship across quintiles within countries. The results are shown in **Table 3**. For the full sample of 51 countries, the two slopes are remarkably similar, at 0.069 and 0.074, respectively. This finding is both surprising and somewhat discouraging. Redistributive and targeted policies within countries have the potential to reduce inequalities in health and education outcomes across income groups, while such redistribution is more difficult across countries. Thus, one might expect the within-country slope to be flatter than the between-country slope. In contrast, we find that the slopes are roughly the same, indicating that income-related differences in the HCI are on average as large within countries as they are across countries. Looking across columns in **Table 3**, we find that roughly the same pattern holds across countries in sub-Saharan Africa, and separately within poorer and richer countries in our sample. In all four cases, the 95% confidence interval around the between-country slope comfortably includes the within-country slope across quintiles.

In the remainder of **Table 3**, we unbundle this by estimating the same between and within regressions for the components of the SES-HCI. While the estimated slope coefficients are different across the different components of the SES-HCI (reflecting differences in the units of the underlying variables), the same general pattern holds: in most cases the slope of the relationship between countries is similar to the cross-country slope. This indicates that on average, these specific human capital outcomes are as unequal across income groups within a country as they are across countries. The one exception to this pattern is child survival rates, where the within-country gradient with income consistently is smaller than the between-country gradient with income.

5.3 What Accounts for Within-Country Human Capital Differences Across SES Quintiles?

In this subsection, we document patterns in, and covariates of, rich-poor gaps in the SES-HCI and its components. In this subsection and the following, we define these gaps between the top and bottom quintiles in a specific way to facilitate interpretation of the relationship between rich-poor gaps in the overall SES-HCI and rich-poor gaps in its components. To motivate our approach, take the logarithm of the SES-HCI as defined in Equations (1)-(4) to arrive at the following expression:

(6) $$\ln(h) = \ln(p) + \phi\left(eys \times \frac{hlo}{625} - 12\right) + \gamma \times nostu$$

where h represents the SES-HCI; p represents the child survival rate; eys represents expected years of school; hlo represents the harmonized test score; $nostu$ represents the fraction of children who are not stunted; and ϕ and γ are parameters measuring the labour market returns to education and health. Also define $qeys \equiv eys \times \frac{hlo}{625}$ as quality-adjusted expected years of school. Written in this way, the logarithm of the HCI is a weighted average of the log of child survival, and the levels of quality-adjusted years of school and the not-stunted rate. This in turn means that rich-poor gaps between the top and bottom quintile of the *logarithm* of the SES-HCI (i.e. the percent difference in the SES-HCI in the top quintile and the bottom quintile) will be a weighted average of gaps between the top and bottom quintile in the *logarithm* of child survival and the *levels* of quality-adjusted years of school and the not-stunted rate, with weights equal to one, $\phi = 0.08$, and $\gamma = 0.35$ as discussed in Section 4.[24] In addition, to allow unbundling

[24] We focus on these simple measures of rich-poor gaps in order to facilitate decomposition of rich-poor gaps in the overall SES-HCI into gaps in the components. However, we also considered three alternative measures of rich-poor gaps in the SES-HCI components: (a) the standard deviation across quintiles within a country, and (b) the slope of a regression of the component values by quintiles on income by quintiles within a country (c) the absolute gap in the top to bottom quintiles. These measures yield broadly similar results and are not reported for reasons of space. We also calculated the concentration index for health outcomes discussed in Wagstaff, Paci and Van Doorslaer (1991) and Kakwani, Wagstaff and Van Doorslaer (1997), for stunting and child mortality. This measure

of rich-poor gaps in quality-adjusted expected years of school, we define gaps in expected years of school and harmonized test scores as differences in *levels* of these variables.

In **Figure 8**, we plot the gaps between the top and bottom quintiles of the SES-HCI and its components against log GDP per capita for our most recent cross-section of 51 countries. This graph reveals quite clear patterns. Reading down the first column, Q5-Q1 gaps in expected years of school are smaller in rich countries than in poor countries. This is because school participation rates tend to be high in the richest quintile across all countries, while there are much more pronounced differences in school participation in the poorest quintile, that are narrower in richer countries. The opposite is true of test scores, where Q5-Q1 gaps tend to be larger in richer countries, although as discussed below this pattern needs to be interpreted with some caution. Combining these two patterns, the former dominates the latter, and Q5-Q1 gaps in quality-adjusted years of school tend to be smaller in rich countries than in poor countries. In the second column of **Figure 8**, we see that Q5-Q1 gaps in child survival and stunting both tend to be lower in rich countries than in poor countries. As with expected years of school, this is due to a much stronger improvement in these outcomes in the poorest quintile as we move from poorer to richer countries. Finally, the bottom-right panel of **Figure 8** shows the pattern in Q5-Q1 gaps for the (log) SES-HCI. As noted above, this pattern will reflect the weighted sum of the patterns for Q5-Q1 gaps in (log) child survival, quality-adjusted years of school, and stunting. These combine into a modest downward-sloping pattern, with smaller Q5-Q1 gaps in richer countries than in poorer countries.

To assess the statistical significance of these correlations, the odd-numbered columns of **Table 4** report the OLS regressions corresponding to the scatterplots **Figure 8**, a regression of the Q5-Q1 gap in the SES-HCI and its components on log GDP per capita, for the most recent cross-section of 51 countries. The pattern of narrowing Q5-Q1 gaps in richer countries for child survival and expected years of school, as well as the pattern of widening Q5-Q1 gaps for harmonized test scores, are all statistically significant at conventional levels. However, the weak negative relationship between the Q5-Q1 SES-HCI gap and per capita GDP across countries is not significant. In the even-numbered columns of **Table 4**, we further probe these cross-country patterns by including measures of country size (log population and log land area,[25] to allow for the possibility of larger Q5-Q1 gaps in human capital outcomes in larger countries) and income inequality (the log difference between the top and bottom quintile shares in income from the

of inequality in health outcomes is less strongly correlated with the simple rich-poor gap we focus on in the paper, but nevertheless also yields broadly similar cross-country patterns that we discuss below.

[25] These data are retrieved from World Bank (2019).

PovcalNet database discussed earlier, to allow for the possibility that Q5-Q1 gaps in human capital outcomes are larger in countries with higher income inequality). The partial correlation between country size and Q5-Q1 human capital gaps is positive in most cases, and significantly so for the overall SES-HCI and non-stunted rates when country size is measured as log population. Q5-Q1 gaps in human capital outcomes are larger in more unequal countries (with the exception of test scores where the correlation is negative but not significant). This intuitive pattern is statistically significant at the 10 percent level or better for the overall SES-HCI, child survival, and the not stunted rate.

Finally, one important qualification about the observed pattern of higher Q5-Q1 gaps in harmonized test scores in rich versus poor countries should be noted. On the one hand, this pattern could be due to underlying factors like the fact that tests in the poorer countries in our sample tend to focus on primary school, while tests in richer countries are more likely to cover secondary school-aged children. If individual differences in learning ability accumulate over time, this could contribute to the observed regularity of greater dispersion in test scores in rich countries. On the other hand, this pattern is also to some extent a consequence of the test score harmonization methodology. Tests are harmonized by (a) first rescaling testing data from individual testing programs to have mean 500 and standard deviation 100 across all students taking that test in all countries, and (b) developing a multiplicative "exchange rate" between testing programs that reflects the ratio of average performance of students in countries participating in two testing programs. This ratio is smaller than one for the testing programs in poorer countries. For example, for EGRA the scaling factor is 0.73 relative to the benchmark of PIRLS. This multiplicative adjustment factor reduces both the mean *and* the dispersion in harmonized test scores in EGRA relative to PIRLS.[26] This in turn contributes to the pattern of lower dispersion in harmonized test scores in poorer countries relative to richer countries. To document this importance of test-specific factors, we augment the regressions for Q5-Q1 gaps in harmonized test scores in columns 7 and 8 of **Table 4** with dummy variables for EGRA, LLECE, PASEC and SACMEQ testing programs (with the various PISA/TIMSS/PIRLS assessments as the omitted category). Doing so reduces the coefficient on log per capita GDP by over half, and it no longer is statistically significant.

[26] The same pattern is true for LLECE, SACMEQ and PASEC where the scaling factors by grade and subject typically fall in the 0.7 to 0.8 range.

5.4 Changes over Time in Human Capital Gaps

In the previous sections, we have explored how gaps in human capital outcomes between better and worse-off households vary across countries, using our most recent cross-section of 51 countries. In this final section, we exploit the more limited panel dimension of our dataset to examine patterns within countries over time in SES-disaggregated human capital outcomes. To do so, we construct a sample of 24 "spells" covering the longest available time span for each country with two or more data points for the SES-HCI and its components. To improve over-time comparability within each of these spells, we restrict attention to spells for which the first and last observation use data from the same testing program. For example, we only consider spells in which the first and last observations are both EGRAs. The full sample of country spells is indicated with highlighted rows in Appendix Table A2. Spells are of varying length: the median spell is 9 years long, and the shortest and longest spells are 4 and 16 years, respectively. To make changes over time comparable across spells of different lengths, we define all changes as annual average changes over the length of the spell.

Figure 9 provides a visual overview of these spells. Each panel refers to one of the components of the SES-HCI or the overall SES-HCI and plots the average annual change in the Q5-Q1 gap (on the vertical axis) against the average annual change in the country-level average across quintiles (on the horizontal axis). In all six panels, there is a modest downward-sloping relationship, indicating a weak tendency for Q5-Q1 gaps in human capital outcomes (on the vertical axis) to narrow as overall average human capital outcomes improve within countries over time (on the horizontal axis). With the exception of test scores, the majority of countries are in the bottom-right quadrants, i.e. where human capital outcomes improve and the Q5-Q1 gap declines. Nonetheless there are important exceptions, notably in the top-right and bottom-left quadrants where mean outcomes improve and dispersion increases or vice-versa.

In **Table 5** we examine correlates of the within-country changes over time in average human capital outcomes and rich-poor gaps in human capital outcomes displayed on the horizontal and vertical axes of **Figure 9**. The odd-numbered columns regress the average annual change in the country-level average SES-HCI and its components on (a) the average annual change in log per capita GDP (i.e. average annual growth in per capita GDP) and (b) the average annual change in inequality, measured as before as the log difference between the top and bottom quintile shares in income/consumption from the PovcalNet database. The even-numbered columns show the same regressions but using the average annual change in the Q5-Q1 gap as the dependent variable. The surprising finding is that within-country improvements in the country average SES-HCI are on average slower in countries with faster growth. This is primarily driven by a similar pattern in country-level average HLO scores, which tend to increase more slowly in faster growing countries. **Figure 10** plots the simple scatterplot between average annual changes in the country-level average SES-HCI (on the vertical axis) against average annual per capita GDP growth (on the horizontal axis).[27] From the graph, it is clear that the data point corresponding to Malawi in the top-left (with rapid improvement in the SES-HCI but slow growth) is influential in this relationship – dropping it reduces the significance of the overall negative correlation from the 1 percent level to the 10 percent level. Overall, though, this somewhat surprising pattern of slower improvements in the SES-HCI in faster-growing countries should be taken with a grain of salt, as the sample is quite small (just 22 countries, with a single "spell" in each country. Also, it is important to note that this correlation should not be misinterpreted as saying anything about the effects of human capital on growth, for at least two reasons: (a) the simple correlation naturally does not have a causal interpretation, and (b) economic theory predicts relationships between the stock of human capital of the entire working age population and economic growth, while the SES-HCI, like the global HCI, thanks to its focus on health and education outcomes among the young, measures only the potential future human capital of the next generation of workers, and not the current stock.

6. Discussion and Conclusions

Inequalities in health and education may reflect barriers to optimal investments in human capital. These inequalities matter intrinsically, and they also affect human capital accumulation of future generations and eventually the aggregate growth process. This paper builds on a large existing literature documenting these inequalities to construct an index of human capital of the next generation of workers, disaggregated by quintiles of socioeconomic status (SES-HCI). This measure follows the methodology of the World Bank global HCI that was launched in October 2018. The HCI measures the human capital that a child born today can expect to attain by her 18th birthday, given the risks of poor health and poor education prevailing in her country.[28] It is constructed by aggregating together indicators of health (child

[27] The sample excludes Zambia and the Democratic Republic of Congo, both of which have very high per capita GDP growth rates in excess of 10 percent per year over their spells, which makes them highly-influential in regressions that include them.
[28] The HCI was introduced in World Bank (2018a,b), and the methodology of the HCI is detailed in Kraay (2019).

survival rates, stunting rates, and adult survival rates) and of the quantity and quality of schooling (expected years of school, and international test scores) using the estimates of their economic returns. Relying on data from DHS/MICS surveys, we compute the SES-HCI for 51 mostly low- and middle-income countries.

We find that differences between socio-economic groups within countries are large, and that within-country variation accounts for roughly one-third of the total variation in human capital outcomes. It is also important to note that our results focus on inequalities across quintile-average outcomes, rather than inequalities across individual-level outcomes, and therefore represent a lower bound on overall inequality in human capital outcomes across individuals. Overall, human capital outcomes increase with income, but when documenting how they do so within and between countries, we find a striking pattern: human capital outcomes increase with income across countries at roughly the same rate as they do within countries, across SES quintiles.

Overall, the gap between human capital of the rich and of the poor decreases weakly with income. This overall trend masks heterogeneity in the behavior of individual measures of human capital. For example, the rich-poor gaps in child survival and expected years of school narrows as countries get richer. In contrast, rich-poor gaps in harmonized test scores tend to be higher in richer countries with higher test scores, although this latter finding is to some extent a consequence of the methodology used to harmonize test scores.

For a small set of countries, we are able to examine trends over time. We find some surprising evidence that improvements in country-level averages of human capital outcomes are slower in faster-growing countries. However, these latter findings must be taken with a grain of salt, since our sample consist of only 22 countries where we have comparable data on all of the components of the SES-HCI at two points in time.

Measuring inequalities in health and education outcomes can be a powerful tool to better target investments in the human capital of those who need it the most. This work documents the existence of large inequalities in human capital across socio-economic groups. Our evidence suggests that countries, especially those where the rich-poor gap has the steepest gradient with income, could make significant improvements in human capital by targeting the most disadvantaged. This implies an important role for governments in easing the constraints faced by the most disadvantaged to increase their – and ultimately their countries' - human capital outcomes. Cash transfers can lessen financial constraints and thus support better access to education and health services; targeted education policies can help to improve access to and quality of learning at all levels; universal health coverage can help prevent and better manage health shocks and the related financial consequences. Yet, we find no evidence that on average redistributive policies *within* countries, arguably more easily implementable than redistribution *across* countries, produce differential effects, because the rate at which inequalities in health and education outcomes are reduced with income is indistinguishable within and across countries. Thus, in addition to highlighting the urgent need to address inequalities in human capital outcomes, our work draws attention to the consonant need for more research on policies and programs that are effective at closing these gaps.

References

Abdul-Hamid, Husein and Syedah Aroob Iqbal (2019). "Inequity in Education: Trends and Lessons from 140 Countries". In progress.

Filmer, Deon and Lant Pritchett (1998). "The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries". *Population and Development Review*. 25(1):85-120.

Filmer, Deon and Lant Pritchett (2001). "Estimating Wealth Effects Without Expenditure Data – Or Tears: An Application to Educational Enrollments in States in India". *Demography*. 38(1):115-132.

Filmer, Deon, Halsey Rogers, Noam Angrist, and Shwetlena Sabarwal (2018). "Learning-Adjusted Years of Schooling (LAYS): Defining a New Macro Measure of Education." World Bank Policy Research Working Paper No. 8591.

Filmer, Deon and Kinnon Scott (2011). "Assessing Asset Indexes". Demography. 49:359-392.

Flabbi, Luca, Gatti, Roberta (2018). "A Primer on Human Capital," World Bank Policy Research Working Paper 8309.

García, Jorge Luis, James J. Heckman, Duncan Ermini Leaf, and María José Prados, (2016). "The Life-cycle Benefits of an Influential Early Childhood Program." NBER Working Paper w22993, National Bureau of Economic Research.

Kakwani, Nanak, Adam Wagstaff and Eddy Van Doorslaer (1997). "Socioeconomic inequalities in health: Measurement, computation and statistical inference." *Journal of Econometrics*. 77(1): 87-104.

Kraay, Aart (2019). "The World Bank Human Capital Index: A Guide." *World Bank Research Observer*. 34(1):1-33.

Masset, Edoardo (2016). "SYNCMRATES: Stata module to compute child mortality rates using synthetic cohort probabilities." Statistical Software Components S458149, Boston College Department of Economics.

Rossi, Federico (2019) "Human Capital and Macro-Economic Development: A Review of the Evidence." World Bank Policy Research Working Paper 8650.

Rutstein, Shea Oscar and Kiersten Johnson (2004). "The DHS Wealth Index". DHS Comparative Reports No. 6. (https://dhsprogram.com/publications/publication-cr6-comparative-reports.cfm).

Patrinos, Harry Anthony, and Noam Angrist. 2018. "A Global Dataset on Education Quality: A Review and an Update (1965–2018)." World Bank Policy Research Working Paper No. 8592.

United Nations Inter-Agency Group for Child Mortality Estimation (UN IGME) (2019). Explanatory Notes: Child mortality trend series to 2018. New York: UNICEF, 2018. (https://childmortality.org/wp-content/uploads/2019/09/UNIGME-Explanatory-Notes_ENGLISH.pdf)

Urquiola, Miguel and Valentina Calderón (2006). "Apples and oranges: Educational enrollment and attainment across countries in Latin America and the Caribbean". *International Journal of Educational Development*, 26(6), 572-590.

Wagstaff, Adam, Patrick Eozenou, Sven Neelsen and Marc-Francois Smitz (2019). “Introducing the World Bank's 2018 Health Equity and Financial Protection Indicators Database,” *The Lancet*, Comment. 7(1): E22-E23.

Wagstaff, Adam (2000). “Socioeconomic Inequalities in Child Mortality: Comparisons Across Nine Developing Countries”. *Bulletin of the World Health Organization*. 78(1).

Wagstaff, Adam (2002). “Child Health on a Dollar a Day: Some Tentative Cross-Country Comparisons”. *Social Science and Medicine*. 57:1529-1538.

Wagstaff, Adam and Naoko Watanabe (2003). "What difference does the choice of SES make in health inequality measurement?" *Health Economics* 12(10): 885-890.

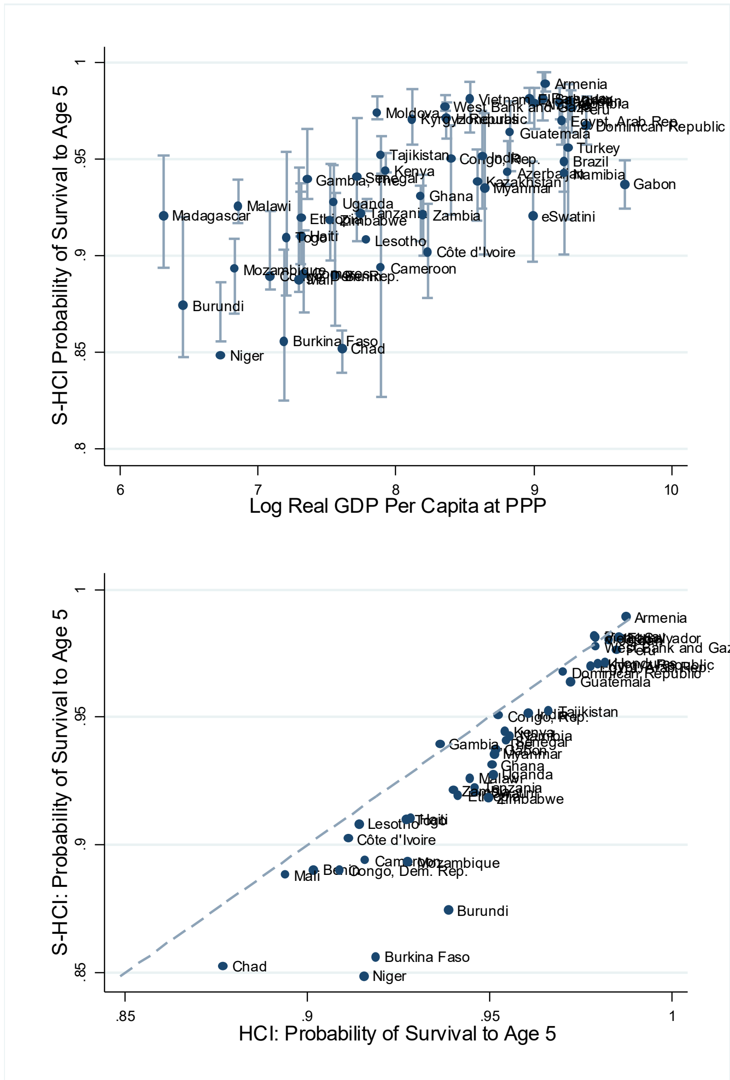
Wagstaff, Adam, Pierella Paci and Eddy Van Doorslaer (1991). "On the Measurement of Inequalities in Health." *Social Science & Medicine* 33(5): 545-557.

World Bank (2018a). World Development Report: The Changing Nature of Work. Washington, World Bank.

World Bank (2018b). “The Human Capital Project”. Washington, World Bank.

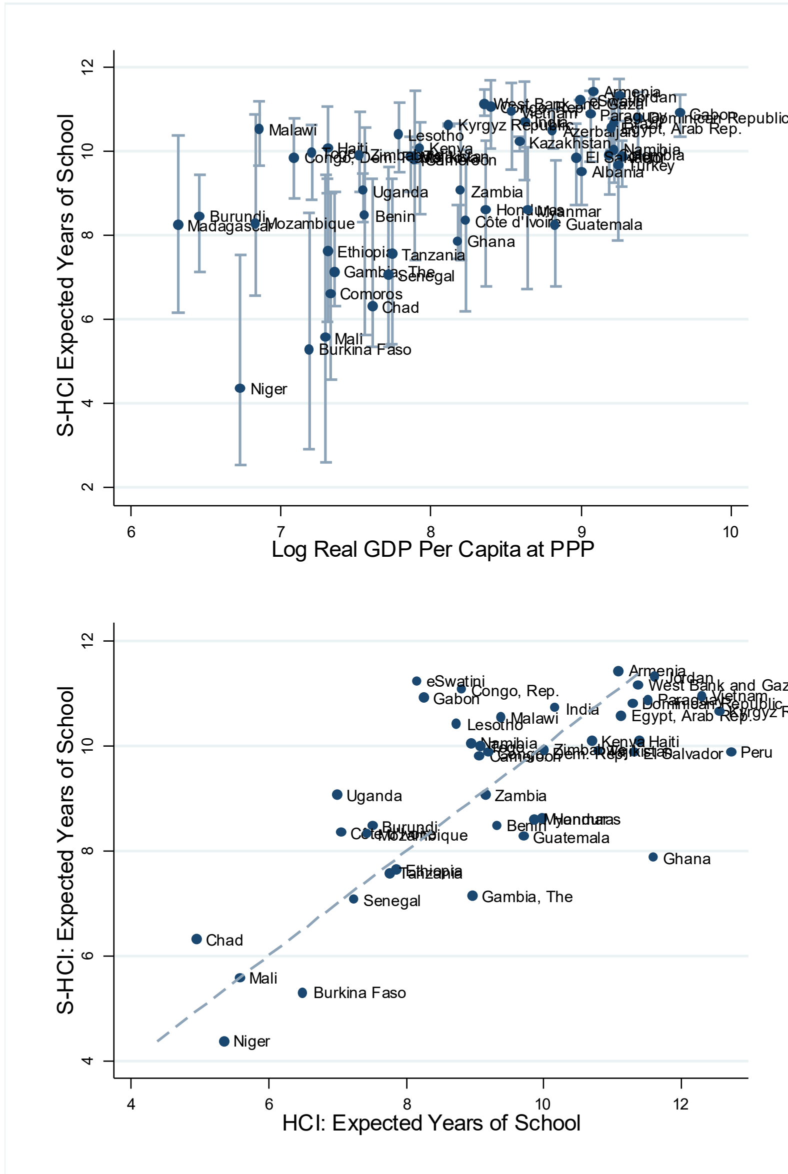
World Bank (2019). World Development Indicators. Washington, World Bank. https://datacatalog.worldbank.org/dataset/world-development-indicators

Figure 1: Child Survival



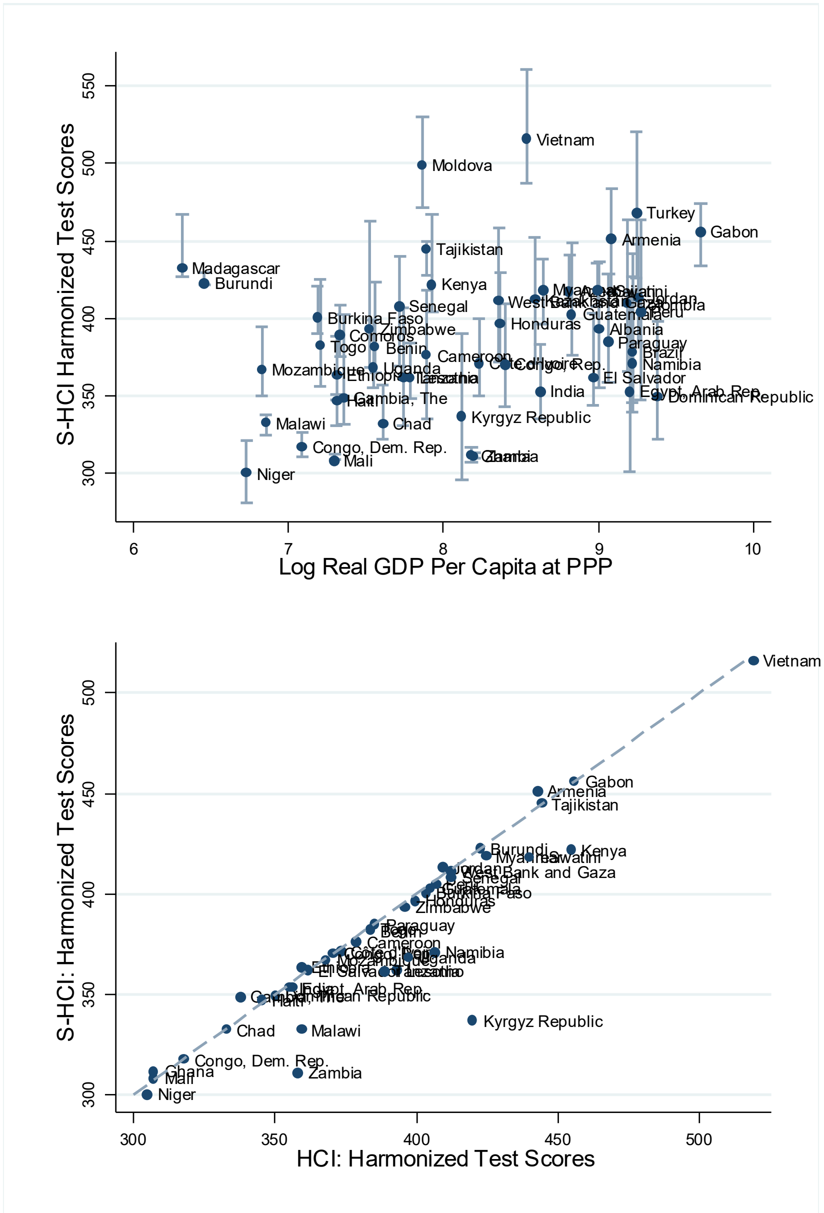
Notes: The top panel reports child survival rates disaggregated by SES quintile (on the vertical axis) against log real GDP per capita (on the horizontal axis) for the most recent cross-section of 51 countries in the SES-HCI dataset. Child survival is defined as probability of survival until age 5. The solid dot indicates the average across quintiles, and the top (bottom) end of the vertical bar indicates the value for the top (bottom) quintile. Note that the country average (dot) may fall below the value for the lowest quintile (bottom end of vertical bar) when middle quintiles have values below the lower quintile (for example, Niger). The bottom panel plots child survival as used in the SES-HCI (on the vertical axis) against child survival as used in the global HCI (on the horizontal axis), for the 42 countries in the most recent cross-section of countries in the SES-HCI dataset for which the SES-HCI data refer to 2010 or later. The dashed line is the 45-degree line.

Figure 2: Expected Years of School



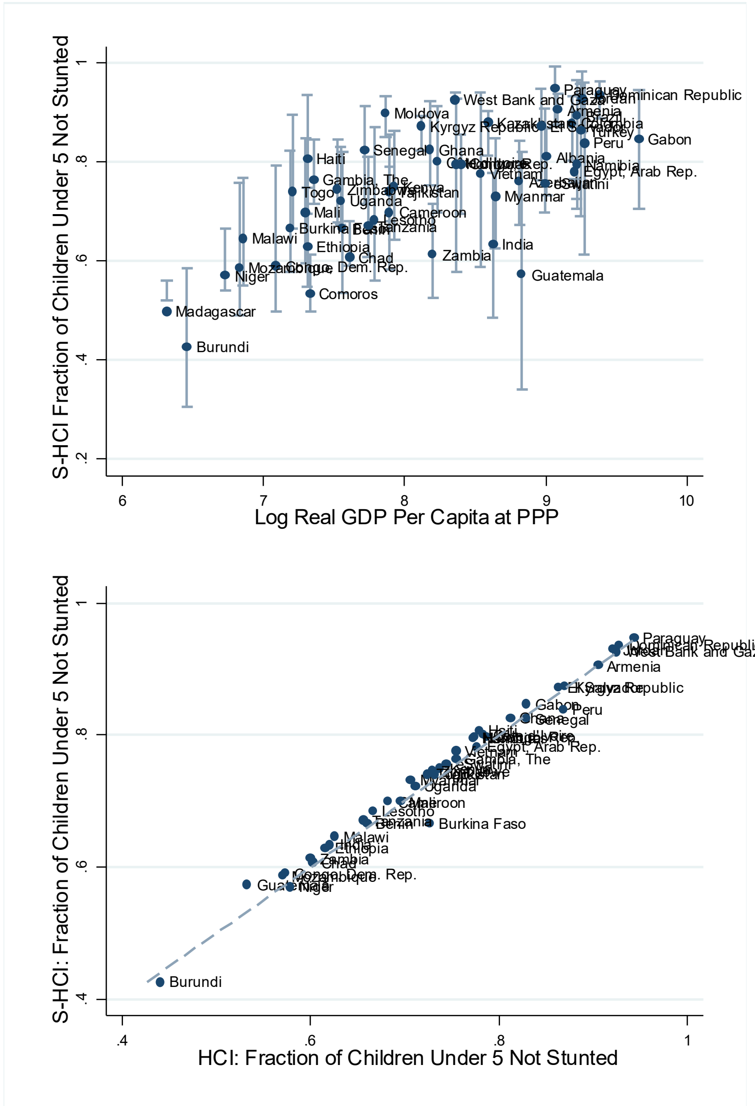
Notes: The top panel reports expected years of school disaggregated by SES quintile (on the vertical axis) against log real GDP per capita (on the horizontal axis) for the most recent cross-section of 51 countries in the SES-HCI dataset. Expected years of school is defined as the sum of enrollment rates by age between ages 6 and 17, for a maximum of 12 years. The solid dot indicates the average across quintiles, and the top (bottom) end of the vertical bar indicates the value for the top (bottom) quintile. The bottom panel plots expected years of school as used in the SES-HCI (on the vertical axis) against expected years of school as used in the global HCI (on the horizontal axis), for the 42 countries in the most recent cross-section of countries in the SES-HCI dataset for which the SES-HCI data refer to 2010 or later. The dashed line is the 45-degree line.

Figure 3: Harmonized Learning Outcomes



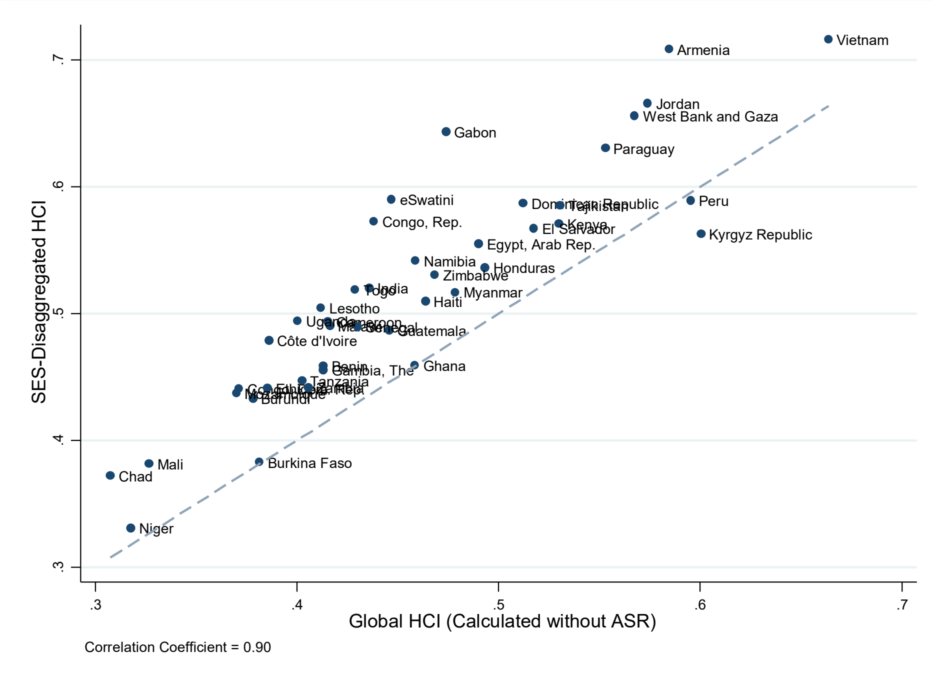
Notes: The top panel reports harmonized learning outcomes disaggregated by SES quintile (on the vertical axis) against log real GDP per capita (on the horizontal axis) for the most recent cross-section of 51 countries in the SES-HCI dataset. Harmonized learning outcomes are measured in TIMSS-equivalent units with 400 corresponding to minimal proficiency and 625 corresponding to high proficiency. The solid dot indicates the average across quintiles, and the top (bottom) end of the vertical bar indicates the value for the top (bottom) quintile. The bottom panel plots harmonized learning outcomes as used in the SES-HCI (on the vertical axis) against harmonized learning outcomes as used in the global HCI (on the horizontal axis), for the 42 countries in the most recent cross-section of countries in the SES-HCI dataset for which the SES-HCI data refer to 2010 or later. The dashed line is the 45-degree line.

Figure 4: Fraction of Children Under 5 Who Are Not Stunted



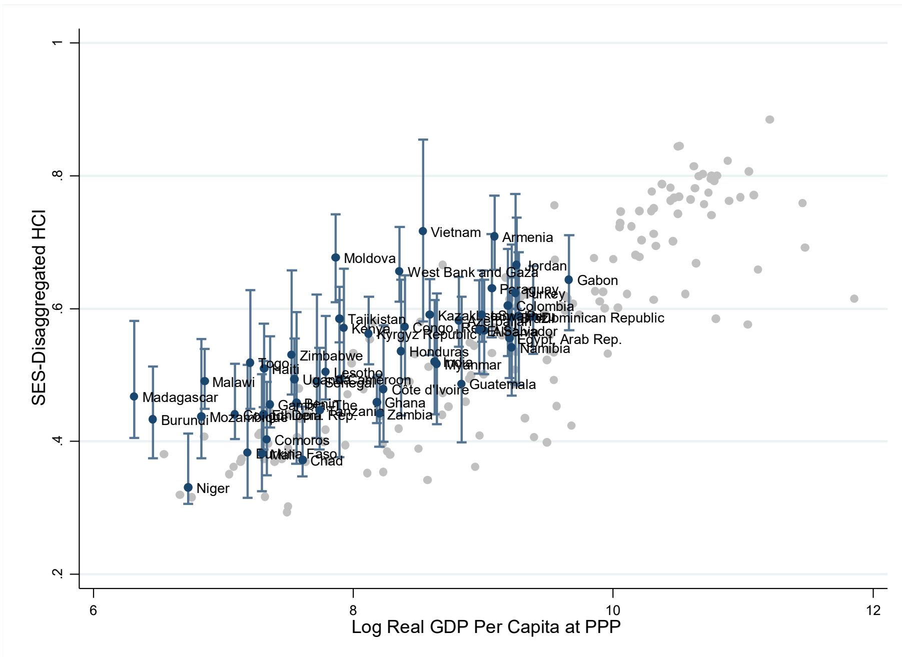
Notes: The top panel reports the fraction of children under 5 who are not stunted disaggregated by SES quintile (on the vertical axis) against log real GDP per capita (on the horizontal axis) for the most recent cross-section of 51 countries in the SES-HCI dataset. Stunting is defined as a child being more than two standard deviations below the reference mean height-for-age according to the WHO child growth standards. The solid dot indicates the average across quintiles, and the top (bottom) end of the vertical bar indicates the value for the top (bottom) quintile. Note that the country average (dot) may fall below the value for the lowest quintile (bottom end of vertical bar) when middle quintiles have values below the lower quintile (for example, Madagascar).The bottom panel plots the not stunted rate as used in the SES-HCI (on the vertical axis) against the not stunted rate as used in the global HCI (on the horizontal axis), for the 42 countries in the most recent cross-section of countries in the SES-HCI dataset for which the SES-HCI data refer to 2010 or later. The dashed line is the 45-degree line.

Figure 5: Comparing the SES-HCI and the HCI



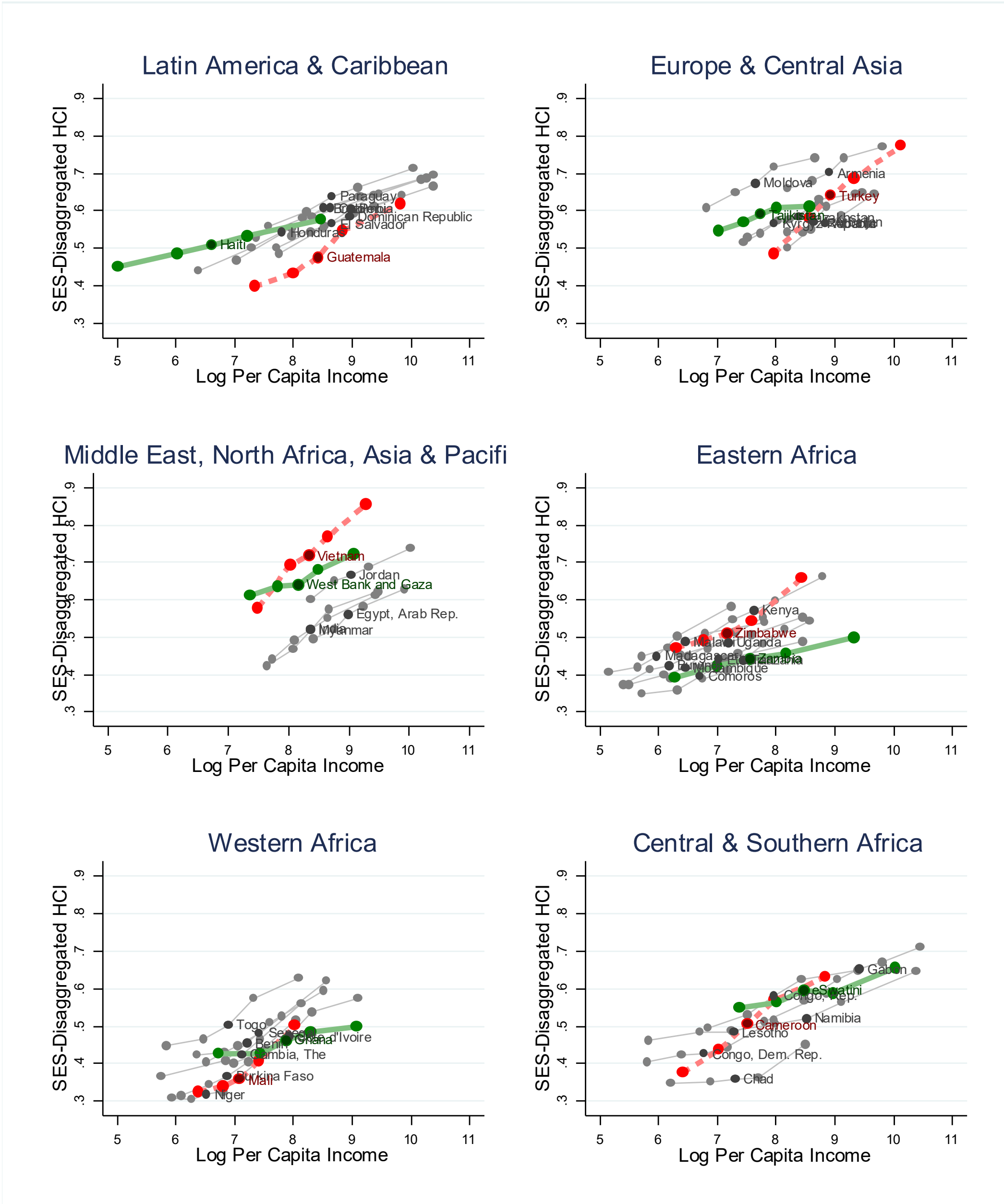
Notes: This graph compares the overall HCI disaggregated by quintiles of socioeconomic status (on the vertical axis) with the global HCI (on the horizontal axis) for the 42 countries in the most recent cross-section of countries in the SES-HCI dataset for which the SES-HCI data refer to 2010 or later. The dashed line is the 45-degree line. For comparison purposes, the global HCI is calculated excluding adult survival rates (ASR) as a proxy for health, in order to be more consistent with the SES-HCI which also does not include ASR.

Figure 6: SES-Disaggregated Human Capital Index



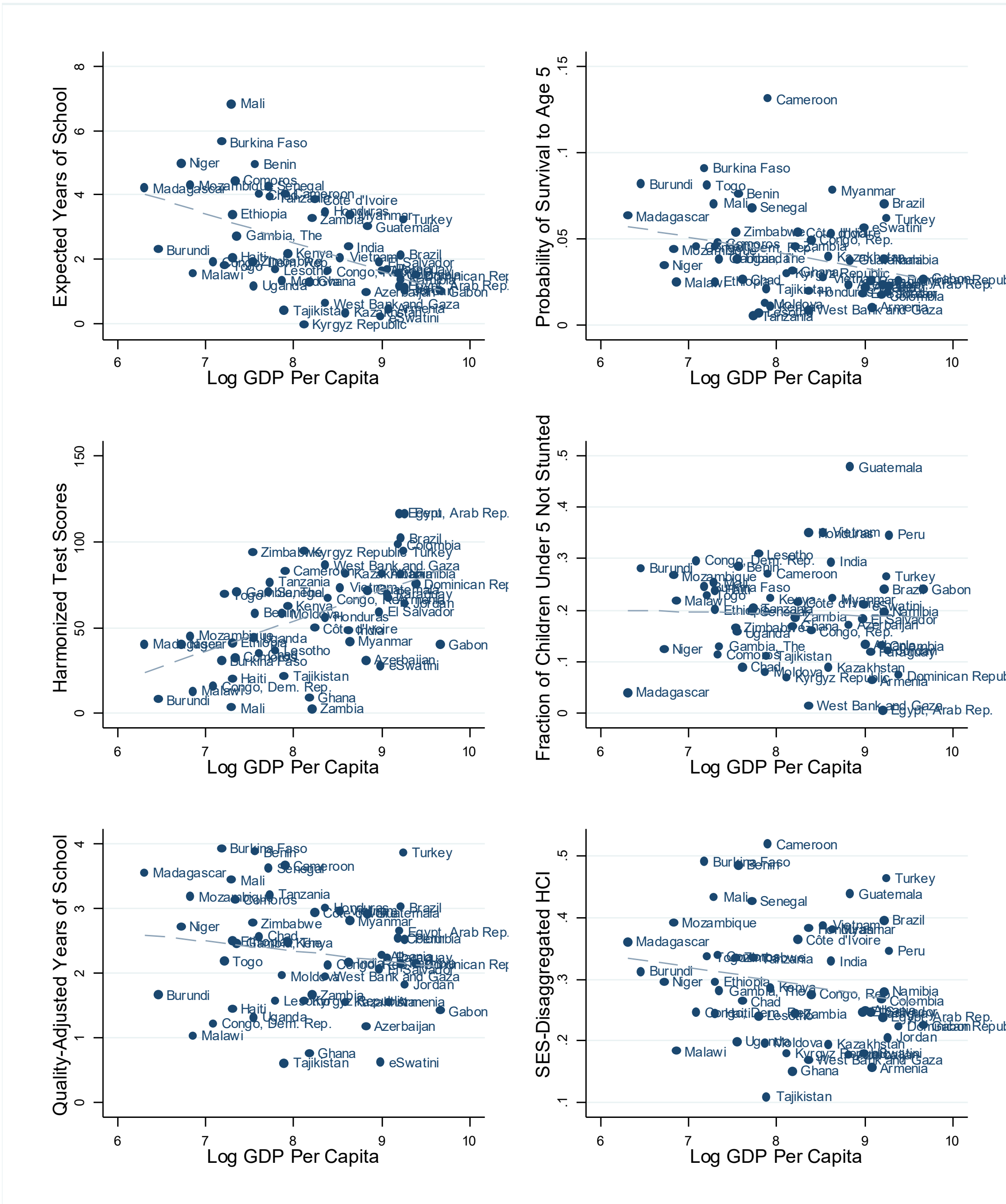
Notes: This graph plots the HCI disaggregated by quintiles of socioeconomic status (on the vertical axis) against log real GDP per capita (on the horizontal axis) for the most recent cross-section of 51 countries in the SES-HCI dataset. The solid dot indicates the average across quintiles, and the top (bottom) end of the vertical bar indicates the value for the top (bottom) quintile. The light grey data points show the global HCI for countries for which the SES-disaggregated HCI is not available.

Figure 7: Human Capital – Income Gradients Within Countries



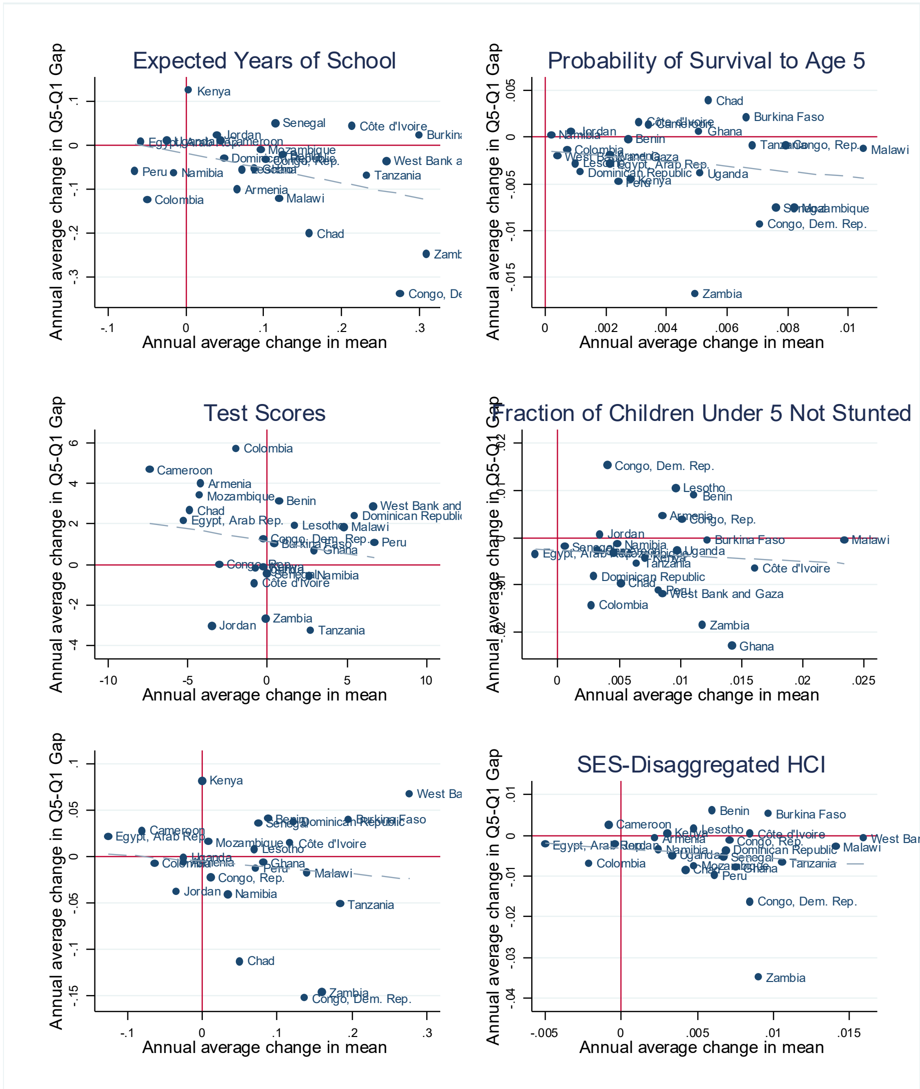
Notes: This graph reports the within-country relationship across quintiles between the SES-disaggregated HCI and log per capita income. Per capita income in each quintile is approximated using the quintile share in income or consumption as reported in the PovcalNet database for the survey nearest to the SES-HCI data, together with GDP per capita as the mean. The upward-sloping lines in each panel trace out the five quintile values for each country. The heavy solid green line (heavy dashed red line) shows the country in each group with the flattest (steepest) within-country gradient between the SES-HCI and log income per capita.

Figure 8: Rich-Poor Gaps in Human Capital Across Countries



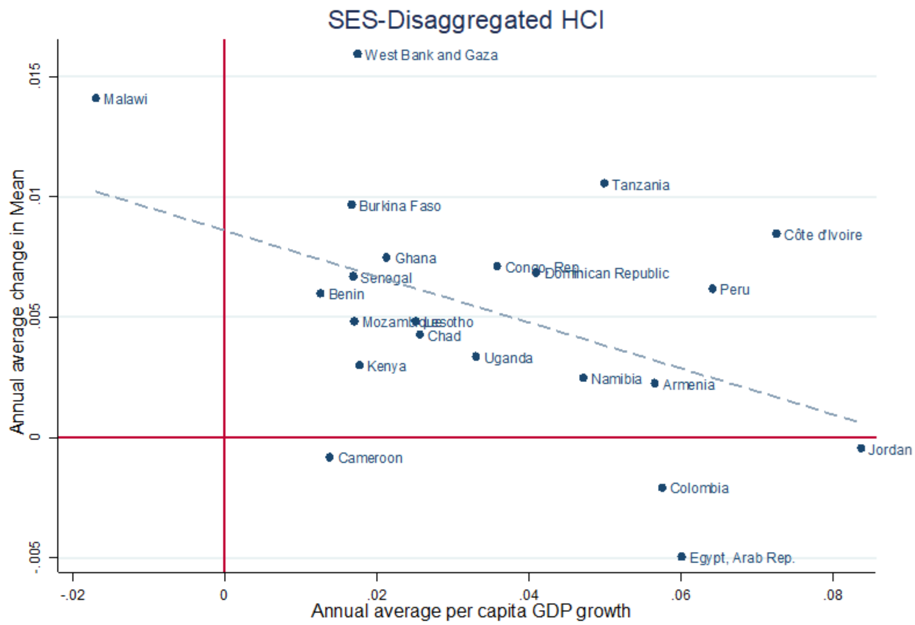
Notes: This graph plots gaps in human capital outcomes between the top and bottom quintiles (on the vertical axis) against log GDP per capita (on the horizontal axis), for the most recent cross-section of 51 countries in the SES-HCI dataset. The Q5-Q1 gaps are defined as (a) the difference between the top and bottom quintiles (for expected years of school, harmonized test scores, quality-adjusted years of school, and the not-stunted rate), and (b) the log-difference between the top and bottom quintiles for child survival, and for the overall HCI.

Figure 9: Changes over Time in Human Capital



Notes: This graph plots the within-country average annual change in the Q5-Q1 gap in the SES-HCI and its components (on the vertical axis) against the average annual change in the country-level average of the SES-HCI and its components (on the horizontal axis). Changes are reported over the longest-available time span for 24 countries with SES-HCI data at two points in time, and for which the initial and terminal period test scores come from the same testing program to ensure over-time comparability. The Q5-Q1 gaps are defined as (a) the difference between the top and bottom quintiles for expected years of school, harmonized test scores, quality-adjusted years of school, and the not-stunted rate, and (b) the log-difference between the top and bottom quintiles for child survival and for the overall HCI. Horizontal (vertical) red lines indicate zero on each axis, to facilitate identifying positive and negative changes.

Figure 10: Changes over Time in Human Capital



Notes: This graph plots the average annual change in the logarithm of the country-level average SES-HCI (on the vertical axis) against the average annual change in log per capita GDP (on the horizontal axis). Changes are reported over the longest-available time span for 22 countries with SES-HCI data at two points in time, and for which the initial and terminal period test scores come from the same testing program to ensure over-time comparability. Horizontal (vertical) red lines indicate zero on each axis, to facilitate identifying positive and negative changes.

Table 1: Summary Statistics

	Q1	Q2	Q3	Q4	Q5
Probability of Survival to Age 5					
Mean	0.92	0.93	0.93	0.94	0.96
SD	0.04	0.04	0.04	0.04	0.03
P10	0.86	0.86	0.88	0.88	0.92
P50	0.92	0.93	0.93	0.94	0.96
P90	0.97	0.98	0.98	0.98	0.99
Expected Year of School Between Ages 6-17					
Mean	8.07	8.78	9.29	9.79	10.46
SD	2.19	2.03	1.88	1.57	0.95
P10	5.36	5.98	6.64	7.55	9.35
P50	8.86	9.52	10.02	10.22	10.64
P90	10.36	10.77	11.19	11.31	11.47
Harmonized Learning Outcomes					
Mean	363.38	373.07	381.27	393.86	419.72
SD	44.41	46.41	46.76	49.02	53.22
P10	309.97	313.74	327.54	336.34	337.92
P50	355.68	360.47	374.37	394.45	425.00
P90	426.42	433.99	446.61	452.54	467.33
Fraction of Children Under 5 Not Stunted					
Mean	0.66	0.71	0.74	0.79	0.85
SD	0.13	0.14	0.14	0.13	0.10
P10	0.50	0.52	0.57	0.60	0.71
P50	0.67	0.72	0.75	0.80	0.86
P90	0.85	0.89	0.90	0.94	0.96
SES-Disaggregated Human Capital Index					
Mean	0.46	0.50	0.52	0.56	0.62
SD	0.08	0.09	0.09	0.10	0.09
P10	0.37	0.40	0.42	0.44	0.50
P50	0.45	0.49	0.52	0.57	0.62
P90	0.57	0.63	0.64	0.68	0.72

Notes: This table presents summary statistics on the overall SES-HCI and its components. Summary statistics refer to the most-recent cross-sectional data for 51 countries. Q1 (Q5) refer to the lowest (highest) SES quintiles. SD refers to standard deviation across countries, and P10, P50, and P90 refer to the 10^{th}, 50^{th} and 90^{th} percentiles across countries.

Table 2: Decomposing the Variance of the SES-HCI

	S-HCI	Probability of Survival to Age 5	Expected Years of School	Harmonized Test Scores	Fraction of Children Under 5 Not Stunted
Variance within countries	40.09	3.55	1.28	635.94	75.64
Variance across countries	78.50	13.61	2.81	2178.49	153.14
Within-country share of variance	**0.34**	**0.21**	**0.31**	**0.23**	**0.33**

Notes: This table reports a decomposition of total variance of the SES-HCI and its components into a within-country component and an across-country component. The first row reports the average across countries of the within-country variance across SES quintiles. The second row reports the variance across countries of the country averages. The sum of these is the total variance, and the last row of the table reports the within-country share of the total variance. Note that the HCI, child survival, and stunting rates are expressed as percentages (i.e. 0-100) for this calculation.

Table 3: Human Capital – Income Gradients Between and Within Countries

	All Countries		Sub-Saharan Africa		High Income		Low Income	
	Within	***Between***	***Within***	***Between***	***Within***	***Between***	***Within***	***Between***
Human Capital Index	0.069***	0.074***	0.066***	0.064***	0.070***	0.085***	0.069***	0.079**
	(0.002)	(0.010)	(0.003)	(0.014)	(0.003)	(0.026)	(0.003)	(0.031)
Probability of Survival to Age 5	0.017***	0.031***	0.019***	0.019***	0.015***	0.020*	0.020***	0.032**
	(0.001)	(0.004)	(0.002)	(0.007)	(0.001)	(0.010)	(0.002)	(0.013)
Expected Years of School	1.066***	1.095***	1.276***	1.085**	0.777***	1.077**	1.407***	0.653
	(0.057)	(0.227)	(0.089)	(0.438)	(0.051)	(0.413)	(0.097)	(0.824)
Harmonized Test Scores	24.907***	17.593**	19.837***	11.961	28.937***	44.516**	20.162***	25.583
	(1.167)	(7.241)	(1.389)	(10.519)	(1.611)	(18.845)	(1.569)	(20.875)
Fraction of Children Not Stunted	0.089***	0.103***	0.091***	0.111***	0.086***	0.078*	0.092***	0.165***
	(0.004)	(0.014)	(0.004)	(0.018)	(0.006)	(0.040)	(0.005)	(0.036)
Number of Observations	255	255	140	140	135	135	120	120
Number of Countries	51	51	28	28	27	27	24	24

Notes: This table reports the slope coefficient from a regression of the SES-disaggregated HCI (first row) and its components (next four rows) on log per capita income. Regressions are estimated for all countries, for countries inSub-Saharan Africa, for countries in the richest half of the sample (High Income), and for countries in the poorest half of the sample (Low Income). Within each country grouping, the "Within" column reports the slope of the within-country regression across quintiles (i.e. a regression with country fixed effects), and the "Between" column reports the slope of the regression of country averages of the dependent variable on country averages of log per capita income. Per capita income in each quintile is approximated using the quintile share in income or consumption as reported in the PovcalNet database for the survey nearest to the SES-HCI data, together with GDP per capita as the mean. The full sample in the first two columns consists of the cross-section of the most recent observation for each country. Robust standard errors reported in parentheses below point estimates. * (**) (***) indicate significance at the 10 (5) (1) percent level.

Table 4: Correlates of Human Capital Gaps

	Dependent Variable is Q5-Q1 Gap In:											
	SES-Disaggregated HCI		**Probability of Survival to Age 5**		**Expected Years of School**		**Harmonized Test Scores**		**Quality-Adjusted Years of School**		**Fraction of Children Not Stunted**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log GDP per capita	-0.022	-0.025*	-0.009***	-0.010***	-0.900***	-0.944***	17.622***	17.637***	-0.148	-0.164	-0.004	-0.005
	(0.013)	(0.013)	(0.003)	(0.003)	(0.198)	(0.198)	(3.850)	(3.769)	(0.139)	(0.133)	(0.016)	(0.017)
Log population		0.019*		0.001		0.029		1.772		0.100		0.028**
		(0.011)		(0.002)		(0.175)		(3.971)		(0.106)		(0.013)
Log land area		0.006		0.002		0.201		1.064		0.076		-0.008
		(0.009)		(0.002)		(0.175)		(2.896)		(0.088)		(0.011)
Inequality		0.049**		0.009*		0.384		-3.082		0.186		0.072***
		(0.023)		(0.005)		(0.357)		(7.229)		(0.220)		(0.022)
Constant	0.477***	0.007	0.115***	0.056	9.711***	6.303**	-87.050***	-122.676**	3.534***	0.691	0.225*	-0.298
	(0.110)	(0.178)	(0.027)	(0.045)	(1.711)	(2.904)	(30.977)	(59.858)	(1.165)	(1.768)	(0.126)	(0.218)
Observations	51	51	51	51	51	51	51	51	51	51	51	51
R-squared	0.039	0.217	0.097	0.171	0.258	0.337	0.277	0.300	0.022	0.117	0.001	0.233

Notes: This table presents cross-country regressions of the Q5-Q1 gap in the SES-disaggregated HCI and its components on log GDP per capita, log population, log land area, and income inequality (measured as log difference between average income in the top versus bottom quintile, with per capita income in each quintile is approximated using the quintile share in income or consumption as reported in the PovcalNet database for the survey nearest to the SES-HCI data, together with GDP per capita as the mean). The Q5-Q1 gaps are defined as (a) the difference between the top and bottom quintiles for expected years of school, harmonized test scores, quality-adjusted years of school, and the not-stunted rate, and (b) the log-difference between the top and bottom quintiles for child survival and for the overall HCI. The sample consists of the cross-section of the most recent observation for each country. Robust standard errors reported in parentheses below point estimates. * (**) (***) indicate significance at the 10 (5) (1) percent level.

Table 5: Changes over Time in Human Capital

Dependent Variable is Average Annual Change in:	SES-Disaggregated HCI		Probability of Survival to Age 5		Expected Years of School		Harmonized Test Scores		Quality-Adjusted Years of Scool		Fraction of Children Not Stunted	
	Mean	*Gap*	*Mean*	*Gap*	*Mean*	*Gap*	*Mean*	*Gap*	*Mean*	*Gap*	*Mean*	*Gap*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average Annual Growth	-0.111***	-0.031	-0.072***	0.018	-1.413	-0.014	-40.328	-24.377	-1.554*	-0.574	-0.102	-0.009
in Per Capita GDP	(0.038)	(0.036)	(0.020)	(0.026)	(0.946)	(0.759)	(29.886)	(20.493)	(0.760)	(0.382)	(0.064)	(0.061)
Average Annual Change	-0.051*	0.033	-0.018	0.020	-0.229	-0.352	-44.014	15.778*	-0.799	-0.190	-0.048	0.079*
in Inequality	(0.029)	(0.036)	(0.013)	(0.016)	(0.583)	(0.526)	(32.919)	(9.001)	(0.655)	(0.380)	(0.049)	(0.038)
Constant	0.009***	-0.002	0.006***	-0.002*	0.135***	-0.030	1.430	2.148***	0.109***	0.024	0.011***	-0.004
	(0.001)	(0.002)	(0.001)	(0.001)	(0.037)	(0.035)	(1.291)	(0.581)	(0.033)	(0.019)	(0.003)	(0.003)
Observations	22	22	22	22	22	22	22	22	22	22	22	22
R-squared	0.306	0.101	0.322	0.051	0.106	0.023	0.132	0.137	0.166	0.102	0.207	0.108

Notes: This table reports the results of regressing the average annual change in the country-level average of the SES-HCI and its components on the average annual changes in log per capita GDP and inequality (even-numbered columns); and the results of regressing the within-country average annual change in the Q5-Q1 gap in the SES-HCI and its components on the average annual changes in log per capita GDP and inequality (odd-numbered columns). Changes are calculated over the longest-available time span for 22 countries with SES-HCI data at two points in time, and for which the initial and terminal period test scores come from the same testing program to ensure over-time comparability. The Q5-Q1 gaps are defined as (a) the difference between the top and bottom quintiles for expected years of school, harmonized test scores, quality-adjusted years of school, and the not-stunted rate, and (b) the log-difference between the top and bottom quintiles for child survival and for the overall HCI. Spells for two countries (Democratic Republic of Congo and Zambia) are excluded because they register very high average annual per capita GDP growth rates in excess of 10 percent per year over their respective spells, making them highly-influential in the regressions. * (**) (***) indicate significance at the 10 (5) (1) percent level.

Appendix Table A1: Country Coverage

	Low income	Lower middle income	Upper middle income	Total
East Asia & Pacific	..	2 (2)	..	2 (2)
Europe & Central Asia	..	4 (6)	4 (5)	8 (11)
Latin America & Caribbean	1 (1)	3 (4)	5 (10)	9 (15)
Middle East & North Africa	..	3 (8)	..	3 (8)
South Asia	..	1 (1)	..	1 (1)
Sub-Saharan Africa	18 (30)	8 (18)	2 (3)	28 (51)
Total	**19 (31)**	**21 (39)**	**11 (18)**	**51 (88)**

Notes: This table reports the country coverage of the SES-disaggregated HCI. The first number in each cell reports the number of countries in the most recent cross section. The second number (in parentheses) in each cell reports the total number of country-year observations

Appendix Table A2: Country Sample and Data Sources

	Year in S-HCI	Child Survival		Stunting		Expected Years of School		Harmonized Test Scores		PovcalNet
Country	**Dataset**	***Year***	***Source***	***Year***	***Source***	***Year***	***Source***	***Year***	***Source***	***Year***
Albania	2008	2008	DHS	2008	DHS	2005	MICS	2009	PISA	2008
Armenia	2000	2000	DHS	2000	DHS	2000	DHS	2003	TIMSS	2001
Armenia	2005	2005	DHS	2005	DHS	2005	DHS	2007	TIMSS	2005
Armenia	2015	2015	DHS	2015	DHS	2015	DHS	2011	TIMSS	2015
Azerbaijan	2006	2006	DHS	2006	DHS	2006	DHS	2006	PISA	2005
Burundi	2010	2010	DHS	2010	DHS	2010	DHS	2014	PASEC	2014
Benin	2006	2006	DHS	2006	DHS	2006	DHS	2006	PASEC	2003
Benin	2014	2014	MICS	2014	MICS	2014	MICS	2014	PASEC	2015
Burkina Faso	2003	2003	DHS	2003	DHS	2003	DHS	2006	PASEC	2003
Burkina Faso	2010	2010	DHS	2010	DHS	2010	DHS	2014	PASEC	2009
Brazil	1996	1996	DHS	1996	DHS	1996	DHS	2000	PISA	1996
Côte d'Ivoire	2011	2011	DHS	2011	DHS	2011	DHS	2006	PASEC	2008
Côte d'Ivoire	2016	2016	MICS	2016	MICS	2016	MICS	2014	PASEC	2015
Cameroon	2004	2004	DHS	2004	DHS	2004	DHS	2006	PASEC	2007
Cameroon	2014	2014	MICS	2014	MICS	2014	MICS	2014	PASEC	2014
Congo, Dem. Rep.	2007	2007	DHS	2007	DHS	2007	DHS	2010	EGRA	2005
Congo, Dem. Rep.	2013	2013	DHS	2013	DHS	2013	DHS	2012	EGRA	2012
Congo, Rep.	2005	2005	DHS	2005	DHS	2005	DHS	2006	PASEC	2005
Congo, Rep.	2014	2014	MICS	2014	MICS	2014	MICS	2014	PASEC	2011
Colombia	2004	2004	DHS	2004	DHS	2005	DHS	2003	PIRLS	2004
Colombia	2009	2009	DHS	2009	DHS	2010	DHS	2009	PISA	2009
Comoros	2000	1996	DHS	2000	MICS	1996	DHS	2006	PASEC	2004
Dominican Republic	2002	2002	DHS	2002	DHS	2000	DHS	2006	LLECE	2002
Dominican Republic	2007	2007	DHS	2007	DHS	2007	DHS	2013	LLECE	2007
Dominican Republic	2013	2013	DHS	2013	DHS	2013	DHS	2015	PISA	2013
Egypt, Arab Rep.	2003	2003	DHS	2003	DHS	2003	DHS	2003	TIMSS	2004
Egypt, Arab Rep.	2008	2008	DHS	2008	DHS	2008	DHS	2007	TIMSS	2008
Egypt, Arab Rep.	2014	2014	DHS	2014	DHS	2014	DHS	2015	TIMSS/PIRLS	2015
Ethiopia	2016	2016	DHS	2016	DHS	2016	DHS	2010	EGRA	2016
Gabon	2012	2012	DHS	2012	DHS	2012	DHS	2006	PASEC	2017
Ghana	1998	1998	DHS	1998	DHS	1998	DHS	2003	TIMSS	1998
Ghana	2003	2003	DHS	2003	DHS	2003	DHS	2007	TIMSS	2006
Ghana	2008	2008	DHS	2008	DHS	2008	DHS	2011	TIMSS	2013
Ghana	2014	2014	DHS	2014	DHS	2014	DHS	2013	EGRA	2017
Gambia, The	2013	2013	DHS	2013	DHS	2013	DHS	2011	EGRA	2015
Guatemala	2014	2014	DHS	2014	DHS	2014	DHS	2013	LLECE	2014
Honduras	2005	2005	DHS	2005	DHS	2005	DHS	2008	EGRA	2005
Honduras	2011	2011	DHS	2011	DHS	2011	DHS	2013	LLECE	2011
Haiti	2012	2012	DHS	2012	DHS	2012	DHS	2013	EGRA	2012
India	2015	2015	DHS	2015	DHS	2015	DHS	2009	PISA	2012
Jordan	2002	2002	DHS	2002	DHS	2002	DHS	2003	TIMSS	2003
Jordan	2009	2009	DHS	2009	DHS	2009	DHS	2009	PISA	2008
Jordan	2012	2012	DHS	2012	DHS	2012	DHS	2012	PISA	2010
Kazakhstan	1999	1999	DHS	1999	DHS	1999	DHS	2009	PISA	2001

Continues on next page. See notes on next page.

Appendix Table A2: Country Sample and Data Sources, cont'd

	Year in S-HCI	Child Survival		Stunting		Expected Years		Harmonized Test Scores		PovcalNet
Country	**Dataset**	***Year***	***Source***	***Year***	***Source***	***Year***	***Source***	***Year***	***Source***	***Year***
Kenya	1998	1998	DHS	1998	DHS	1998	DHS	2000	SACMEQ	1997
Kenya	2014	2014	DHS	2014	DHS	2014	DHS	2007	SACMEQ	2016
Kyrgyz Republic	2014	2014	MICS	2014	MICS	2014	MICS	2009	PISA	2014
Lesotho	2004	2004	DHS	2004	DHS	2004	DHS	2000	SACMEQ	2003
Lesotho	2014	2014	DHS	2014	DHS	2014	DHS	2007	SACMEQ	2010
Moldova	2005	2005	DHS	2005	DHS	2005	DHS	2007	PIRLS	2005
Madagascar	2008	2008	DHS	2008	DHS	2008	DHS	2006	PASEC	2010
Mali	2015	2015	MICS	2015	MICS	2015	MICS	2015	EGRA	2010
Myanmar	2015	2015	DHS	2015	DHS	2015	DHS	2014	EGRA	2015
Mozambique	2003	2003	DHS	2003	DHS	2003	DHS	2000	SACMEQ	2003
Mozambique	2011	2011	DHS	2011	DHS	2011	DHS	2007	SACMEQ	2009
Malawi	2000	2000	DHS	2000	DHS	2000	DHS	2000	SACMEQ	1998
Malawi	2006	2006	MICS	2006	MICS	2006	MICS	2007	SACMEQ	2004
Malawi	2010	2010	DHS	2010	DHS	2010	DHS	2010	EGRA	2010
Malawi	2015	2015	DHS	2015	DHS	2015	DHS	2012	EGRA	2016
Namibia	2000	2000	DHS	2000	DHS	2000	DHS	2000	SACMEQ	2004
Namibia	2013	2013	DHS	2013	DHS	2013	DHS	2007	SACMEQ	2015
Niger	2012	2012	DHS	2012	DHS	2012	DHS	2014	PASEC	2011
Peru	2000	2000	DHS	2000	DHS	2000	DHS	2000	PISA	2000
Peru	2003	2003	DHS	2003	DHS	2004	DHS	2006	LLECE	2003
Peru	2012	2012	DHS	2012	DHS	2012	DHS	2015	PISA	2012
Paraguay	2016	2016	MICS	2016	MICS	2016	MICS	2013	LLECE	2016
West Bank and Gaza	2010	2010	MICS	2010	MICS	2010	MICS	2007	TIMSS	2010
West Bank and Gaza	2014	2014	MICS	2014	MICS	2014	MICS	2011	TIMSS	2017
Senegal	2005	2005	DHS	2005	DHS	2005	DHS	2006	PASEC	2006
Senegal	2014	2014	DHS	2014	DHS	2014	DHS	2014	PASEC	2011
El Salvador	2014	2014	MICS	2014	MICS	2014	MICS	2007	TIMSS	2014
eSwatini	2014	2014	MICS	2014	MICS	2014	MICS	2007	SACMEQ	2009
Chad	2004	2004	DHS	2004	DHS	2004	DHS	2006	PASEC	2003
Chad	2014	2014	DHS	2014	DHS	2014	DHS	2014	PASEC	2011
Togo	2013	2013	DHS	2013	DHS	2013	DHS	2014	PASEC	2015
Tajikistan	2012	2012	DHS	2012	DHS	2012	DHS	2016	EGRA	2009
Turkey	1998	1998	DHS	1998	DHS	1998	DHS	2003	PISA/PIRLS	1994
Turkey	2003	2003	DHS	2003	DHS	2003	DHS	2012	PISA	2003
Tanzania	1999	1999	DHS	1999	DHS	1999	DHS	2000	SACMEQ	2000
Tanzania	2009	2009	DHS	2009	DHS	2010	DHS	2007	SACMEQ	2007
Tanzania	2015	2015	DHS	2015	DHS	2015	DHS	2013	EGRA	2012
Uganda	2000	2000	DHS	2000	DHS	2000	DHS	2000	SACMEQ	2000
Uganda	2016	2016	DHS	2016	DHS	2016	DHS	2007	SACMEQ	2017
Vietnam	2013	2013	MICS	2010	MICS	2013	MICS	2015	PISA	2014
Zambia	2001	2001	DHS	2001	DHS	2001	DHS	2000	SACMEQ	2003
Zambia	2007	2007	DHS	2007	DHS	2007	DHS	2007	SACMEQ	2007
Zambia	2013	2013	DHS	2013	DHS	2013	DHS	2011	EGRA	2015
Zimbabwe	2015	2015	DHS	2015	DHS	2015	DHS	2007	SACMEQ	2011

Notes: This table reports the country sample as well as the names and timing of data sources for the four components of the SES-HCI. Each row corresponds to an observation in the dataset. The columns report the data sources and the year in which the data are measured (recognizing that the SES-HCI combines data from different sources in different years). Gray-shaded pairs of observations indicate sample used for over-time comparisons in Section 5.4. DHS: Demographic and Health Surveys; MICS: Multiple Indicator Cluster Surveys; TIMSS: Trends in International Maths and Science Study; PIRLS: Progress in International Reading Literacy Study; PISA: Programme for International Student Assessment; SACMEQ: Southern and Eastern Africa Consortium for Monitoring Educational Quality; PASEC: Program of Analysis of Education Systems; LLECE: Latin American Laboratory for Assessment of the Quality of Education; EGRA: Early Grade Reading Assessments. Last column labelled PovcalNet indicates year of household survey from which quintile share in income/consumption are taken, as retrieved from the World Bank's PovcalNet database.