WPS8593

Policy Research Working Paper 8593

Background Paper to the 2019 World Development Report

Methodology for a World Bank Human Capital Index

Aart Kraay

World Bank Group
Development Economics
Development Research Group
September 2018

Abstract

This paper describes the methodology for a new World Bank Human Capital Index (HCI). The HCI combines indicators of health and education into a measure of the human capital that a child born today can expect to obtain by her 18th birthday, given the risks of poor education and health that prevail in the country where she lives. The HCI is measured in units of productivity relative to a benchmark of complete education and full health, and ranges from 0 to 1. A value of *x* on the HCI indicates that a child born today can expect to be only $x \times 100$ percent as productive as a future worker as she would be if she enjoyed complete education and full health. The methodology of the HCI is anchored in the extensive literature on development accounting.

This paper—prepared as a background paper to the World Bank's *World Development Report 2019: The Changing Nature of Work*—is a product of 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/research. The author may be contacted at akraay@worldbank.org.

Methodology for a World Bank Human Capital Index

Aart Kraay[1]

JEL Codes: I1, I2, O1, O4

Keywords: human capital, health, education, development accounting

---

[1] World Bank Development Research Group, akraay@worldbank.org. This paper was prepared as a background paper for the World Development Report 2019 and for the World Bank's Human Capital Project. It has benefited from extensive discussions with Roberta Gatti, Simeon Djankov and David Weil (Brown). Particular thanks to Rachel Glennerster (DFID), Bill Maloney, Mamta Murthi and Martin Raiser for peer review comments; to Chika Hayashi (UNICEF) and Espen Prydz for guidance on stunting data; to Noam Angrist (Oxford), Harry Patrinos and Syedah Aroob Iqbal for the harmonized test score data; to Deon Filmer and Halsey Rogers for extensive discussions on converting test scores into learning-adjusted school years; to Husein Abdul-Hamid, Anuja Singh (UNESCO) and Said Ould Ahmedou Voffal (UNESCO) for help with enrollment data; to Patrick Eozenou and Adam Wagstaff for help with DHS data; and to Krycia Cowling (IHME), Nicola Dehnen and Ritika D'Souza for tireless research assistance. Valuable comments were provided by Sudhir Anand (Oxford), George Alleyne (PAHO), Ciro Avitabile, Francesco Caselli (LSE), Matthew Collins, Shanta Devarajan, Patrick Eozenou, Tim Evans, Jed Friedman, Emanuela Galasso, Michael Kremer (Harvard), Lant Pritchett (Harvard), Federico Rossi (Johns Hopkins), Michal Rutkowski, Jaime Saavedra, Adam Wagstaff, and Pablo Zoido-Lobatón (IDB). This paper has also benefitted from the discussion at two workshops on measuring the contribution of health to human capital held at the World Bank on March 1, 2018 and May 14, 2018, and a Bank-wide review meeting held June 11, 2018. The data used in this paper have benefitted from an extensive consultation process organized by the office of the World Bank Chief Economist for Human Development, which resulted in many expansions and refinements to the school enrollment and stunting data used in the HCI. The HCI will be published in the 2019 World Development Report and accompanying special report on human capital. *The views expressed here are the author's, and do not reflect those of the World Bank, its Executive Directors, or the countries they represent.*

1. Introduction

Effective investments in human capital are central to development, delivering substantial economic benefits in the long term. However, the benefits of these investments often take time to materialize and are not always very visible to voters. This is one reason why policymakers may not sufficiently prioritize programs to support human capital formation. At the 2017 Annual Meetings, World Bank management called for a Human Capital Project (HCP) to address this incentive problem through a program of advocacy and analytical work intended to raise awareness of the importance of human capital and to increase demand for interventions to build human capital in client countries.

The advocacy component of the HCP features a Human Capital Index (HCI) that measures the human capital that a child born today can expect to attain by age 18, given the risks to poor health and poor education that prevail in the country where she lives. The HCI is designed to highlight how investments that improve health and education outcomes today will affect the productivity of future generations of workers. The HCI measures current education and health outcomes since they can be influenced by current policy interventions to improve the quantity and quality of education, and health.

The main text of this paper provides a nontechnical description of the components of the HCI (Section 2) and how they are combined into an aggregate index (Section 3). This is followed by a description of the index and its interpretation (Section 4). Section 5 discusses how the index can be linked to aggregate per capita income differences and growth, and Section 6 concludes. A lengthy technical appendix provides details on index methodology and data, as well as citations to the relevant literature.

2. Components of the Human Capital Index

Imagine 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 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 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 goal of the HCI is to quantitatively illustrate the key stages in this trajectory and their consequences for the productivity of the next generation of workers, with these three components:

Component 1: Survival. This component of the index reflects the unfortunate reality that not all children born today will survive until the age when the process of human capital accumulation through formal education begins. It is measured using under-5 mortality rates taken from the UN Child Mortality Estimates (**Figure 1**), with survival to age 5 as the complement of the under-5 mortality rate. Data on under-5 mortality are available for 198 countries, and much of the variation across countries in child mortality rates reflects differences in mortality in the first year of life.

Component 2: Expected Learning-Adjusted Years of School. This component of the index combines information on the quantity and quality of education. The quantity of education is measured as the number of years of school a child can expect to obtain by age 18 given the prevailing pattern of enrolment rates. It is calculated as the sum of age-specific enrollment rates between ages 4 and 17. Age-specific enrollment rates are approximated using school enrollment rates at different levels: pre-primary enrollment rates approximate the age-specific enrollment rates for 4 and 5 year-olds; the primary rate approximates for 6-11 year-olds; the lower-secondary rate approximates for 12-14 year-olds; and the upper-secondary rate approximates for 15-17 year-olds. Data to construct this measure is available for 194 countries (**Figure 2**). The quality of education reflects new work at the World Bank to harmonize test scores from major international student achievement testing programs (**Figure 2**). The database covers over 160 countries. These are combined into a measure of expected learning-adjusted years of school, using the conversion metric proposed in the 2018 World Development Report (**Figure 3**).

Component 3: Health There is no single broadly-accepted, directly-measured, and widely-available metric of health that is analogous to years of school as a standard metric of educational attainment. In the absence of such a measure, two proxies for the overall health environment are used to populate this component of the index: (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 (**Figure 4**). Adult survival rates are calculated by the UN Population Division for 197 countries. In the context of the HCI they are used as a proxy for the range of non-fatal health outcomes that a child born today would experience as an adult if current conditions prevail into the future. Stunting serves as an indicator for the pre-natal, infant and early childhood health environment, summarizing 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. Data on the prevalence of stunting is reported in the UNICEF-WHO-World Bank Joint Malnutrition Estimates. This dataset contains 132 countries with at least one estimate of stunting in the past 10 years. The considerations leading to the choice of these two proxy measures for the overall health environment are detailed in Appendix A3.

3. Aggregating the Components into a Human Capital Index

The health and education components of human capital all have intrinsic value that is undeniably important but difficult to quantify. This in turn makes it challenging to combine the different components into a single index. One solution that permits aggregation is to interpret each component in terms of its contribution to worker productivity, relative to a benchmark corresponding to complete education and full health.

In the case of survival, the relative productivity interpretation is very stark, since children who do not survive childhood never become productive adults. As a result, the expected productivity as a future worker of a child born today is reduced by a factor equal to the survival rate, relative to the benchmark where all children survive.

In the case of education, the relative productivity interpretation is anchored in the large empirical literature measuring the returns to education at the individual level. A rough consensus from this literature is that an additional year of school raises earnings by about 8 percent. This evidence can be used to convert differences in learning-adjusted years of school across countries into differences in worker productivity. For example, compared with a benchmark where all children obtain a full 14 years of school by age 18, a child who obtains only 9 years of education can expect to be 40 percent less productive as an adult (a gap of 5 years of education, multiplied by 8 percent per year). Details on the education component of the HCI are provided in Appendix A2.

In the case of health, the relative productivity interpretation is based on the empirical literature on health and income, in two steps. The first step relies on the evidence on health and earnings among adults. Many of these studies have used adult height as a proxy for overall adult health, since adult height reflects the accumulation of shocks to health through childhood and adolescence. These studies focus on the relationship between adult height and earnings across individuals within a country. A baseline estimate from these studies is that the improvements in overall health that are associated with an additional centimeter of height raise earnings by 3.4 percent. However, representative data on adult height are not widely available across countries. Constructing an index with broad cross-country coverage requires a second step in which the relationship between adult height and more widely-available summary health indicators such as stunting rates and adult survival rates is estimated. Putting the estimates from these two steps together results in a "return" to reduced stunting and a "return" to improved adult survival rates. Baseline estimates suggest that an improvement in overall health that is associated with a reduction in stunting rates of 10 percentage points raises worker productivity by 3.5 percent. Similarly, an improvement in overall health that is associated with an increase in adult survival rates of 10 percentage points raises productivity by 6.5 percent. In countries where data on both stunting and adult survival rates are available, the average of the improvements in productivity associated with both health measures is used as the health component of the HCI. When stunting data is not available (most commonly for rich countries), only adult survival rates are used. Details on the health component of the HCI are provided in Appendix A3

Figure 5 and **Figure 6** show the components of the HCI expressed in terms of worker productivity relative to the benchmark of complete education and full health. The vertical axis in each graph runs from zero to one. The distance between a country's value and one shows how much productivity is lost due to the corresponding component of human capital falling short of the benchmark of complete education and full health. The benchmark of "complete education" is defined as 14 learning-adjusted years of school. The benchmark of "full health" is defined as 100 percent adult survival and no stunting. Under the assumptions spelled out in the technical appendix, multiplying together the three components expressed in terms of relative productivity results in a human capital index that measures the overall productivity of a worker relative to this benchmark. The index ranges from zero to one, and a value of x means that a worker of the next generation will be only $x \times 100$ percent as she would be under the benchmark of complete education and full health. Equivalently, the gap between x and one measures the shortfall in worker productivity due to gaps in education and health relative to the benchmark.

4. The Human Capital Index

The overall human capital index is shown in **Figure 7**. The units of the HCI have the same interpretation as the components measured in terms of relative productivity. Consider for example a country such as Morocco, which has a HCI equal to around 0.5. This means that, if current education and health conditions in Morocco persist, a child born today will only be half as productive as she could have been relative to the benchmark of complete education and full health. The HCI exhibits substantial variation across countries, ranging from 0.3 in the poorest countries to 0.9 in the best performers.

All of the components of the HCI are measured with some error, and this uncertainty naturally has implications for the precision of the overall HCI. To capture this imprecision, the HCI estimates for each country are accompanied by upper and lower bounds that reflect the uncertainty in the measurement of the components of the HCI. As described in more detail in Section A4.4, these bounds are constructed by calculating the HCI using lower- and upper-bound estimates of the components of the HCI. The resulting uncertainty intervals are shown in **Figure 8**, as vertical ranges around the value of the HCI for each country. These upper and lower bounds are a tool to highlight to users that the estimated HCI values for all countries are subject to uncertainty, reflecting the corresponding uncertainty in the components. In cases where these intervals overlap for two countries, it indicates that the differences in the HCI estimates for these two countries should not be over-interpreted since they are small relative to the uncertainty around the value of the index itself. This is intended to help to move the discussion away from small differences in country ranks on the HCI, and towards more useful discussion around the level of the HCI itself and what it implies for the future productivity of children born today.

Another feature of the HCI is that it can be disaggregated by gender, for the 126 countries where gender-disaggregated data on the components of the index are available. Gender gaps are most pronounced for survival to age 5, adult survival, and stunting, where girls on average do better than boys in nearly all countries. Expected years of school is higher for girls than for boys in about two-thirds of countries, as are test scores. The gender-disaggregated overall HCI is shown in **Figure 9**. Overall, HCI scores are higher for girls than for boys in the majority of countries. The gap between boys and girls tends to be smaller and even reversed among poorer countries, where gender-disaggregated data also is less widely available.

The HCI uses returns to education and health to convert the education and health indicators into worker productivity differences across countries. The higher are these returns, the larger are the resulting worker productivity differences. The size of the returns also influences the relative contributions of education and health to the overall index. For example, if the returns to education are high while the returns to health are low, then cross-country differences in education will account for a larger portion of cross-country differences in the index. The information in **Figure 5** and **Figure 6** provides a sense of the relative contributions of the different components of the HCI. Learning-adjusted years of school range from around 3 to a potential maximum of 14. This gap implies that children in countries near the bottom of the distribution of expected years of school will only be 40 percent as productive as future workers as children with complete high-quality education. The productivity gaps associated with differences in health outcomes across countries are somewhat smaller. Using adult survival rates as a proxy for overall health, future worker productivity in countries with the worst health outcomes is about 75 percent of what it could be if children enjoyed full health. Using stunting rates, the comparable figure is around 85 percent.

Although different assumptions about the returns to education and health will affect countries' relative positions in the index, in practice these changes are small since the health and education indicators are strongly correlated across countries. This is illustrated in **Figure 10**, which compares the baseline index with three alternatives based on different values for the return to health, using adult survival rates as the health indicator. The top two panels consider weights based on low-end and high-end estimates from the empirical literature on the returns to height, while the bottom panel arbitrarily assumes that cross-country differences in health and education have equally-sized contributions to productivity differences (which implies a return to health almost three times as large as the baseline). In all cases, the correlation of the baseline index with the index based on alternative weights is very high, indicating that the precise choice of weights does not matter greatly for countries' relative positions on the index.

5. Connecting the Human Capital Index to Future Income Levels and Growth

The HCI is measured in terms of the productivity of next generation of workers, relative to the benchmark of complete education and full health. This gives the units of the index a natural interpretation: a value of x for a particular country means that the productivity as a future worker of a child born today is only a fraction x of what it could be under the benchmark of complete education and full health. The relative productivity units of the HCI make it straightforward to connect the index to scenarios for future aggregate per capita income and growth. Imagine a "status quo" scenario in which the expected learning-adjusted school years and health as measured in the HCI today persist into the future. Over time, new entrants to the workforce with "status quo" health and education will replace current members of the workforce, until eventually the entire workforce of the future has the expected learning-adjusted school years and level of health captured in the current human capital index. This can be compared with a scenario in which the entire future workforce benefits from complete high-quality education and enjoys full health. Per capita GDP in this scenario will be higher than in the "status quo" scenario, through two channels: (a) a direct effect of higher worker productivity on GDP per capita, and

(b) an indirect effect reflecting greater investment in physical capital that is induced by having more productive workers.

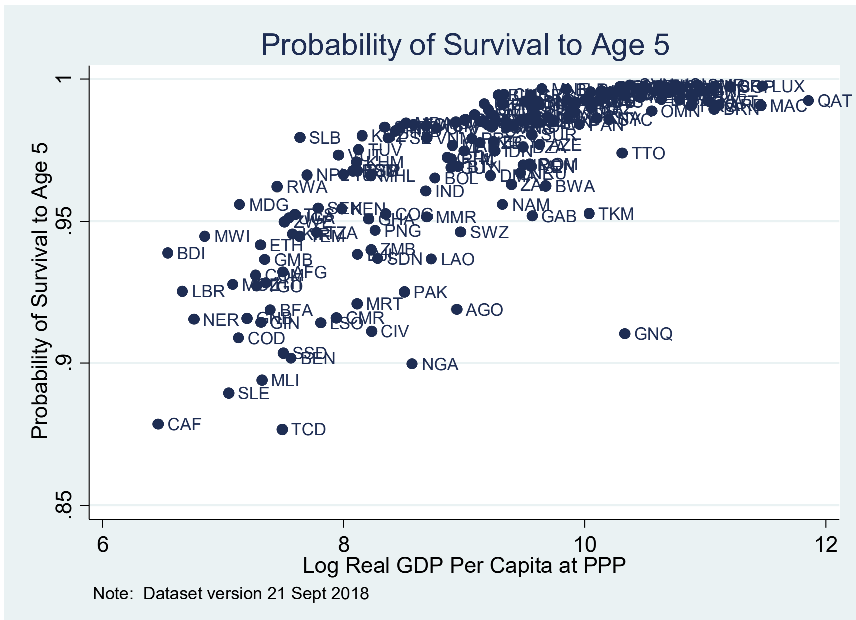
Under standard assumptions from the macro development accounting literature (that are detailed in Appendix A5), projected future per capita GDP will be approximately $1/x$ times higher in the "complete education and full health" scenario than in the "status quo" scenario for a country where the value of the HCI is x. For example, a country such as Morocco with an HCI value of 0.5 could in the long run have future GDP per capita in this scenario of complete education and full health that is approximately $1/0.5$ or two times higher than in the status quo scenario. What this means in terms of average annual growth rates of course depends on how "long" the long run is. For example, under the assumption it takes 50 years for these scenarios to materialize, then a doubling of future per capita income relative to the status quo corresponds to roughly 1.4 percentage points of additional growth per year.

6. Conclusions and Caveats

Like all cross-country benchmarking exercises, the HCI has limitations. Components of the HCI such as stunting and test scores are measured only infrequently in some countries, and not at all in others. Data on test scores come from different international testing programs that need to be converted into common units, and the age of test takers and the subjects covered vary across testing programs. Moreover, test scores may not accurately reflect the quality of the whole education system in a country, to the extent that tests-takers are not representative of the population of all students. Reliable measures of the quality of tertiary education do not yet exist, despite the importance of higher education for human capital in a rapidly-changing world. Data on enrollment rates needed to estimate expected school years often have many gaps and are reported with significant lags. Socio-emotional skills are not explicitly captured. Child and adult survival rates are imprecisely estimated in countries where vital registries are incomplete or non-existent.

One objective of the HCI is to call attention to these data shortcomings, and to galvanize action to remedy them. Improving data will take time. In the interim, and recognizing these limitations, the HCI should be interpreted with caution. The HCI provides rough estimates of how current education and health will shape the productivity of future workers, and not a finely-graduated measurement of small differences between countries. Naturally, since the HCI captures outcomes, it is not a checklist of policy actions, and right type and scale of interventions to build human capital will be different in different countries. Although the HCI combines education and health into a single measure, it is too blunt a tool to inform the cost-effectiveness of policy interventions in these areas – which should instead be assessed based on careful cost-benefit analysis and impact assessments of specific programs. Since the HCI uses common estimates of the economic returns to health and education for all countries, it does not capture cross-country differences in how well countries are able to productively deploy the human capital they have. Finally, the HCI is not a measure of welfare, nor is it a summary of the intrinsic values of health and education – rather it is simply a measure of the contribution of current health and education outcomes to the productivity of future workers.

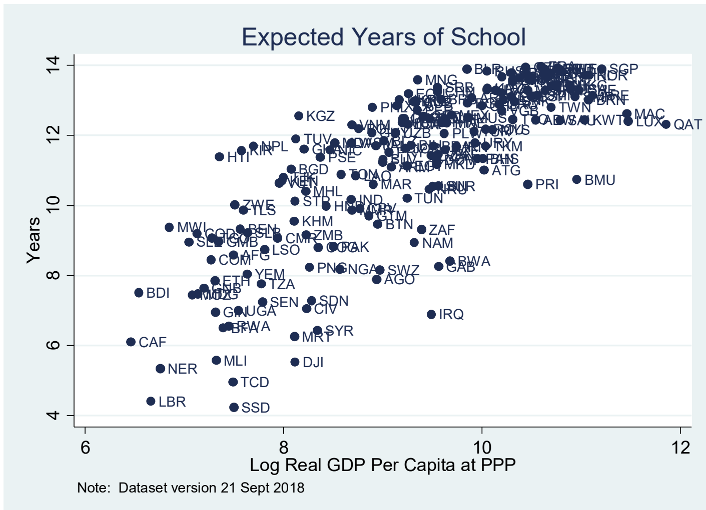
Figure 1: Probability of Survival to Age 5



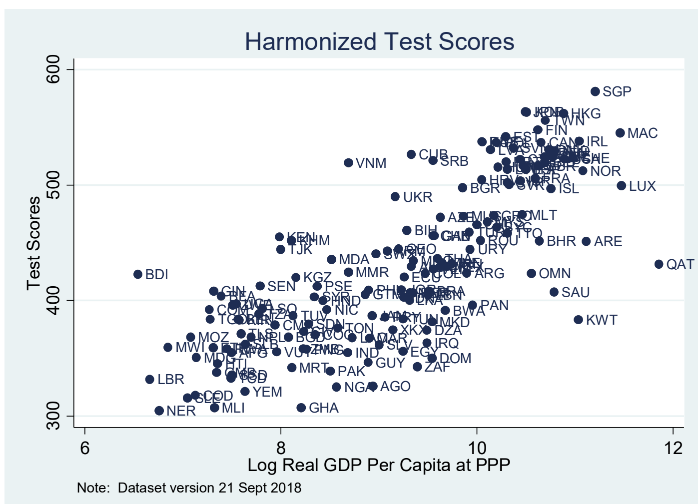
Notes: Probability of survival until age 5 is one minus the under-5 mortality rate. Estimates of under-5 mortality rates are taken from the UN Inter-Agency Group on Child Mortality Estimation (www.childmortality.org), and supplemented with data provided by World Bank staff. Real GDP per capita adjusted for differences in purchasing power parity is taken from the Penn World Tables (Feenstra et. al. (2015)), with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for all countries.

Figure 2: Quantity and Quality of Education Data

Panel A: Expected Years of School By Age 18

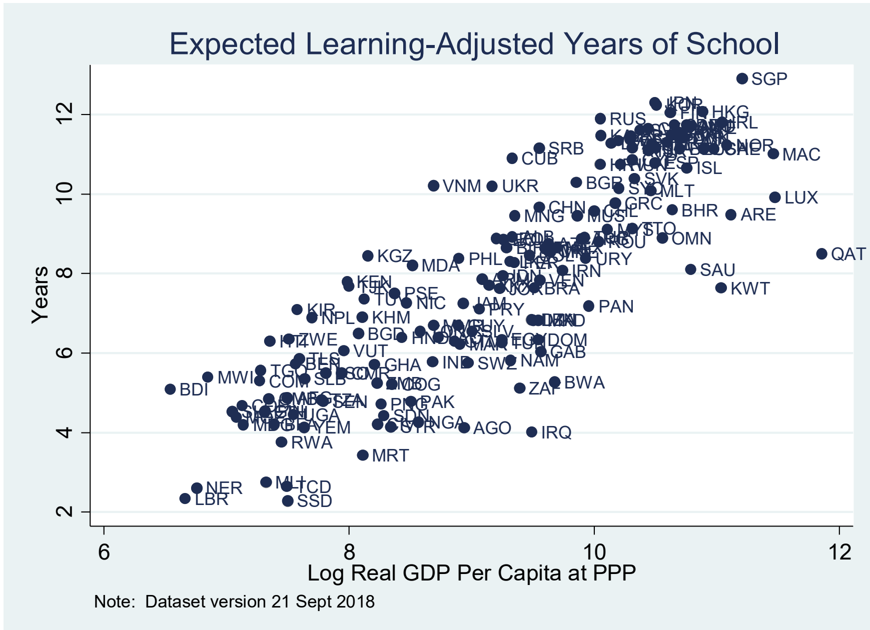


Panel B: Harmonized Test Scores



Notes: Expected years of school are calculated using repetition-adjusted enrollment rates by school level to proxy for age-specific enrollment rates up to age 18. Enrollment rates are taken from the UNESCO Institute for Statistics, and extensively revised/updated/expanded with estimates provided by World Bank staff. Harmonized test scores are taken from Patrinos and Angrist (2018) and are measured in TIMSS-equivalent units, i.e. a mean of 500 and a standard deviation of 100 across students in OECD countries. Real GDP per capita adjusted for differences in purchasing power parity is taken from the Penn World Tables (Feenstra et. al. (2015)) , with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for all countries.

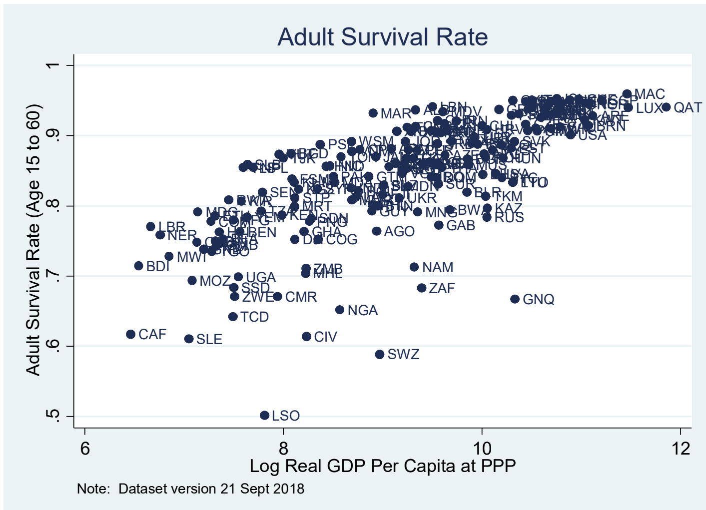
Figure 3: Expected Learning-Adjusted Years of School



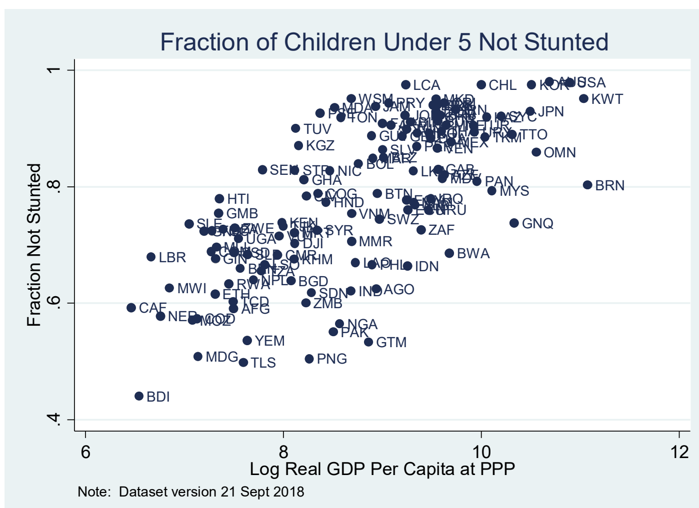
Notes: Learning-adjusted years of school are measured as expected years of school (top panel of Figure 2) multiplied by the ratio of each country's harmonized test score (bottom panel of Figure 2) to a benchmark score of 625, corresponding to the threshold of advanced attainment set by TIMSS. Real GDP per capita adjusted for differences in purchasing power parity is taken from the Penn World Tables (Feenstra et. al. (2015)) , with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for all countries.

Figure 4: Health Indicators

Panel A: Adult Survival Rate

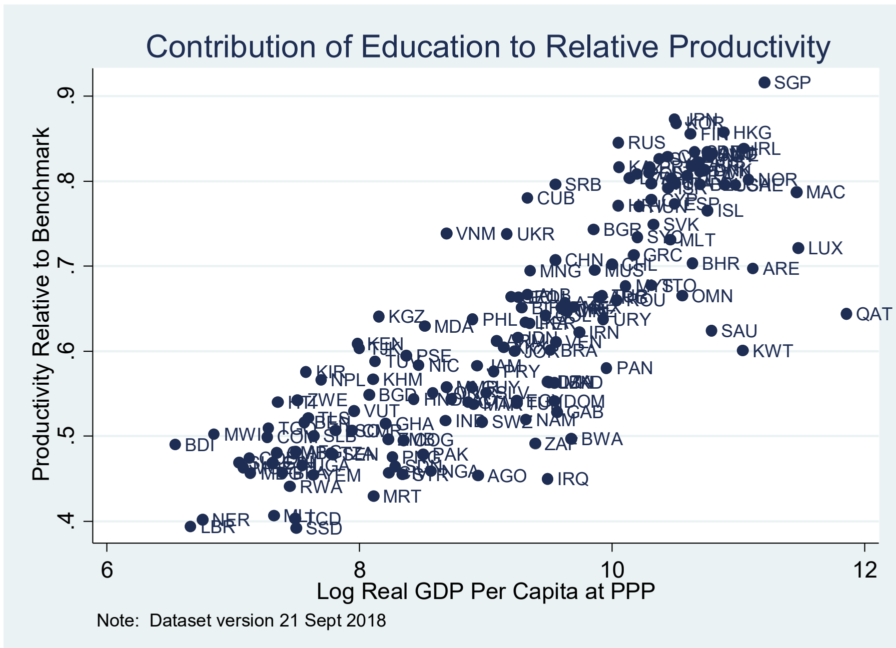


Panel B: Fraction of Children Under 5 Not Stunted



Notes: Adult survival rates are estimated by the UN Population Division and refer to the fraction of 15 year-olds who survive to age 60. Stunting rates are taken from the WHO-UNICEF-World Bank Joint Malnutrition Estimates and refer to the fraction of children under 5 who are more than two reference standard deviations below the reference median height for their age. Data are supplemented with estimates provided by World Bank staff. The graph reports the complementary proportion of children who are not stunted. Real GDP per capita adjusted for differences in purchasing power parity is taken from the Penn World Tables (Feenstra et. al. (2015)) , with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for all countries.

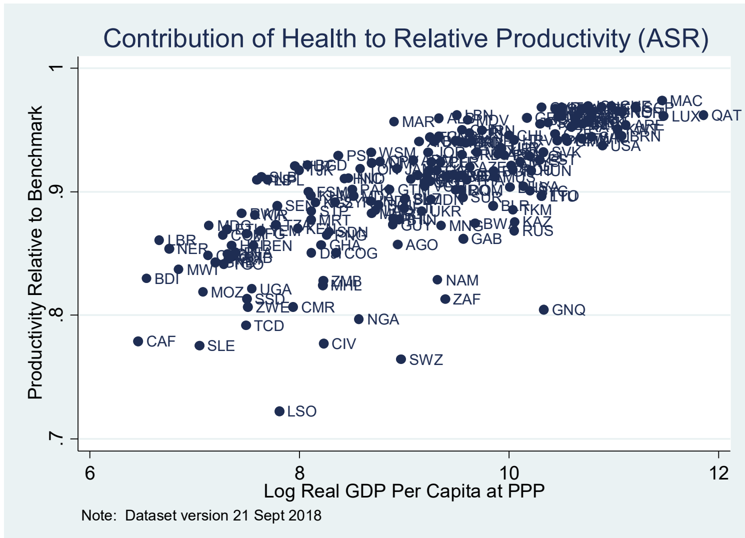
Figure 5: Contribution of Education to Productivity



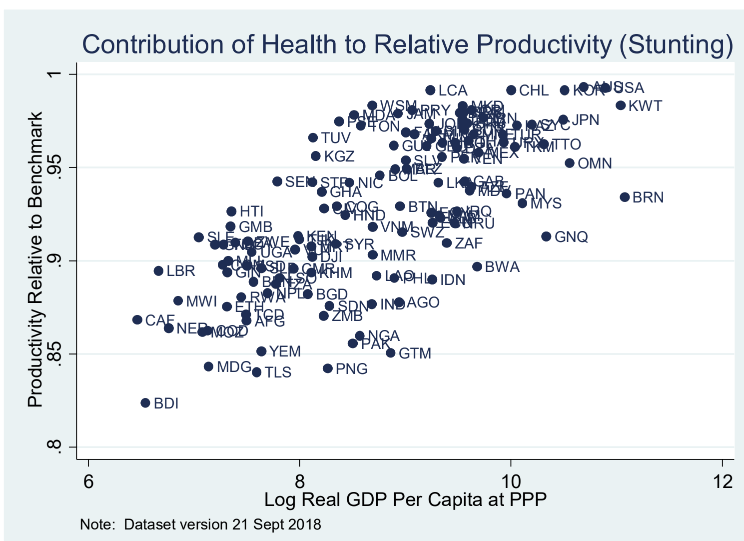
Notes: This graph reports the contribution of cross-country differences in learning-adjusted years of school to cross-country differences in worker productivity. The vertical axis measures the productivity of a worker relative to the benchmark of complete education. Differences in years of school are converted to productivity differences using estimates of the returns to school detailed in Appendix A2. Real GDP per capita adjusted for differences in purchasing power parity is taken from the Penn World Tables (Feenstra et. al. (2015)) , with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for all countries.

Figure 6: Contribution of Health to Productivity

Panel A: Based on Adult Survival Rates

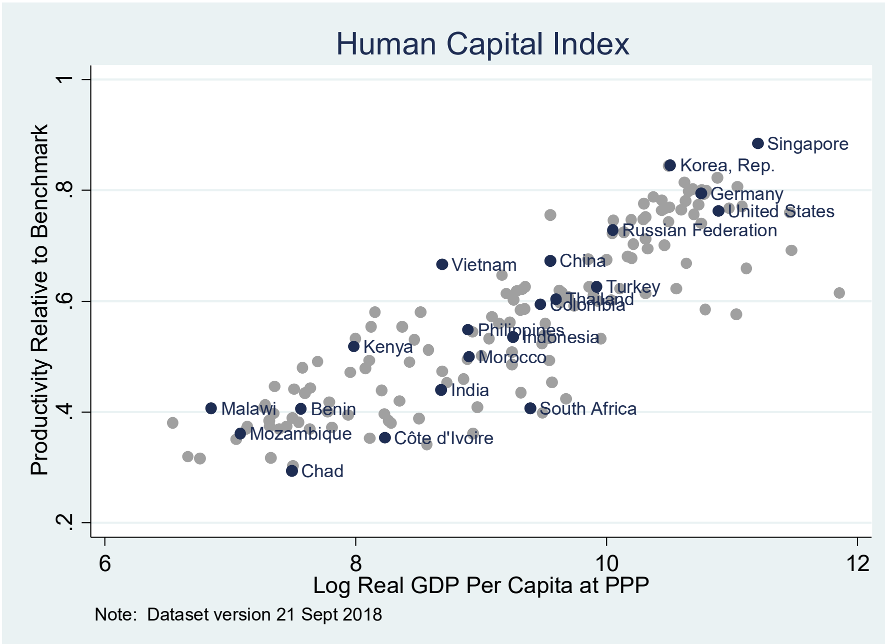


Panel B: Based on Stunting Rates



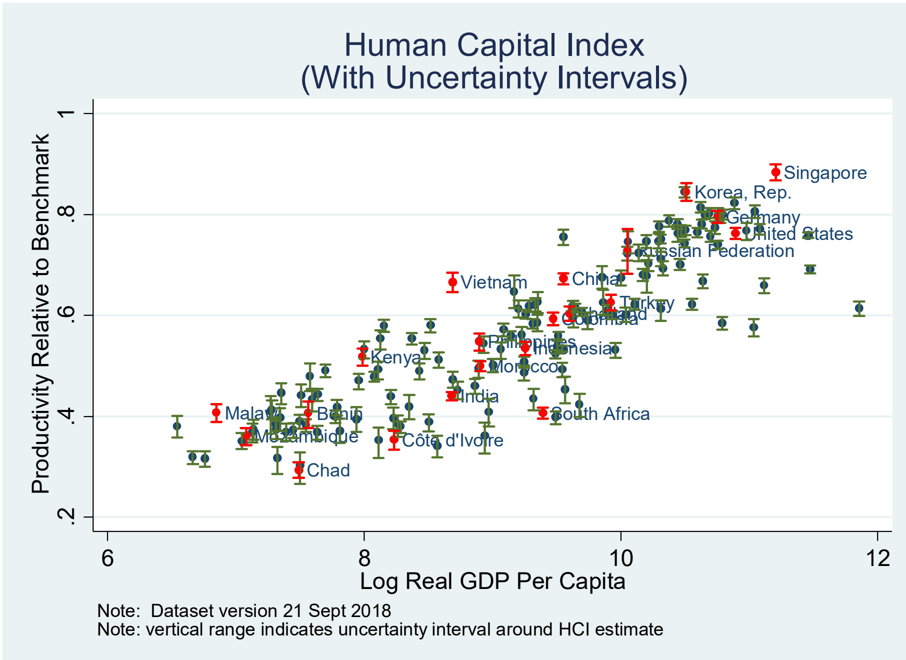
Notes: This graph reports the contribution of cross-country differences in health outcomes to cross-country differences in worker productivity. The vertical axis measures the productivity of a worker relative to the benchmark of full health. Differences in health outcomes are converted to productivity differences using estimates of the returns to health detailed in Appendix A3. Real GDP per capita adjusted for differences in purchasing power parity is taken from the Penn World Tables (Feenstra et. al. (2015)) , with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for all countries.

Figure 7: The Human Capital Index



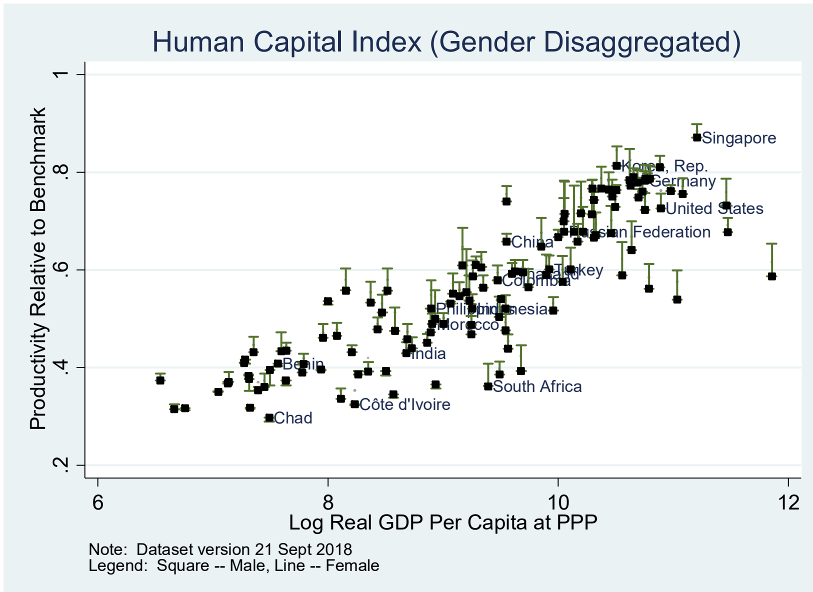
Notes: This figure reports the Human Capital Index. The vertical axis measures productivity relative to the benchmark of complete education and full health. A value of x on the vertical axis means that the productivity as a future worker of a child born today is only $x \times 100$ percent what it would be in the benchmark of complete education and full health. Real GDP per capita adjusted for differences in purchasing power parity is taken from the Penn World Tables (Feenstra et. al. (2015)) , with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for 157 countries. Selected countries are labelled for illustrative purposes.

Figure 8: The Human Capital Index, With Uncertainty Intervals



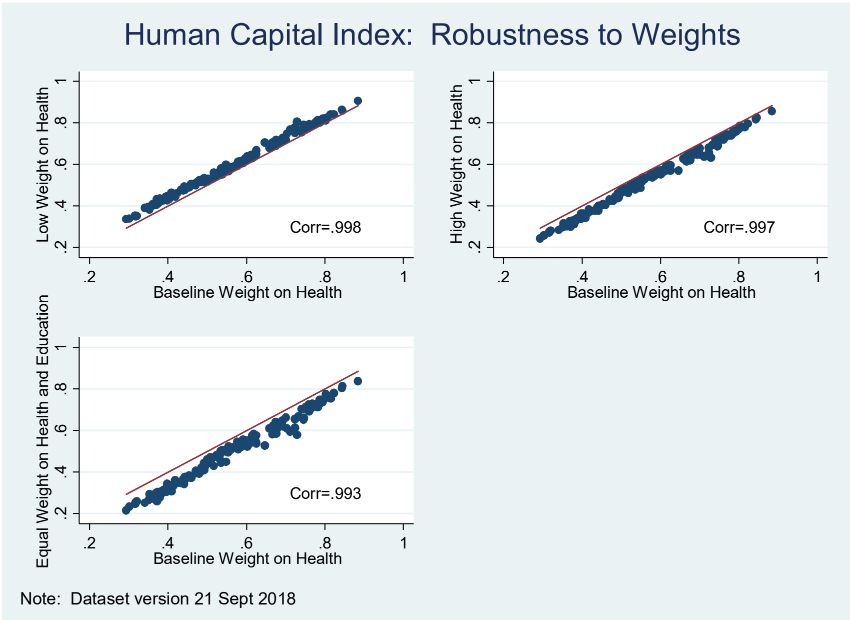
Notes: This figure reports the Human Capital Index. The vertical axis measures productivity relative to the benchmark of complete education and full health. A value of x on the vertical axis means that the productivity as a future worker of a child born today is only $x \times 100$ percent what it would be in the benchmark of complete education and full health. Uncertainty intervals around estimates are shown as vertical ranges for each country. Real GDP per capita adjusted for differences in purchasing power parity is taken from the Penn World Tables (Feenstra et. al. (2015)) , with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for 157 countries. Selected countries are labelled for illustrative purposes.

Figure 9: Gender Differences in the Human Capital Index



Notes: This figure reports the Human Capital Index. The vertical axis measures productivity relative to the benchmark of complete education and full health. A value of x on the vertical axis means that the productivity as a future worker of a child born today is only $x \times 100$ percent what it would be in the benchmark of complete education and full health. Real GDP per capita adjusted for differences in purchasing power parity is taken from the Penn World Tables (Feenstra et. al. (2015)) , with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for 131 countries where gender-disaggregated data for all of the HCI components is available. Selected countries are labelled for illustrative purposes.

Figure 10: Effect of Changing Weight on Health in the Human Capital Index



Notes: This graph shows the effect of changing the weights on the health and education components of the HCI. In each panel the horizontal axis corresponds to the HCI with baseline weights. In the top-left (top-right) panel the vertical axis corresponds to the HCI assuming a low-end (high-end) estimate for the return to health from the empirical literature, as discussed in Appendix A3. The bottom-left panel assumes a much larger value for the return to health that generates the same gap in productivity between best and worst performers as is observed between the best and worst performers in learning-adjusted years of school. Real GDP per capita adjusted for differences in purchasing power parity is taken from the Penn World Tables (Feenstra et. al. (2015)) , with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for all countries.

Technical Appendix: Detailed Methodology for The Human Capital Index

A1. Basic Framework

This section sets out a simple framework used by the development accounting literature to measure human capital and uses it to motivate the Human Capital Index (HCI).[2] This literature begins from the observation that the productivity of an individual worker is higher the more educated she is and the healthier she is. This gain in productivity represents the contributions of health and education to her human capital.

Let s_i represent the years of school of an individual worker i, and let z_i be a measure of her health. The human capital of a worker is:

$$h_i = e^{\phi s_i + \gamma z_i} \tag{1}$$

Section A2 discusses how years of school s_i are measured and adjusted for differences in quality as reflected in performance on international student achievement tests. Section A3 discusses the mapping from unobserved "latent" health z_i to observable health indicators. The parameters ϕ and γ represent the "returns" to an additional unit of education and health. For example, when education is measured as years of school, this formulation implies that an additional year of school raises the human capital of the worker by $100 \times \phi$ percent. As detailed in Sections A2 and A3, plausible values for ϕ and γ can be drawn from the large microeconometric literature that has estimated returns to education and health using individual-level data.

The expected future human capital of a child born today is:

[2] Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999) are early examples of the development accounting approach, and Caselli (2005) and Hsieh and Klenow (2012) provide surveys. See also Caselli (2014) for an application of this methodology to Latin America, commissioned by the LAC region of the World Bank. The discussion of the contribution of health to human capital draws heavily on Weil (2007) and Ashraf, Lester and Weil (2009). Galasso and Wagstaff (2016) use the development accounting approach to assess the macroeconomic costs of stunting.

(2) $$h_{NG} = pe^{\phi s_{NG}+\gamma z_{NG}}$$

where s_{NG} and z_{NG} represent her expected future education and health; p is the probability that a child born today survives; and NG represents the Next Generation of workers.[3] Multiplying by p captures the loss in future productivity per child born today due to premature mortality, since children who do not survive do not grow up to become productive adults. The survival probability p is the complement of the under-5 mortality rate.[4] As discussed below, expected future education and health are measured based on the current outcomes. For example, expected future education will be measured as the number of years of school a child progressing through the education system is likely to obtain given prevailing enrollment rates at different levels. Similarly, expected future health will be measured under the assumption that current health conditions prevail into the future.

Human capital in Equation (1) expresses human capital in units of productivity relative to a worker with $s_i = z_i = 0$, in which case $h_i = 1$. To express the HCI in more intuitive units, rescale Equation (1) by dividing by a benchmark level of human capital corresponding to complete education and full health. Let p^*, s^* and z^* represent these benchmark values. For survival, a natural benchmark is $p^* = 1$. For years of school, the benchmark is $s^* = 14$ years of school, corresponding the maximum possible number of years of school achieved by age 18 by a child who starts school at age 4. For health the natural benchmark corresponding to full health is $z^* = 1$.

With this notation, the HCI is:

[3] Formally, let h_{t+k} represent human capital at some future date $t + k$. Expected future human capital is given by $E_t[h_{t+k}] = pE_t\left[e^{\phi s_{t+k}}\right]E_t\left[e^{\gamma z_{t+k}}\right] \geq pe^{\phi E_t[s_{t+s}]}e^{\gamma E_t[z_{t+s}]}$, where p is the probability a child does not survive to become a future worker, in which case her human capital as a future worker does not materialize. The first equality requires the assumption of independence between education and health outcomes across individuals, and $E_t\left[e^{\phi s_{t+k}}\right]$ and $E_t\left[e^{\gamma z_{t+k}}\right]$ should be interpreted as expectations conditional on survival (and assuming that human capital conditional on not surviving is zero). The second inequality is due to the convexity of the human capital function. Since only the "likely future values" of health and education, $E_t[s_{t+k}]$ and $E_t[z_{t+k}]$, are observable, and not the entire distribution of possible future outcomes which would be required to calculate $E_t\left[e^{\phi s_{t+k}}\right]$ and $E_t\left[e^{\gamma z_{t+k}}\right]$, the last term serves as a lower bound on expected future human capital. Naturally, given the convexity of the human capital function, a higher variance of education and health across individuals, and a higher covariance between the two, increases the gap between the lower bound and the expectation. To keep notation simple, s_{NG} and z_{NG} denote the likely future values $E_t[s_{t+k}]$ and $E_t[z_{t+k}]$ that represent the expected education and health of the next generation of workers.

[4] Data on under-5 mortality are produced by the UN Child Mortality Estimates. Most of the cross-country variation in mortality under 5 is due to cross-country variation in under-1 mortality rates.

(3) $$HCI = \frac{p}{p^*} \times e^{\phi(s_{NG}-s^*)} \times e^{\gamma(z_{NG}-z^*)}$$

The HCI is the product of three easily-interpretable components, each measuring productivity relative to the benchmark of full health and complete education. The first term, $\frac{p}{p^*}$, captures forgone future productivity due to child mortality, since children who do not survive never become productive adults. As a result, the average productivity as a future worker of a child born today is reduced by a factor equal to the survival rate, relative to the benchmark where all children survive. The second term, $e^{\phi(s_{NG}-s^*)}$, reflects foregone future productivity due to children completing less than a full 14 years of school. The third term, $e^{\gamma(z_{NG}-z^*)}$, reflects the reduction in future worker productivity due to poor health. Multiplying these three terms together gives the overall productivity of a worker relative to the benchmark of complete education and full health.

This approach is closely linked to standard measures of the average human capital per worker of the current workforce that have been widely used in the development accounting literature:

(4) $$h_{CW} = e^{\phi s_{CW}+\gamma z_{CW}}$$

where h_{CW} represents the average human capital of the current workforce, and s_{CW} and z_{CW} represent the average levels of education and health in the current workforce. The only difference between this measure and the expected human capital of the next generation in Equation (2) is that the term reflecting the probability of survival is not required. This is because the measure of human capital of the current workforce measures the average human capital of workers who are currently living.

While measures of the human capital stock like those in Equation (4) are standard in the development accounting literature (see for example Weil (2008)), they are less well suited to the communications and advocacy purpose of the HCI. This is because measures of the human capital of the existing workforce – and most particularly the education component, reflect the educational opportunities that were available to current workers in the past when they were school-aged children, and so now are largely beyond the influence of current and future policy interventions. Instead, the HCI measures how current health and education outcomes – that are amenable to improvement through current and future policy efforts – shape the likely future human capital of children born today.

Measures of the monetary value of human capital based on the present value of future earnings of individuals, analogous to estimates of the value of physical capital as the present value of future returns, also exist. Naturally, these measures are conceptually closely related. Suppose for example that log wages of individual i at some future time t are given by a health-augmented Mincer equation like $lnw_{it} = \phi s_i + \gamma z_i + g_i t$, where g_i represents future trend growth in wages for the individual. Treating the unskilled wage as the numeraire, human capital measured as the present value of future wages is simply $\frac{h_i}{\delta - g_i}$, where δ represents the discount rate, and h_i is the measure of individual human capital in Equation (1). Human capital measures along these lines have a long history (see for example Jorgenson and Fraumeni (1998)), and are extensively discussed in the context of satellite national accounts in UN (2016). Measures of human capital along these lines in a cross-country setting have been developed since 2012 in the United Nations University "Inclusive Wealth Index" study (UNU (2012)), as well as in the latest edition of the World Bank's "Changing Wealth of Nations" report (World Bank (2018)). The key incremental difficulty in constructing these measures relative to measures of h_i is coming up with plausible measures of future earnings growth, g_i. Because the difference between the growth rate and the discount rate is small and enters in the denominator of this measure, small changes in assumed growth rates are magnified into large changes in measured human capital.[5]

[5] For example, if the discount rate is five percent, changing the assumed growth rate from three to four percent per year has the effect of doubling the measured human capital stock.

A2: Education

A2.1 Data on Expected Years of School

Expected number of years of school (EYS) that a child who starts school at age 4 would attain by her 18th birthday is calculated using the methodology described in UNESCO et. al (2014), Section 2.2 and Annex 2.2. Conceptually, expected years of school achieved by age A is simply the sum of age-specific enrollment rates over all ages in this age range, i.e.

$$EYS = \sum_{a=4}^{17} ENR_a \tag{5}$$

where ENR_a is the enrollment rate of children aged a. Unfortunately however, age-specific enrollment rates are not systematically available for a broad cross section of countries. Instead, more readily-available data on enrollment rates by level of school are used to approximate enrollment rates in different age brackets. Specifically, pre-primary enrollment rates approximate the age-specific enrollment rates for 4 and 5 year-olds; the primary rate approximates for 6-11 year-olds; the lower-secondary rate approximates for 12-14 year-olds; and the upper-secondary rate approximates for 15-17 year-olds. Naturally, cross-country definitions in school starting ages and duration of different levels of school imply that these will only be approximations to the number of years of school a child can expect to complete by age 18.

The ideal measure of enrollment rates for this calculation is the "total net enrollment rate" (TNER), which measures the fraction of children in the theoretical age range for a given level of school, who are in school at any level. For example, if the theoretical age range for lower secondary school is 12 to 14 years, then the TNER measures all children aged 12 to 14 who are enrolled in any level of school as a fraction of all children aged 12 to 14. In this way, the TNER best approximates the age-specific enrollment rates for ages 12 through 14 since it captures the enrollment status of all 12 to 14 year-olds, irrespective of what level of school they are in. Unfortunately however, data on TNER are missing for many countries and years in the UNESCO database, and, depending on the country and year, one or more of three other enrollment rates are more widely available. These are (i) the "adjusted net enrollment rate" (ANER), measuring the fraction of children in the theoretical age range for a given level of school who are in school at that level or the level above; (ii) the "net enrollment rate" (NER), measuring the fraction of children in the theoretical age range for a given level of school who are in school at that level; and (iii) the "gross enrollment rate" (GER), measuring the number of children of any age who are enrolled in a given level, as a fraction of the number of children in that age range.

To maximize country coverage, the following procedure populates the enrollment rate series used to calculate expected years of school:

- Available data for all four enrollment rates for all four levels of school were retrieved, combining information from the three external data platforms maintained by the UNESCO Institute of Statistics. For pre-primary, TNERs are not reported as there is no level below pre-primary, and ANER is available only for the age corresponding to one year before the official start of primary school.
- Gaps in each enrollment rate for each level and country were filled by taking the most recently available data, going backwards up to 10 years for each country-year observation.
- Within each level, the preferred enrollment rate available in the filled-in data as of 2017 was chosen, in the following order of priority: TNER, ANER, NER, and GER. The filled-in series for this enrollment rate was then used for all years for this level of school. Note that in some cases differences in availability of enrollment measures means that different enrollment rates are used for different levels of school. However, with a level for a given country, same type of enrollment rate is used.
- Data on repetition rates for primary, lower-secondary, and upper-secondary school were retrieved, and filled in with up to 10 years of lags in the same way as for enrollment rates. These are needed to adjust enrollment rates obtained in the previous step for repetition. Adjusting for repetition is important because failing to do so would count students repeating a grade as gaining an additional year of school, and in some school systems at some levels repetition rates are as high as 25 percent. Data on repeaters is available for most countries included in the HCI; However for a few countries where data on repetition is not reported by UNESCO, out of necessity repetition rates are assumed to be zero.
- Finally, in a few cases where only GERs were available and repetition-adjusted GERs exceeded 100 percent. These are topcoded at 100 percent.

The resulting measure of expected school years approximates the number of years of school that a child can expect to attain by her 18th birthday if she starts school at age 4, for a maximum of 14 years. Data on this measure of expected years of school are reported in the top panel of **Figure 2**.

Conceptually this calculation corresponds to the measure of "school life expectancy" (SLE) calculated by UNESCO Institute for Statistics.[6] However, the implementation here differs in that UIS uses gross enrollment rates to calculate school life expectancy, whereas here total net enrollment rates are used wherever possible. This reflects a tradeoff. On the one hand, gross enrollment rates are more widely available and typically have longer time-series coverage in the UIS data. On the other hand, total net enrollment rates conceptually correspond more closely to the age-specific enrollment rates in Equation (5). A further reason to use total net enrollment rates is that for some countries the repetition-adjusted gross enrollment rates reported by UNESCO are – sometimes implausibly – well above 100 percent. While gross enrollment rates (the number of students enrolled at a given grade level as a fraction of children of the theoretical age for that grade level) can in principle exceed 100 percent if some children start school early or late, it is difficult for this timing effect to generate gross enrollment rates that are persistently much above 100 percent.[7] A practical consequence of using gross enrollment rates above 100 percent to calculate SLE is that SLE can then exceed the statutory duration of school. This is the case for primary and secondary SLE as reported by UIS for about one-quarter of

---

[6] When more granular data such as enrollment rates by age are available, UNESCO uses them to calculate SLE. When they are not, UNESCO calculates SLE using school level-specific gross enrollment rates (which are the most widely available in the UNESCO data), using Equation (5). Whether age-specific or level-specific enrollment rates are used makes little practical difference for the calculations. Using the data on gross enrollment rates and duration by level of school as reported by UNESCO in Equation (5), and restricting attention to primary and secondary, reproduces the UNESCO estimates of SLE for primary plus secondary almost exactly.

[7] To see why, consider an education system in which a fraction a_t of each cohort of students start school at the official age; a fraction l_t starts one year late and therefore are above age-for-grade; and a fraction $1 - a_t - l_t$ students do not start school at all. Gross enrollment in first grade at time t is the number of students in first grade divided by the cohort of children aged 6 who are supposed to be in first grade, i.e. $GER_t = \frac{a_t \times N_t + l_{t-1} \times N_{t-1}}{N_t}$ where N_t is the number of children in the cohort of new six year-olds at time t. With a stationary population $N_t = N_{t-1}$ and stationary starting age shares $a_t = a$ and $l_t = l$, it follows that $GER = a + l < 1$, even though there could be many over-aged children in the grade. With stationary starting age shares the only way to generate $GER = 1$ is if $a + l$ is close to one (i.e. near-universal enrollment) and $N_{t-1} > N_t$ i.e. a shrinking school-aged population. The other way to generate $GER > 1$ is with time-varying starting age shares. If for example there is a big expansion in enrollment to include children who should have started in the previous year but did not, then with stable cohort size it follows that $GER = a_t + l_{t-1}$ and it is possible that $a_t + l_{t-1} > 1$ if enough children from the previous cohort who did not enroll at the correct age now enroll, i.e. if l_{t-1} is large enough. One factor contributing to high gross enrollment rates in some countries is that they can include adults who are returning to complete secondary school, which can inflate gross enrollment rates beyond 100 percent, making it a poor proxy for enrollment rates among 14-17 year-olds

countries in 2015. In contrast, here enrollment rates below 100 percent are used at all levels, with the result that the maximum possible expected years of school is fixed at 14 years.

A2.2 Harmonizing Test Scores

The school quality adjustment is based on a new large-scale effort to harmonize international student achievement tests from several multi-country testing programs. A detailed description of the test score harmonization exercise is provided in Patrinos and Angrist (2018). This paper updates and expands the dataset described in Altinok, Angrist, and Patrinos (2018). This earlier dataset harmonized scores from three major international testing programs (Trends in International Maths and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), and Programme for International Student Assessment (PISA)), as well as three major regional testing programs (Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), Program of Analysis of Education Systems (PASEC), and Latin American Laboratory for Assessment of the Quality of Education (LLECE)). Patrinos and Angrist (2018) subsequently update this dataset with more recent rounds of PIRLS, PASEC and SACMEQ, and also substantially expand the cross-national coverage of the database by including Early Grade Reading Assessments (EGRA). The expanded dataset covers over 160 countries.

Test scores from these different testing programs can be converted into common units (or "harmonized") using a variety of methodologies, as described in detail in Patrinos and Angrist (2018) and Altinok, Angrist and Patrinos (2018). In the version of the data used here, the numeraire units are those of the international standardized achievement tests. This includes TIMSS (for math and science) and PISA and PIRLS (for reading, at secondary and primary level, respectively). These tests are expressed in units with a mean of 500 and a standard deviation across students of 100 points. The harmonization method is based on the ratio of country-level average scores on each program to the corresponding country-level scores in the numeraire testing program, for the set of countries participating in both the numeraire and the other testing program. For example, consider the set of countries that participate in both PISA and TIMSS assessments. The ratio of average PISA scores to average TIMSS scores for this set of countries provides a conversion factor for PISA into TIMSS scores, that can then be used to convert all countries' PISA scores into TIMSS scores. Altinok, Angrist and Patrinos (2018) and Patrinos and Angrist (2018) refer to the set of common countries as the "doubloon countries", the resulting conversion factor as the "doubloon index", and the test scores in common units as "harmonized learning outcomes (HLOs)". In the version of the data used here, the doubloon index is calculated pooling all doubloon observations between 2000 and 2017 and therefore is constant over time. This ensures that within-country over-time fluctuations in harmonized test scores for a given testing program reflect only changes in the test scores themselves, and not changes in the conversion factor between tests.[8] Harmonization is done at the subject x grade level. The country-level test scores used in the HCI average across subjects and grades.

In most cases, the tests are designed to be nationally representative. There are however some notable cases where they are not, including:

- In a number of countries, EGRA assessments are not nationally-representative and are identified as EGRANR in the data documentation.
- In India, the 2009 PISA was administered in two states (Himachal Pradesh and Tamil Nadu). However, a comparison with state-level scores for all of India in the 2012/2013 National Achievement Survey (NAS) suggests that the average NAS score for these two states is quite similar to the national average NAS score, indicating that the 2009 PISA scores probably are roughly representative of India as a whole.[9]
- PISA scores for China in 2009 and 2012 are based only on reported data for Shanghai, and in 2015 for Beijing, Shanghai, Jiangsu and Guangdong (B-S-J-G). Shanghai and B-S-J-G are both considerably richer than China as a whole. Since test scores tend to improve with income within and across countries, reported PISA scores are unlikely to be nationally representative. Corroborating evidence can be found in Gao et. al. (2017) who implement PIRLS assessments in Shaanxi and rural Jiangxi and Guizhou provinces, the latter being among the poorest areas in China. As noted in Gao et. al. (2017), test scores in these areas are among the lowest PIRLS scores observed globally. Extrapolations using average household per capita income using (a) Shanghai and B-S-J-G PISA scores, and (b) PIRLS scores in Gao et. al. (2017), result in similar estimates of nationally-representative test scores. Converting these extrapolations to HLO units and averaging them together results in an HLO of 456 (as opposed to 532 in B-S-J-G alone). This extrapolated estimate is used as the HLO for China.[10]

[8] The one exception to this is the 2007 and 2014 PASEC rounds, which were not designed to be intertemporally comparable, and for which different doubloon countries are available in 2007 and 2014.
[9] See the appendix of Patrinos and Angrist (2018) for details.
[10] See the appendix of Patrinos and Angrist (2018) for details.

The next step is to create a single time series of harmonized learning outcomes (HLOs) for each country, combining HLOs from different testing programs in a way that balances coverage and comparability over time. This is done in three main steps:

- International Testing Programs: Data from TIMSS and PIRLS are combined into a single testing program, recognizing that both tests are carried out by the same parent organization, the International Association for the Evaluation of Educational Achievement (IEA), and use common units. TIMSS assessments have been carried out in 2003, 2007, 2011 and 2015, and PIRLS in 2001, 2006, 2011, and 2016. PIRLS assessments are shifted to the nearest TIMSS year, and then a combined TIMSS/PIRLS score is created by averaging the two (for countries participating in both), and otherwise using whichever of the two is available. TIMSS/PIRLS and PISA scores are chosen by taking whichever of the two is available for each country-year, and the average of the two in 2003 and 2015 (when the two testing programs were both carried out in the same year, for the set of common countries). The end result is that the HLO data as used in the HCI has 175 country-year observations between 2000 and 2017 based on TIMSS/PIRLS, 278 country-year observations based on PISA, and 68 country-year observations based on an average of PISA and TIMSS/PIRLS.
- Regional Testing Programs (LLECE, SACMEQ, PASEC). Scores from these programs are used in the combined HLO measure only for those country-year observations where no data from international testing programs is available. This adds 25 country-year observations from LLECE, 37 from SACMEQ, and 30 from PASEC. Six country-year observations from LLECE and one observation from PASEC are dropped because the occur in country-years for which an international test is available.
- Early Grade Reading Assessment (EGRA). The HLO database contains data from 75 EGRA assessments carried out in 48 low- and middle-income countries. There are 27 countries where EGRA is the only available international test, and so for these countries the HLO simply is based on the EGRA assessment. The remaining 22 countries also have data from one of the major international and/or regional testing programs. For these countries, EGRA data are included in the time series of test scores for the country only if (a) the additional test substantially expanded the time series coverage of the combined HLO, and (b) the additional test did not

create an implausibly large change in the combined HLO score. The end result of this step is to add 54 country-year observations based on EGRA assessments.[11]

Finally, the HLO dataset includes one observation for Sri Lanka in 2013 based on the national assessment, which was linked to TIMSS. These steps result in a combined HLO score with 668 country-year observations covering 162 countries. The median country has three time series observations between 2000 and 2017. The HCI is based on a cross-section of test scores in 2017, taking the most recent test score available in the period 2006-2017. For 133 of these countries, the most recently-available test score falls in the five-year period 2013-2017. A cross-section of most recently-available test scores are displayed in the bottom panel of **Figure 2**. Test scores range from just below 600 in the best-performing countries to just above 300 in the worst-performing countries.

A2.3 Adjusting Expected Years of School for Quality

The next step is to transform test scores into a quality-adjustment factor for years of school. There are (at least) five options for doing so:[12]

- The first is to exploit the fact that in some settings the same test is administered to children in different grades. If children in different grades are on average similar in their abilities, the difference in average performance between children in the two grades can be interpreted a measure of the "productivity" of the additional year in terms of higher test scores. This gradient can then be compared across countries to infer cross-country differences in average school quality. One application of this is done directly by PISA, and exploits the fact that PISA tests are administered to children at age 15, irrespective of what grade they are in. While most PISA test takers are in 9^{th} or 10^{th} grade (depending on the country), some are in earlier or later grades. Comparing the difference in PISA scores across grades within a country provides an estimate of how many PISA points are gained as a result of one more year of school in that country.

[11] There is one exception to the rule of giving preference to international assessments over EGRAs. This is for Yemen, which participated in TIMSS in 2007 and 2011. Test scores in both years were extremely low (low 200s), and TIMSS caveated the results, noting that too few students achieved minimum performance standards for the country-level average scores to be reliably estimated. These TIMSS observations are droped and the 2011 EGRA for Yemen is used instead.

[12] For an early effort to incorporate cross-country differences in test scores into measures of human capital, see Caselli (2005), Section 4.2. Studies such as Hanushek and Woessman (2010) have incorporated tests scores into cross-country growth regressions. Hanushek, Ruhose, and Woessman (2015) use test scores as a measure of cognitive ability in a development accounting exercise across US states, and Caselli (2014) performs a similar exercise across countries in Latin America.

Estimates from the latest PISA round suggest that on average roughly 30 PISA points are in this sense equivalent to an additional year of school.[13] This would imply that the observed 300 point range in test scores in the data correspond to $\frac{300}{30} = 10$ years of school. A drawback of this approach is that it does not control for selection. For example, if high-ability students are more likely to be in 10th grade than 9th grade at age 15, then this calculation overstates the return in terms of PISA points to an additional year of school. A conceptually-related approach is taken by Kaarsens (2014), who exploits the fact that the 1995 TIMSS round was administered to some 3rd and 7th grade students as well as the usual 4th and 8th grade students for this program. He finds that a year of school in the US is worth approximately three times as much as in low-income countries.

- Evans and Yuan (2017) suggest a related approach which instead looks at slope of the relationship between test score obtained by adults and the number of years of education they obtained as children. The main advantage of this approach is looks at the effect of education on learning over a wider range than just the one-year comparisons in the PISA approach, and also controls for selection issues. They find that a one standard deviation improvement in Skills Towards Employability and Productivity (STEP) scores – a testing program for adults in developing countries – is equivalent to between 5 and 7 years of school. The difficulty in applying these results in this context is that it is unclear how to map standard deviations of STEP scores into standard deviations of PISA scores. Under the somewhat arbitrary assumption that the standard deviations are the same, their results would imply that the 300 PISA point differences (or 3 standard deviations) observed across countries translate into a very large gap of 15 to 21 years of school.[14]
- A third approach is suggested in World Bank (2017) and developed further in Filmer, Rogers, Angrist, and Sabarwal (2018). They propose scaling actual average school years in a country by the ratio of the country's test score to a benchmark value of top performance, and explain how that this scaling factor can be justified under the assumption that grade-learning trajectories are linear through the origin. A natural benchmark value for top performance is a TIMSS score of 625, which corresponds to the TIMSS "advanced" international benchmark. Since country-level average test scores range from around 300 in the worst-performing countries to 600 in the best-

[13] See OECD (2016), p. 64.
[14] A similar exercise is done in Hanushek and Zhang (2009), who use adult test scores for 14 mostly OECD countries and relate them to years of educational attainment of the test takers.

performing countries, the ratio of test scores relative to the leader is 300/625=0.48 in the worst-performing countries. This implies that a year of school in the worst-performing countries is "worth" only about half as much as a year of school in the best-performing countries.

- A fourth approach is suggested in Schoellman (2012), who relates estimated returns to education of immigrants in the United States to test scores in their countries of origin, and finds a strong positive relationship. He then introduces this relationship into a model in which agents take education quality into account when choosing their investment in years of school. His estimates suggest that the return to a year of school needs to be doubled relative to standard benchmarks to take into account the fact that in countries where school attainment is low, this reveals an endogenous response to low quality. While this argument does not provide a direct mapping from individual country test scores, it suggests a factor-of-two quality adjustment which is similar to World Bank (2017) and Filmer, Rogers, Angrist, and Sabarwal (2018).
- A fifth approach can be anchored in the small literature that has jointly estimated returns to years of school and returns to cognitive ability, where cognitive ability is proxied using test scores. For example, Hanushek and Yang (2009) estimate Mincerian regressions of log earnings on education and test scores and find a return per year of school of 10 percent, and a return per standard deviation of test scores of 10 to 20 percent (as cited in Hanushek, Ruhose and Woessmann (2015)) . The ratio of these two estimated returns implies that one standard deviation of test scores is equivalent to 1 to 2 years of school in terms of its effects on earnings. Under the somewhat arbitrary assumption that a standard deviation of test scores in the Hanushek and Yang (2009) study is equivalent to a PISA standard deviation of 100, this would imply that the 300 PISA point gap in test scores between the worst and best performing countries is equivalent to approximately 3 to 6 years of school. At the other extreme, Caselli (2014) cites estimates in Vogl (2014) that suggest a return to school of 7.2 percent and a return per 100 PISA points of 1.4 percent, suggesting that the 300 PISA point gap in test scores across countries is equivalent to only about $\frac{1.4}{7.2} \times 3 = 0.6$ years of school.

Absent a clear empirical consensus on the equivalence between test scores and years of school, the HCI calculations in this document use the mapping from test scores to quality differences suggested by Filmer, Rogers, Angrist, and Sabarwal (2018) because of its simplicity and ease of communication. The calculations are implemented using the most recently available test scores for each country. This amounts to the assumption that the quality of education that children will receive in the future is the same as what is reflected in the most recently-observed test scores.

It is also worth noting that for most countries, test scores are observed only at one grade level. This means that out of necessity quality measured at one grade level is assumed to be representative of the entire school system. This assumption however may however be questionable in countries where there is a lot of attrition as students move to higher grades. In this case, test scores observed in later grades reflect an element of selection, to the extent that lower-ability students are more likely to drop out of school before reaching the level at which the test is implemented. This implies that observed average test scores in later grades likely overstate the quality of education in lower grades. Adjusting for this is difficult, as it requires some way to approximate the distribution of test scores among those students who did not take the test. However, a partial adjustment for this in the overall HCI comes through the fact that in such cases, the enrollment rates used to calculate expected years of school will also be lower, which in turn lowers the learning-adjusted years of school measure.

A final caveat worth noting is that test scores are only an imperfect measure of learning. There are the usual concerns that test scores measure performance only on those items of the curriculum that are covered in the test. Beyond this, test scores also respond to other factors, including intrinsic motivation on the part of test takers. For example, extensive empirical evidence from schoolchildren in Chicago suggests that small immediate financial rewards for good performance have substantial effects on students' standardized test scores (Levitt, List, Neckerman and Sadoff (2016)). To the extent that there are cross-country differences in students' intrinsic motivation when taking standardized tests, this will be conflated with the quality interpretation of the test scores. Student test scores also reflect the influence of the home environment. De Philippis and Rossi (2017) use US PISA data to study test scores of children attending the same school, but whose parents immigrated from different countries. Children of parents who immigrated from countries with high test scores also tend to do better on test scores themselves, holding constant the quality of education they received by focusing on children in the same schools.

A2.4 Returns to Education

Values for the return to education parameter ϕ can be anchored in the vast empirical literature that estimates Mincer (1958)-style regression of log wages on years of school. Returns to education naturally vary across levels of education, by gender, and across countries. However, in the interests of generating a simple and transparent index that focuses on the variation in the quantity and quality of education across countries, the HCI uses a single benchmark value of $\phi = 0.08$, or 8 percent per year of school, for all countries and all levels of school. To put this value in perspective, consider the following sets of estimates from the existing literature:

- Montenegro and Patrinos (2014) estimate returns to education using household survey data from 139 countries. They find an overall average return to an additional year of school of 10.1 percent, and disaggregated returns of 10.6 (7.2) (15.2) for primary (secondary) (tertiary). Averaging the primary and secondary returns which are most relevant for the HCI gives a value of 8.9 percent. Jedwab and Islam (2018) update these estimates in background analysis for the 2019 World Development Report, and find an average return to school of around 8 percent as well.
- In a highly-influential review of the development accounting literature, Caselli (2005) summarizes the empirical consensus on returns to school as 13 percent (for less than four years), 10 percent (for four to eight years) and 7 percent (for more than eight years). This parameterization was also adopted in the Penn World Tables 9.0 estimates of human capital (Feenstra, Inklaar and Timmer (2015)). Assuming that primary and secondary school each last six years, the baseline parameterization implies a return to primary school of $\frac{2}{3} \times 0.13 + \frac{1}{3} \times 0.10 = 0.12$ and a return to secondary school of $\frac{1}{3} \times 0.10 + \frac{2}{3} \times 0.07 = 0.08$.

The value of $\phi = 0.08$ is deliberately chosen to be at the low end of this range. This is because the vast majority of estimates of the return to school do not also control for health, while the human capital measure in Equation (1) is intended to reflect the partial effects of education and health on worker productivity. To the extent that empirical studies of the return to education are unable to control of health (or the factors determining health), the resulting estimates may overstate the partial effect of education on productivity. For more discussion of this point see Caselli (2014), who advocates using a conservative estimate of the return to school in a similar setting.

A final issue concerns the fact that the vast majority of empirical estimates of the returns to school use simply the number of years of education, and do not adjust for the quality of education. This may be a concern if the returns to learning-adjusted years of school differ from returns to years of school. One study that addresses this point is Hanushek and Zhang (2009), who generate quality-adjusted years of school using an approach similar to Evans and Yuan (2017), and then estimate Mincerian returns to quality-adjusted years of school in a sample of mostly OECD countries. While the country-by-country estimates of returns to quality-adjusted school years differ somewhat from the returns to unadjusted school years, on average they are fairly similar (Hanushek and Zhang (2009), Figure 2). This provides some justification for applying returns estimated using "raw" years of education to expected learning-adjusted years of school as is done here.

A3: Health

A3.1 Basic Methodology and Data Considerations

The measure of health z in the description of the HCI in Section A1 should be interpreted as a scalar index of "latent" health that summarizes the aspects of health that matter for worker productivity. Latent health cannot be observed directly, and so implementing the HCI requires a mapping from unobserved "latent" health z to directly-observable health indicator that serve as proxies, as well as an estimate of the corresponding "return" to health, γ. Weil (2007) proposes a strategy for doing so by recognizing that wages as well as observable summary indicators of health status such as adult height both respond to unobserved latent health. In the case of wages, this could reflect channels such as improved physical strength enabling greater work effort, as well as the effects of better health on better cognitive skills, both of which then are rewarded with higher wages.[15] A large literature has also argued that trends in average adult height within a country can serve as a proxy for trends in the overall health conditions in a country.[16] Poor *in utero* and early childhood nutrition and health lead to stunting among children, which in turn is reflected in shorter adult height, as well as a greater incidence of poor health outcomes among adults.

A key advantage of adult height as an observable health indicator is that there are many micro-econometric estimates of the "return" to height obtained from extended Mincer regressions of log wages on education, height and other controls. Weil (2007) develops a latent variable representation of the relationship between unobserved latent health, wages, and height, and demonstrates that this can be used to replace unobserved latent health and its return, $\gamma \times z$, with observed height and its estimated return, $\gamma_{HEIGHT} \times HEIGHT$, in the expressions for the health component of human capital in Equation (1). The interpretation in this case is not that height directly makes workers more productive. Rather, the correct interpretation is that if latent health improves in such a way that height increases by 1 cm, then this will lead to an increase in worker productivity of γ_{HEIGHT} percent, i.e. height serves as an observable proxy for unobservable latent health.

[15] In fact, Case and Paxson (2008) study data on test scores, height, and earnings in the US and the UK and conclude that all of the effect of height on earnings operates through cognition in their sample.
[16] Considering within-country over-time trends in height is crucial here, since cross-country differences in average stature also reflect cross-country differences in genetic predisposition for height among different ethnicities.

A practical problem with this approach is that cross-country data on adult height are relatively scarce. Moreover, the interpretation of the cross-country variation is clouded by genetic differences in populations of different countries. To address this problem, Weil (2007) suggests using a more widely-available summary indicator of health, adult survival rates (ASR). The basic insight is that ASR also improves within countries over time with improvements in latent health, in the same way that adult height does. As a result, the within-country over-time relationship between improvements in adult height and ASR can be used to transform the “return” to height into a “return” to ASR, and furthermore, ASR can be more meaningfully compared across countries .[17] Specifically, this means that $\gamma \times z$ in the expression for human capital can be replaced with $\gamma_{ASR} \times ASR$, where $\gamma_{ASR} = \gamma_{HEIGHT} \times \beta_{HEIGHT,ASR}$ and $\beta_{HEIGHT,ASR}$ is the slope coefficient from a regression of adult height on ASR. Intuitively, this slope coefficient captures how both adult height and adult survival rates improve when latent health improves, and this relationship can be used to convert the “return” to height into a “return” to ASR.[18] Again, however, the correct interpretation is not that there is a labour market “return” to adult survival rates. Rather, the correct interpretation is that when latent health improves to the extent that ASR increases by one percentage point, then worker productivity increases by γ_{ASR} percentage points.

A complementary strategy to solve the problem of data scarcity for adult height is to instead use data on stunting in childhood, which has become increasingly available, particularly in low-income countries where stunting is common and recognized as an important marker of poor early childhood development outcomes. Although country coverage of stunting data is less complete than for ASR, a benefit of stunting as a proxy for health is that there is direct evidence on the links between height in childhood and adult height. Evidence from cohort studies that track individuals over time provide evidence that height deficits in childhood persist into adulthood. This relationship can be used to create a link between stunting rates and likely future adult height, which analogously is referred to as $\beta_{HEIGHT,STUNTING}$. This can then be used to derive an alternative measure of the contribution of health to future adult productivity, $\gamma_{STUNTING} \times STUNTING$, where $\gamma_{STUNTING} = \gamma_{HEIGHT} \times \beta_{HEIGHT,STUNTING}$. The same caveats of interpretation apply to this measure as do to γ_{ASR}.

[17] Considering the within-country over-time variation controls for time-invariant factors contributing to cross-country differences in height.
[18] An underlying assumption here is that latent health is a single scalar index and that the observed indicators all respond to this single underlying measure. See Weil (2007), Section III.C for a discussion of the consequences of relaxing this assumption.

The HCI uses stunting and ASR as two alternative observable proxies for the overall health environment. Absent a strong view on which of these is a better proxy, in countries where both are available, a simple average of their contributions to productivity in the HCI is used, i.e. $\gamma \times z$ is replaced with $(\gamma_{ASR} \times ASR + \gamma_{STUNTING} \times STUNTING)/2$. In the (mostly richer) countries where data on stunting are not available, only $\gamma_{ASR} \times ASR$ is used. Since both $\gamma_{ASR} \times ASR$ and $\gamma_{STUNTING} \times STUNTING$ are measured in the same units, the unavailability of one or the other should not make a systematic difference for the calculation of the contribution of health to productivity.

Both ASR and stunting rates are imperfect proxies for latent health. The choice of these measures is largely dictated by the scarcity of alternative widely-accepted and broadly-available cross-country data on non-fatal health outcomes that could be used instead. The main alternative would be to use the measure of "years lived with disability" (YLD), that is constructed by the WHO and by the Institute for Health Metrics and Evaluation (IHME) Global Burden of Disease project. These estimates draw on available survey-based data on the prevalence of individual health conditions, that are then aggregated across conditions using disability weights. The measure of YLD, together with "years of life lost" (YLL) based on age-specific death rates due to various conditions, make up the well-known "disability-adjusted life years" (DALY) measure.[19]

Although the calculation of YLD reflects the best efforts of the profession to piece together and extrapolate the limited available direct information on non-fatal health outcomes into a comprehensive measure, it is not well-suited for the HCI for three main reasons:

- The first is that YLD nearly uncorrelated with per capita income across countries. This is to some extent by design. As noted above YLD is based on measures of the prevalence of non-fatal health conditions. As countries become richer and health care improves, conditions that

[19] To understand the mechanics of YLL, YLD, and DALY, consider this very stylized setting. Individuals with a expected years of life remaining get sick with probability q. Conditional on being sick, there are two outcomes (i) death, with probability m, and (ii) survival, but with disability that results from having been sick, and that lasts for a_s years. In this setting, $YLL = q \times m \times a$ measures years of life lost: i.e. the probability of getting sick and dying, $q \times m$, times the remaining number of years of life lost, a. Similarly, $YLD = q \times (1 - m) \times a_s \times d$ is the number of years lived with disability, i.e. the probability of getting sick but surviving, $q \times (1 - m)$, times the number of years the disability lasts, a_s, times a "disability weight" d between zero and one that measures the severity of the disability. Note that while both YLD and YLL decline when q, the incidence or risk of getting sick falls, YLL and YLD move in opposite directions when the mortality rate among the sick, m, decreases. In this case YLL falls but YLD increases since now more survivors live with the disability that follows from having been sick. Note that $DALY = YLL + YLD = q(m \times a + (1 - m) \times a_s \times d)$ unambiguously declines when both q and m decline (since $a_s \times d < a$).

previously were fatal become non-fatal, and the survivors of these conditions are recorded as living with any subsequent disability. This makes it challenging to use this particular measure as an indicator of health that leads to improvements in aggregate productivity.

- The second is that data on YLD are extremely heavily imputed and extrapolated to compensate for the scarcity of the primary sources measuring the relevant non-fatal health conditions. While such extrapolation is unavoidable if some estimate is required (no matter how imprecise)[20], it makes it difficult to use this measure in a policy advocacy index since the links from the imputed health indicators to primary country data sources recognizeable by policymakers in the country are complex. In fact, in many cases, imputed estimates of specific health conditions have no directly-measured counterparts in the country itself. [21] In this case, the policy advocacy role of the HCI can be undermined by relying on heavily-imputed data, since it risks "papering over" data gaps, reducing the incentives to fill these gaps with new data collection work.
- The third is that using imputed data makes it difficult to track and interpret changes over time in the index. To oversimplify, if health conditions are imputed using a measure of living standards such as GDP per capita, then changes in GDP per capita (i.e. growth) will lead to changes in imputed health indicators that may have little to do with actual health outcomes or policy interventions to improve them.

While these limitations of more comprehensive but imputed indicators of non-fatal health outcomes are real, the limitations of the data on ASR and stunting used in the HCI are real as well.   Measurement of ASR requires data on death rates by age. While these are readily available in countries with strong vital registries, in roughly the poorest quarter of countries such data are missing or incomplete. In these countries, the UN Population Division instead estimates death rates by age by linking together the limited available age-specific mortality data with "model life tables" capturing the typical pattern of distribution of deaths by age. As noted above, data on stunting are increasingly available, but primarily

[20] Commendably the GBD estimates of YLD are accompanied by confidence bands, which are very wide given the underlying data limitations. For example, for most countries the confidence bands for adult YLD encompass nearly the entire range of point estimates of YLD, which transparently indicates the substantial uncertainty associated with these estimates.

[21] These gaps are illustrated using metadata reporting the dates of the primary sources underlying the GBD estimates of the top 10 global causes of YLD. Taking 2017 as a reference point, calculate the fraction of countries for which no data source was available for each cause in the previous 10 years (so that the data necessarily were imputed/extrapolated beyond 10 years). This fraction is greater than 75 percent for developing countries, for 8 out of the top 10 global causes of YLD (anemia and diabetes are the only two exceptions).

for developing countries through household surveys such as the DHS, and typically are infrequently collected.

The rest of Section 3 details the specific steps of this procedure for calibrating the effects of health on worker productivity. Section A3.2 discusses micro-econometric estimates of the return to height, γ_{HEIGHT}. Section A3.3 discusses estimates of the relationship between adult height and adult survival rates, $\beta_{HEIGHT,ASR}$. Section A3.4 discusses how to calibrate estimates of the change in adult height attributable to reductions in stunting, $\beta_{HEIGHT,STUNTING}$.

A3.2 Estimates of the Return to Height

Weil (2007) uses a baseline estimate of $\gamma_{HEIGHT} = 0.034$, i.e. one additional centimeter of height raises earnings by 3.4 percent. This evidence is taken from two previous studies comparing height and earnings within twin pairs in the United States and in Norway. The main strength of these twin studies is their identification strategy of relying in random variations in birthweight between twins as a plausibly exogenous source of variation in their eventual adult heights. However, one might reasonably be concerned with the external validity of these findings, since these studies are based on pairs of twins in two advanced economies (Norway and the United States). To assess this concern, it is useful to briefly consider three other sets of estimates of returns to height from 19 other studies covering a range of developed and developing countries. All of these other estimates are based on instrumental variables regressions of log wages on height, with instruments of varying degrees of plausibility. Conditional on the validity of the instruments, these should all recover the effect of height on wages conditional on education, either because education is included in the regression, or because the instrument is uncorrelated with omitted education (under the null hypothesis that the exclusion restriction holds).

The first set can be found in Table 1 of Weil (2007), which reports estimates of the return to height conditional on education from three studies in Colombia, Ghana, and Brazil in the 8 to 9 percent per year range. The second set is summarized in Table 1 of Galasso and Wagstaff (2016), who summarize five studies in developing countries, none of which overlap with those in the first set. They arrive at mean return to height of about 1.5 percent per year. The third set of studies are summarized in Horton and Steckel (2011). Their Table 1 reports estimates from studies for 8 advanced economies not covered in the previous two sets, with a mean return to height of 0.5 percent per centimeter. Their Table 2 reports studies for developing countries. The three studies not included in the previous sets of results provide returns to height ranging from 1.4 percent to 4.5 percent per centimeter. In what follows, the Weil (2007) preferred value of 3.4 percent is used as the baseline. A reasonable range of estimates has 6.8 percent as the upper bound (corresponding to the mean estimated return to height across the 5 studies with estimates greater the Weil (2007) benchmark), and a lower bound of 1 percent as the lower bound (corresponding to the mean estimated return to height in the remaining 13 studies with estimated returns below 3.4 percent).[22]

A3.3 The Relationship Between Adult Height and Adult Survival Rates

The second key ingredient in the calculation is the estimated relationship between height and adult survival, $\beta_{HEIGHT,ASR}$. Weil (2007) estimates this using long-run historical data on stature and survival rates for 10 advanced economies over the 20th century, where there is considerable variation within countries over time in adult height. In his sample, average height varies from around 164 cm to 180 cm, and he obtains an estimate of $\beta_{HEIGHT,ASR} = 19.2$. To assess the robustness of this finding, the same relationship is estimated using data on female height collected in 172 DHS surveys covering 65 developing countries between 1991 and 2014.[23] In this sample, female height exhibits comparable variation to the historical dataset in Weil (2007), ranging from 148 cm to 163 cm. In the roughly half of the sample corresponding to non-Sub-Saharan African countries, a country-fixed effects regression of height on adult survival results in a slope coefficient of 19 and a standard error of 3.6, which is extremely close to the Weil (2007) baseline estimate of 19.2. In Sub-Saharan Africa, the slope coefficient is close to zero and very imprecisely estimated. This likely reflects the large swings in adult mortality rates due to the AIDS epidemic. To assess this, the relationship is re-estimated for all countries, but excluding observations above the Sub-Saharan Africa median for aids-related death rates. This gives a very similar estimate of 18.3 with a standard error of 3.4. These estimates lie in the same vicinity as the baseline estimate in Weil (2007). This evidence suggests that the Weil (2007) estimate of $\beta_{HEIGHT,ASR} = 19.2$ is reasonable to use in the baseline "return" to ASR of $\gamma_{ASR} = \gamma_{HEIGHT} \times \beta_{HEIGHT,ASR} = 0.034 \times 19.2 = 0.65$.

[22] All of the discussion here relies on the literature on the returns to height. While less extensive, there are well-identified econometric estimates of the economic returns to a number of other measures of health, including iron deficiency anemia, malaria, and hookworm infestation. Subsequent updates of this background paper will investigate how the estimates of overall health to productivity change when these other estimates of individual health conditions and their returns are used in this framework.

[23] The data come from the Health Equity and Financial Protection (HEFPI) Project at the World Bank. Patrick Eozenou and Adam Wagstaff kindly made this data available.

A3.4 The Relationship Between Stunting and Adult Height

An alternative approach to incorporating health into the human capital index is to use measures of stunting in childhood directly as the observed proxy for latent health. Stunting is measured as the fraction of children under five years old whose height is more than two reference standard deviations below the reference median, where the reference median and standard deviation are taken from WHO standards for normal healthy child development. Creating a link from stunting to the contribution of latent health to productivity, requires evidence on the relationship between the proportion of children who are stunted in childhood and average attained height of the population in adulthood. Combining this relationship with the estimated labour market returns to height creates a link from stunting in childhood to worker productivity in adulthood operating through the channel of increased height. This subsection discussed two complementary approaches to obtaining this relationship

The first is a calibration based on a simple variant on the calculations and estimates in Galasso and Wagstaff (2016). They cite a number of cohort studies that provide evidence that having been stunted as a child reduces attained adult height by approximately 6 centimeters. Under the assumption that average adult height conditional on stunting status in childhood does not change with the stunting rate, they calculate the change in average adult height due to the elimination of stunting as this difference of 6 centimeters multiplied by the fraction of the adult population that was stunted in childhood.

This estimate may however be conservative because it assumes no change in the adult height of children who were not initially stunted, even though these children are likely also to benefit from the improvements in health that reduce the proportion of children who are stunted. These wider effects can be captured with an alternative calibration of how the mean of the distribution of adult height shifts when childhood stunting falls. Let x represent adult height and q_c represent the fraction of adults who were stunted as children, i.e. $q_c \equiv P[x_c < z_c]$ where x_c denotes height in childhood when stunting is measured, and z_c represents the corresponding age-specific height threshold for stunting in childhood. Next consider three simplifying assumptions: (i) adult height is normally distributed, i.e. $x \sim N(\mu, \sigma^2)$; (ii) the fraction of adults who were stunted as children is the same as the fraction of children who were stunted when these adults were children, i.e. $q_c = q \equiv P[x < z]$, where z is the adult height threshold corresponding to z_c in childhood; and (iii) the ordering of children by height in the under-5 age group where stunting is measured persists into adulthood. Assumption (ii) enables the use of observed data on stunting in childhood to measure the proportion of adults who were stunted as children, although this requires abstracting from "catchup growth" as well as higher rates of mortality among stunted children, both of which would lead to $q < q_c$. Assumption (iii) ensures that the same group of individuals who were stunted as children are also stunted as adults. This assumption can be rationalized by the high correlation between childhood and adult height.

As noted above, data on q_c is available, which by Assumption (ii) is equal to stunting in adulthood, q. Estimates of the mean difference in adult height between adults who were not, and who were, stunted as children, d, also exist and Assumption (iii) ensures that adults who were stunted as children are also stunted as adults. Together with Assumption (i) of normality, this implies two moment conditions relating the data on q and d to the parameters of the distribution of adult height, μ and σ:

$$q = P[x < z] = F\left(\frac{z-\mu}{\sigma}\right) \tag{6}$$

$$d = E[x|x > z] - E[x|x < z] = \frac{\sigma f\left(\frac{z-\mu}{\sigma}\right)}{F\left(\frac{z-\mu}{\sigma}\right)\left(1 - F\left(\frac{z-\mu}{\sigma}\right)\right)} \tag{7}$$

where $F(.)$ and $f(.)$ denote the normal distribution and density functions, and (7) relies on the properties of the truncated normal distribution.[24]

These two equations can be used to calibrate the changes in average adult height μ associated with reduced stunting rates q. One simple way for doing so is to use Equation (6) to trace out the relationship between μ and q for a fixed value of the standard deviation of height, σ. Another way of doing so is to solve Equations (6) and (7) to obtain this expression for mean adult height as a function of the rate of stunting q:

[24] Specifically, $E[x|x < z] = \mu - \frac{\sigma f\left(\frac{z-\mu}{\sigma}\right)}{F\left(\frac{z-\mu}{\sigma}\right)}$ and $E[x|x > z] = \mu + \frac{\sigma f\left(\frac{z-\mu}{\sigma}\right)}{1-F\left(\frac{z-\mu}{\sigma}\right)}$.

(8) $$\mu = z - \frac{dq(1-q)F^{-1}(q)}{f\left(F^{-1}(q)\right)}$$

This expression can be used to trace out the relationship between μ and q for a fixed value of the mean difference in height between adults who were and were not stunted as children, d. Both of these methods can be contrasted with the assumption in Galasso and Wagstaff (2016) in which the only effect on adult average height comes through a reduction in the stunting rate, i.e. $\mu = z - dq$

Figure A3.4.1 Calibrated Relationship Between Adult Height and Stunting

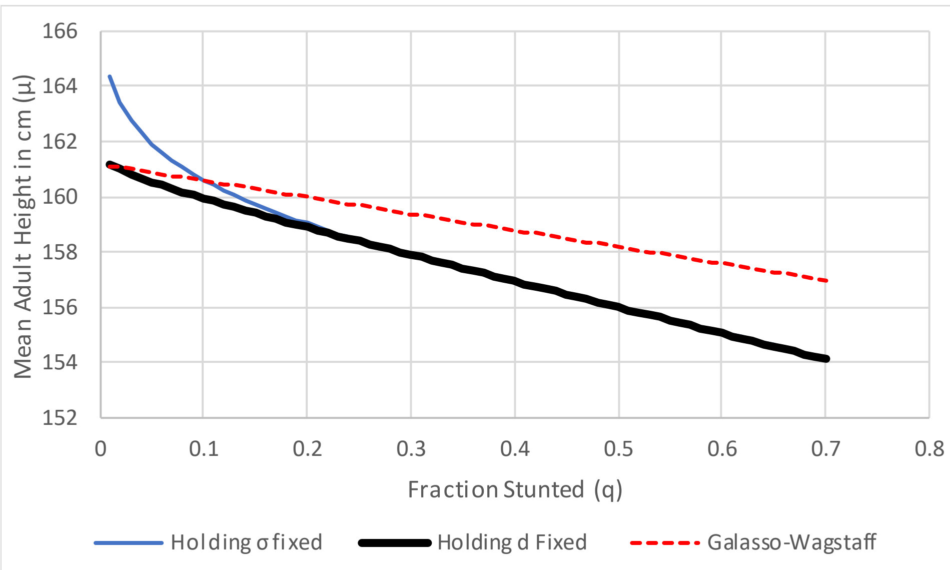


Figure A3.4.1 plots the relationship between average adult height (on the vertical axis) and stunting (on the horizontal axis) implied by these three methods. To plot these graphs, set $z = 156$ centimeters, corresponding to the WHO height-for-age z-score of -2 for 19 year-olds (average across male and female). The thin blue line plots the relationship between mean adult height and stunting holding fixed the standard deviation of height.[25] The heavy black line shows the same relationship in

[25] The value at which σ is held fixed matters for the calculation. To make the series comparable, set σ so that Equations (6) and (7) hold at a stunting rate of $q = 0.25$ and the observed height difference of $d = 6$ in the study by Victoria et. al. (2008) cited by Galasso and Wagstaff (2016). Victoria et. al. (2008) report on young adult health outcomes observed in the mid-2000s in cohort studies that have tracked respondents since childhood. Data from the WHO-UNICEF-WB Joint Malnutrition Estimates databased indicate that stunting rates in the early 1990s (when the respondents were children) in the five countries were 19% (Brazil), 66% (Guatemala), 62% (India), 43% (Philippines) and 32% (South Africa). This range of values for the stunting rate is represented in the horizontal axis of the figure. In contrast, the value of z does not matter for the analysis since it only shifts the relationship between stunting and adult height up and down, without changing the slope.

Equation (8), which holds fixed the height differential d, while the dashed red line shows the relationship assumed in Galasso and Wagstaff (2016) which holds fixed mean height among adults who were and were not stunted in childhood and varies only the proportion stunted. Except at low rates of stunting, the first two methods give a very similar relationship between mean adult height and stunting rates in childhood. Moreover, this relationship is steeper than the relationship assumed in Galasso and Wagstaff (2016). This is because their approach does not take into account the increases in height among individuals who were not initially stunted as the stunting rate declines. The slope of the dashed red line is $-d = -6$, while the average slope of the other two lines over the range where they coincide is -10.2. Consequently, a reduction in the rate of stunting q by ten percentage points raises attained adult height by 10.2×0.1 or approximately one centimeter, or approximately 40 percent larger than in the calibrations of Galasso and Wagstaff (2016).

The main advantage of this calculation is that it provides a very simple way to calibrate the response of mean adult height to stunting in childhood, using only data on childhood stunting rates and the estimate of the adult height differential from cohort studies. An alternative approach to inferring shifts in the mean of the distribution of height associated with reductions in stunting is to estimate them directly. This can be done using the same cross-country panel of DHS surveys described in the previous subsection. These surveys contain data on the incidence of stunting, as well as average attained height of children of different ages. A country-fixed-effects regression of average height of two-year-olds on the fraction of children who are stunted yields a slope coefficient of -0.12 and a standard error of 0.012. This implies that a reduction in the stunting rate of 10 percentage points is associated with an increase in average height among two-year-olds of 1.2 centimeters. Under the assumption that height deficits in two-year-olds persist into adulthood, this implies a reduction in average adult height of about the same amount. This estimate is slightly larger than but quite close to the one obtained by the calibration approach discussed above. To be conservative, the HCI uses the smaller of the two estimates by setting $\beta_{HEIGHT,STUNTING} = -10.2$, with an overall "return" to reduced stunting of $\gamma_{STUNTING} = \gamma_{HEIGHT} \times \beta_{HEIGHT,STUNTING} = -0.034 \times -10.2 = 0.35$.

A4 The Human Capital Index

A4.1: Putting the Pieces Together

This section draws together the discussion of the previous sections to summarize the overall HCI, which is the product of three components:

(9) $$HCI = Survival \times School \times Health$$

Using the notation from Equation (3), the three components of the index are formally defined as:

(10) $$Survival \equiv \frac{p}{p^*} = \frac{1 - Under\ 5\ Mortality\ Rate}{1}$$

(11) $$School \equiv e^{\phi(s_{NG}-s^*)} = e^{\phi(Expected\ Years\ of\ School \times \frac{Harmonized\ Test\ Score}{625} - 14)}$$

(12) $$Health \equiv e^{\gamma(z_{NG}-z^*)} = e^{(\gamma_{ASR}\times(Adult\ Survival\ Rate-1)+\gamma_{Stunting}\times(Not\ Stunted\ Rate-1))/2}$$

The baseline values for the returns to education and health are $\phi = 0.08$, $\gamma_{ASR} = 0.65$ and $\gamma_{Stunting} = 0.35$ as discussed in the previous sections. The probability of survival until age 5 is shown in **Figure 1**. The education component of the index is shown in **Figure 5**, and the health component of the index is shown in **Figure 6**, separately for adult survival rates and stunting. Expected learning-adjusted years of school range from around 3 years to close to 14 years in the best-performing countries. This gap in expected learning-adjusted years of school implies a gap in productivity relative to the benchmark of complete education of $e^{\phi(3-14)} = e^{0.08(-11)} = 0.4$, i.e. the productivity of a future worker in countries with the lowest expected years of learning-adjusted school is only 40 percent of what it would be under the benchmark of complete education.

For health, adult survival rates range from 60 to 95 percent, while the fraction of children not stunted ranges from around 60 percent to over 95 percent. Using ASR this implies a gap in productivity of $e^{\gamma_{ASR}(0.6-1)} = e^{0.65(-0.4)} = 0.77$, i.e. productivity of a future worker using the ASR-based measure of health is only 77% of what it would be under the benchmark of full health. Using the fraction of children not stunted, this implies a gap in productivity of $e^{\gamma_{Stunting}(0.6-1)} = e^{0.35(-0.4)} = 0.87$, i.e. productivity of a future worker using the stunting-based measure of health is only 87% of what it would be under the benchmark of full health.

The overall HCI is shown in **Figure 7**, and ranges from around 0.3 in the lowest countries to around 0.9 in the highest countries. This means that in countries with the lowest value of the human capital index, the expected productivity as a future worker of a child born today is only 30 percent of what it would be under the benchmark of complete education and full health.

A4.2: Robustness To Alternative Weights

The calibrated returns to education and health, i.e. ϕ , γ_{ASR}, and $\gamma_{Stunting}$, determine both the range of the HCI as well as the relative weights on education and health in the HCI. The higher are the returns to education and health, the greater are the productivity differences implied by the differences in learning-adjusted school years and health. In addition, higher (lower) values of the returns to health relative to education place greater (lower) weight on the health component of the HCI. To the extent that countries have different relative positions in the education and health measures included in the HCI, changing the relative weights on health and education can change countries' relative positions in the overall HCI. However, these changes in relative positions are not very large because, not surprisingly, the education and health measures included in the HCI are fairly highly correlated across countries.

This can be seen in **Figure 10**, which shows the correlation between the baseline HCI reported in **Figure 7** and three alternative versions corresponding to three alternative estimates of the return to height (which in turn feed into γ_{ASR} and $\gamma_{Stunting}$). The baseline assumed return to an additional centimeter of height is $\gamma_{Height} = 0.034$ or 3.4 percent. As discussed in Section A3.2, a reasonable range of values from the empirical literature goes from 1 percent to 6.8 percent. Alternative versions of the HCI using these estimates are shown in the top left and top right panels of **Figure 10**. They are correlated with the baseline HCI at 0.99 in both cases.

Another way of assessing the robustness of the index to alternative weighting schemes is to consider the (arbitrary) benchmark in which the education and health components of the index simply are assumed to have equally-large effects on worker productivity. Specifically, let s_{max} and s_{min} denote the largest and smallest observed values for learning-adjusted years of school across countries, and similarly let z_{max} and z_{min} denote the larges and smallest values of the health measure. Then setting

$\frac{\gamma}{\phi} = \frac{s_{max} - s_{min}}{z_{max} - z_{min}}$ corresponds to the assumption that moving from the bottom to the top of the distribution of countries in health has the same effect on worker productivity as moving from the bottom to the top of the distribution of education. The range of observed outcomes for learning-adjusted years of school is about 11 years, while the range of observed outcomes for adult survival rates is about 0.5, i.e. $\frac{\gamma}{\phi} = 22$. Using the baseline value of $\phi = 0.08$ and using $\gamma = \gamma_{ASR} = \gamma_{HEIGHT} \times \beta_{HEIGHT,ASR}$ implies $\gamma_{HEIGHT} = 0.09$ or 9 percent per centimeter (holding fixed $\beta_{HEIGHT,ASR} = 19.2$), which is much higher than is observed in the empirical literature. An alternative version of the HCI using this higher return to height, which in turn implies equal weights on education and health, is shown in the bottom-left panel of **Figure 10**. Again, the correlation with the baseline HCI is very high at 0.99. Overall this suggests that countries' relative positions on the HCI are fairly robust to changes in the calibrated returns to health and education that determine the relative weights on the components of the HCI.

A4.3: Gender Disaggregation

The components of the HCI, and therefore also the HCI itself, can be disaggregated by gender for 126 countries. Gender gaps are most pronounced for survival to age 5, adult survival, and stunting, where girls on average do better than boys in nearly all countries. Expected years of school is higher for girls than for boys in about two-thirds of countries, as are test scores. The gender-disaggregated overall HCI is shown in **Figure 9**. It is calculated by using the gender-disaggregated components to evaluate the overall HCI, while keeping the returns to health and education constant. As a result, the gender differences in this figure reflect only gender differences in the components of the HCI. Overall, HCI scores are higher for girls than for boys in the majority of countries. The gap between boys and girls tends to be smaller and even reversed among poorer countries, where gender-disaggregated data also is less widely available.

A4.4: Uncertainty Intervals for the HCI and Its Components

All of the components of the HCI are measured with some error, and this imprecision naturally has implications for the precision of the overall HCI. This section briefly describes how imprecision in the components of the HCI is measured, and the implications for imprecision in the overall HCI. Formal measures of imprecision are available for each of the components of the HCI, with the exception of expected years of school, as follows:

- Under-5 mortality rates: The UN Child Mortality Estimates program reports 90 percent uncertainty intervals for under-5 mortality rates. These uncertainty intervals reflect imprecision in the primary data sources (e.g. vital registries, household surveys, etc.) as well as imprecision attributable to the smoothing mechanism that is used to generate annual estimates of these rates. For the median country in 2017, the 90 percent uncertainty interval is equal to 0.01 or a range of 1 percentage point, while the median estimate of under-5 mortality is 2 percent. For countries with higher estimated mortality rates, the uncertainty intervals can be larger: for example, the 75th (90th) percentile of uncertainty intervals are 3.2 (5.3) percentage points wide.
- Harmonized learning outcomes (HLOs): As described above in Section A2.2, the calculation of HLOs involves the application of a test x subject x grade-specific conversion factor to the country-level average test score in its original units. This means that there are two distinct sources of uncertainty in the HLO calculation: (a) uncertainty around the country-level average scores in their original units, as reflected in the reported standard error around the country-level average , and (b) uncertainty in the calculation of the conversion factor. The HLO database quantifies the combination of these two sources of uncertainty through bootstrapping. Specifically, 1000 random draws are taken from the distribution of the test x subject x grade-specific original score at the country level, assuming that the country-level mean (across students) score is normally distributed. Then the HLOs are calculated using the 1000 samples of original scores, and the 2.5th and 97.5th percentiles of the resulting bootstrapped HLOs are reported as upper and lower bounds. The HLOs used in the HCI are further aggregated to the country-year level as described in Section A2.2. This aggregation is done using the reported HLO estimates at the test x subject x grade level, and then repeated using the lower and upper bounds of the test x subject x grade-level scores. The median HLO score as used in the HCI in 2017 is 424 TIMSS-equivalent points, and the median range of the uncertainty interval is fairly narrow at 12 points. However, this range is larger for testing programs such as PASEC and SACMEQ which have few "doubloon" observations on which the conversion factor is based, so that uncertainty coming from variation in the conversion factor is larger.
- Adult Survival Rates: Adult survival rates (ASR) are compiled by the UN Population Division using a similar process to the under-5 mortality rates described above. While there is uncertainty in the primary estimates of mortality as well as the process for data modeling, UNPD does not report uncertainty intervals. Instead, uncertainty intervals produced in the IHME Global Burden of Disease modelling process for ASR are used. The point estimates for ASR in the

IHME and UNPD data are quite similar for most countries. The ratio of the upper (lower) bound to the point estimate of ASR in the IHME data is applied to the point estimate of ASR in the UNPD data to obtain upper (lower) bounds on ASR. The median uncertainty interval is 4.4 percentage points wide, while the median adult survival rate is 86 percent. Uncertainty intervals are substantially smaller (larger) for countries with higher (lower) ASR. The 25th and 75th percentiles of the width of the uncertainty interval are 2.5 and 7.2 percentage points respectively.

- Stunting: The UNICEF-WHO-World Bank Joint Malnutrition Estimates reports 95 percent confidence intervals around estimates of stunting for about 40 percent of observations – primarily those where the JME team has access to the record-level survey data and can do reanalysis. These also correspond to the set of surveys for which gender-disaggregated stunting rates are available, and confidence intervals are reported for all gender-disaggregated rates. For the median observation, the 95 percent confidence interval is just under four percentage points wide. Absent better alternatives, for the remaining observations in the JME database, confidence intervals are imputed using the fitted values a regression of the width of the confidence interval on the stunting rate. Looking at the cross-section of most recently-available data for all countries in 2017, and after this imputation, the 95 percent confidence interval is 3.5 percentage points wide, while the median stunting rate is 22 percent.

Transforming the uncertainty intervals for the individual components of the HCI into uncertainty intervals for the overall HCI is complicated by the fact that there is no information on the joint distribution of uncertainty across components of the HCI. To see why this matters, note that if measurement error were uncorrelated across the different components, then the uncertainty intervals for the overall HCI would be smaller than for the components since over-estimates of some components would be offset by under-estimates of other components. If by contrast measurement error were highly correlated across components, then uncertainty intervals for the overall HCI would be larger than for the individual components, as over-estimates on one component would be compounded by over-estimates on other components, and vice versa.

Absent any information on the extent of correlation of measurement error across components, the HCI uses the simple approach of constructing a lower (upper) bound of the uncertainty interval for the overall HCI by assuming that each of the components is at its lower (upper) bound. This approach is conservative in the sense that it amounts to assuming that the measurement error is highly correlated across components of the HCI. On the other hand, these intervals understate the degree of uncertainty around the overall HCI scores because they do not capture (a) uncertainty around the estimates of expected years of school (for which uncertainty intervals are not available) and (b) uncertainty around the estimates of the returns to education and health that are used to transform the components of the HCI into contributions to productivity.

The resulting uncertainty intervals are shown in **Figure 8**, as vertical ranges around the value of the HCI for each country. The uncertainty intervals are moderate in size: the median uncertainty interval across all countries has a width of 0.03, while the HCI scores range from around 0.3 to 0.9. For some countries with less precise component data, the uncertainty intervals can be larger: the 75th and 90th percentiles of the width of the uncertainty interval are 0.04 and 0.05 respectively.

Although crude, these uncertainty intervals are a useful way of indicating to users that the values of the HCI for all countries are imprecise and subject to errors, reflecting the corresponding imprecision in the components. This should not be too surprising given the various limitations of the component data described in previous sections. The uncertainty intervals can also serve as an antidote against the tendency to over-interpret small differences between countries. While the uncertainty intervals constructed here do not have a rigorous statistical interpretation, they do signal that if for two countries overlap substantially, the differences between their HCI values are not likely to be all that practically meaningful. This is intended to help to move the discussion away from small differences in country ranks on the HCI, and towards more useful discussion around the level of the HCI itself and what it implies for the productivity of future workers.

A5: Linking the Human Capital Index To Future Income Levels and Growth

This section provides illustrative links from human capital to growth anchored in the logic of the development accounting literature (see for example Caselli (2005) and Hsieh and Klenow (2011)). It follows much of this literature in adopting a simple Cobb-Douglas form for the aggregate production function:

$$y = Ak_p^{\alpha}k_h^{1-\alpha} \tag{13}$$

where y is GDP per worker; k_p and k_h are the stocks of physical and human capital per worker; and A is total factor productivity; and α is the output elasticity of physical capital. When thinking about how changes in human capital may affect income levels in the long run, it is useful to re-write the production function as follows:

$$y = \left(\frac{k_p}{y}\right)^{\frac{\alpha}{1-\alpha}} A^{\frac{1}{1-\alpha}} k_h \tag{14}$$

In this formulation, GDP per worker is proportional to the human capital stock per worker, holding constant the level of total factor productivity and the ratio of physical capital to output, $\frac{k_p}{y}$. This formulation can be used to answer the following question: *"By how much does an increase in human capital raise output per worker, in the long run after taking into account the increases physical capital that is likely to be induced by the increase in human capital?"*. The answer to the question is that output per worker increases equiproportionately to human capital per worker, i.e. a doubling of human capital per worker will also lead to a doubling of output per worker in the long run.

Linking this framework to the human capital index requires a few further steps. First, following much of the existing literature, assume that the stock of human capital per worker that enters the production function, k_h, is equal to the human capital of the average worker.[26] Second, the human

[26] This is by no means innocuous, because it embodies the strong assumption that workers with different levels of human capital are perfectly substitutable after taking into account their individual productivity differences. To take a highly simplified and memorable example (due to David Weil) of where perfect substitutability breaks down, note that although the educational human capital of four unskilled workers probably is lower than that of one PhD,

capital of the next generation, as measured in the HCI, and the human capital stock that enters the production function, need to be linked. This can be done by considering the scenarios outlined in the main text. Imagine first a "status quo" scenario in which the expected learning-adjusted years of school and health as measured in the HCI today persist into the future. Over time, new entrants to the workforce with "status quo" health and education will replace current members of the workforce, until eventually the entire workforce of the future has the expected learning-adjusted years of school and level of health captured in the current human capital index. Let $k_{h,NG} = e^{\phi s_{NG}+\gamma z_{NG}}$ denote the future human capital stock in this baseline scenario. Contrast this with a scenario which the entire future workforce benefits from complete education and enjoys full health, resulting in a higher human capital stock $k_h^* = e^{\phi s^*+\gamma z^*}$.

It is possible to compare eventual steady-state GDP per worker levels in the two scenarios using Equation (14), assuming that levels of TFP and the physical capital-to-output ratio are the same in the two scenarios, to obtain:

$$\frac{y}{y^*} = \frac{k_{h,NG}}{k_h^*} = e^{\phi(s_{NG}-s^*)+\gamma(z_{NG}-z^*)} \tag{15}$$

This expression is the same as the human capital index in Equation (3), except for the term corresponding to survival to age 5 (since children who do not survive do not become part of the future workforce). This creates a close link between the human capital index and growth. Disregarding the (small) contribution of the survival probability to the HCI, Equation (15) says that a country with an HCI equal to x could have future GDP per worker that would be $1/x$ times higher in the future if its citizens enjoyed complete education and full health (corresponding to $x = 1$). For example, a country such as Morocco with a HCI value of aroudn 0.5 could in the long run have future GDP per worker in this scenario of complete education and full health that is $\frac{1}{0.5} = 2$ times higher than in the status quo scenario. What this means in terms of average annual growth rates of course depends on how "long" the long run is. For example, under the assumption that it takes 50 years for these scenarios to

the four unskilled workers are undoubtedly more productive when it comes to moving a piano. See Jones (2014) for alternative human capital aggregators that relax the assumption of perfect substitutability. Jones (2014) argues that allowing for complementarities between workers of different skill levels substantially increases the role of human capital in accounting for cross-country income differences. However, Caselli and Ciccone (2017) point out that this interpretation ignores the important role of cross-country differences in productivity in driving the skill premia that in turn drive the conclusions in Jones (2014).

materialize, then a doubling of future per capita income relative to the status quo corresponds to roughly 1.4 percentage points of additional growth per year.

The calibrated relationship between the HCI and future income levels described here is simple because it focuses only on steady-state comparisons. In related work, Collin and Weil (2018) elaborate on this by developing a calibrated growth model that traces out the dynamics of adjustment to the steady state. They use this model to trace out trajectories for per capita GDP and for poverty measures for individual countries and global aggregates, under alternative assumptions for the future path of human capital. They also calculate "equivalent" increases in investment rates in physical capital that would be required to deliver the same increases in output associated with improvements in human capital.

References

Altinok, Nadir, Noam Angrist, and Harry Patrinos (2018). "Global Dataset on Education Quality 1965-2015". World Bank Policy Research Working Paper No. 8314.

Ashraf, Quamrul, Ashley Lester and David Weil (2009). "When Does Improving Health Raise GDP?", in Daron Acemoglu, Kenneth Rogoff and Michael Woodford, eds. NBER Macroeconomics Annual.

Case, Ann and Christina Paxson (2008). "Stature and Status: Height, Ability, and Labour Market Outcomes". Journal of Political Economy. 116(3): 499-532.

Caselli, Francesco (2005). "Development Accounting". Handbook of Economic Growth.

Caselli, Francesco (2014). "The Latin American Efficiency Gap". Manuscript, London School of Economics.

Caselli, Francesco and Antonio Ciccone (2017). "The Human Capital Stock: A Generalized Approach – Comment". Manuscript, London School of Economics.

Collin, Matthew, and David N. Weil. (2018). "The Effect of Increasing Human Capital Investment on Economic Growth and Poverty: A Simulation Exercise." World Bank Policy Research Working Paper No. 8590.

De Philippis, Marta and Federico Rossi (2017). "Parents, Schools, and Human Capital Differences Across Countries". Manuscript, Johns Hopkins University.

Evans, David and Fei Yuan (2017). "The Economic Return to Interventions That Increase Learning". Background Paper for World Development Report (2018). https://www.riseprogramme.org/sites/www.riseprogramme.org/files/inline-files/Evans%20Returns%20to%20Learning%20-%20v2017-06-09%20CLEAN.pdf.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" American Economic Review, 105(10), 3150-3182, available for download at www.ggdc.net/pwt.

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

Gakidou, Emmanuela, Krycia Cowling, Rafael Lozano and Chris Murray (2010). “Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis”. The Lancet. 376(9745):18-24.

Galasso, Emanuela and Adam Wagstaff (2016). “The Economic Costs of Stunting and How to Reduce Them”. World Bank Policy Research Note 17/5.

Gao, Qiufeng, Yaojiang Shi, Hongmei Yi, Cody Abbey, and Scott Rozelle (2017). “Reading Achievement in China’s Rural Primary Schools: A Study of Three Provinces”. Stanford University, Freeman Spogli Institute Working Paper, available at https://fsi.stanford.edu/publication/reading-achievement-chinas-rural-primary-schools-study-three-province.

Hall, Robert and Chad Jones (1999). “Why Do Some Countries Produce So Much More Output Per Worker Than Others?”. Quarterly Journal of Economics. 114(1):83-116.

Hanushek, Eric and Lei Zhang (2009). "Quality-Consistent Estimates of International Schooling and Skill Gradients." Journal of Human Capital 3(2): 107-143.

Hanushek, Eric, Jens Ruhose, and Ludger Woessmann (2015). “Human Capital Quality and Aggregate Income Differences: Development Accounting for US States”. Hoover Institutions Economics Working Paper No. 15112.

Horton, Sue and Richard Steckel (2011). Copenhagen Consensus Assessment Paper on Malnutrition.

Hsieh, Chang-Tai and Peter Klenow (2010). “Development Accounting”. American Economic Journal (Macroeconomics). 2(1):207-223.

Jedwab, Remi and Asif Mohammed (2018). “Returns to Education and Experience”. Background Paper for 2019 World Development Report, in progress.

Jones, Benjamin (2014). “The Human Capital Stock: A Generalized Approach”. American Economic Review. 104(11):3752-3777.

Jorgenson, Dale W., and Barbara M. Fraumeni (1989). “The Accumulation of Human and Nonhuman Capital, 1948-1984. In Robert E. Lipsey and Helen S. Tice, eds. The Measurement of Savings, Investment and Wealth Studies in Income and Wealth. Chicago: University of Chicago Press.

Kaarsen, Nicolai (2014). “Cross-Country Differences in the Quality of Schooling”. Journal of Development Economics. 107: 215-224.

Klenow, Peter and Andres Rodriguez-Clare (1997). “The Neoclassical Revolution in Growth Economics: Has It Gone Too Far?”. NBER Macroeconomics Annual.

Lagakos, David, Benjamin Moll, Tommaso Porzio, Nancy Qian and Todd Schoellman (2018). "Life Cycle Wage Growth Across Countries". Journal of Political Economy, forthcoming.

Levitt, Steven, John List, Susanne Neckerman and Sally Nadoff (2016). "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance." American Economic Journal: Economic Policy. 8(4): 183-219.

Manuelli, Rodolfo and Ananth Seshadri (2014). "Human Capital and the Wealth of Nations". American Economic Review. 104(9):2736-2762

Mincer, Jacob (1958). "Investment in Human Capital and Personal Income Distribution". Journal of Political Economy. 66(4):281-302.

Montenegro, Claudio and Harry Patrinos (2014). "Comparable Estimates of Returns to Schooling Around the World". World Bank Policy Research Working Paper No. 7020.

OECD (2016). "PISA 2015 Results (Volume 1). Excellence and Equity in Education". Paris, OECD. http://dx.doi.org/10.1787/9789264266490-en.

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.

Ravallion, Martin (2011). "Mashup Indices of Development". World Bank Research Observer. 27:1-32.

Schoellmann, Todd (2012). "Education Quality and Development Accounting". Review of Economic Studies. 79:388-417.

UN (2016). "Guide on Measuring Human Capital" (https://www.unece.org/index.php?id=44704).

UNESCO, World Bank, UNICEF and Global Partnership for Education (2014). "Education Sector Analysis Methodological Guidelines". (https://www.globalpartnership.org/content/methodological-guidelines-education-sector-analysis-volume-1).

United Nations University (2012). "Inclusive Wealth Report". Cambridge: Cambridge University Press.

Victora, C. G., et al. (2008). "Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital. The Lancet. 371(9609):340-57.

Weil, David (2007). "Accounting for the Effect of Health on Economic Growth". Quarterly Journal of Economics. 156-1306.

World Bank (2018). "The Changing Wealth of Nations".