The World Bank Human Capital Index: A Guide

Aart Kraay

This paper provides a guide to the new World Bank Human Capital Index (HCI), situating its methodology in the context of the development accounting literature. The HCI combines indicators of health and education into a measure of the human capital that a child born today can expect to achieve by her 18^{th} 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.

JEL codes: I1, I2, O1, O4
Keywords: human capital, health, education, development accounting.

Introduction

Investments in human capital deliver substantial economic returns 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. To address this incentive problem, in October 2018 the World Bank published a new human capital index (HCI) designed to call attention to the future economic consequences of shortfalls in investment in human capital, especially for the young.[1] This paper describes the HCI, situating its methodology in the extensive literature on development accounting.

The HCI measures the human capital that a child 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 follows 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 may not

The World Bank Research Observer

doi: 10.1093/wbro/lkz001 34:1–33

survive to her fifth birthday. Even if she does reach school age, there is a further risk that she may not start school, let alone complete the full cycle of 14 years of school from preschool 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 HCI quantitatively illustrates the key stages in this trajectory and their consequences for the productivity of the next generation of workers. The HCI consists of three components: (1) survival, measured as the probability of survival to age five; (2) school, which combines a measure of the number of years of school a child born today can expect to attain given prevailing enrollment rates with a measure of the quality of education based on international student achievement tests; and (3) health, which uses childhood stunting rates and adult survival rates as proxies for the overall health environment.

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. Rather than rely on ad hoc aggregation with arbitrary weights, the HCI uses techniques from the literature on development accounting to convert measures of health and education into contributors to worker productivity, relative to a benchmark of 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. The contributions of education and health to productivity are calibrated using microeconometric estimates of the labor market returns to education and health. The resulting HCI ranges from zero to one, and a value of x means that a child born today can expect to be only $x \times 100$ percent as productive as a future worker as she would be under the benchmark of complete education and full health.

Although the HCI is intended primarily as a tool to communicate in a simple way the likely future economic consequences of shortfalls in education and health among the young, its units and methodology are anchored in the extensive academic literature on development accounting. This literature quantifies the contributions of physical and human capital to output differences across countries through an accounting framework centered on an aggregate production function that transforms factors of production into output per worker. Given assumptions on the shape of the aggregate production function and measures of factors of production, this literature sheds light on the relative importance of factors of production and levels of productivity as proximate sources of income differences across countries. Much of the attention in this literature has focused on developing suitable measures of factors of production. The HCI draws heavily on techniques from this literature to quantify the contributions of education and health to worker productivity. In addition to summarizing the methodology of the HCI, this paper selectively reviews the relevant parts of this literature on measuring human capital. More comprehensive reviews of the development accounting literature in general can be found in Caselli (2005) and Hsieh and Klenow (2010). Bleakley (2010), and Weil (2014) provide reviews focusing on health, and Rossi (2018) provides a recent review focusing on human capital.

The rest of this paper proceeds as follows. The next section summarizes the basic framework for measuring human capital used in the development accounting literature. The paper next discusses in further detail the measurement of education and health, and their corresponding labor market returns and briefly describes the resulting HCI. The final three sections describe how the machinery of development accounting creates a direct connection between the HCI and future income levels and growth, relate the HCI to a number of other existing measures of human capital and development, and provide conclusions and caveats.

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 and healthier she is. This gain in productivity represents the contributions of health and education to her human capital. A simple expression for the human capital of an individual worker *i* is:

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

where s_i represents her education measured in years of school (adjusted for quality or learning); z_i is a measure of her health; and the parameters ϕ and γ represent the returns to education and health, that is, $\phi \times 100$ and $\gamma \times 100$ are the percent increases in human capital attributable to a unit increase in education and health. The following two sections discuss how s_i and z_i are measured, and how the corresponding returns are retrieved from existing microeconometric evidence on the returns to education and health.

Equation (1) is the building block for various aggregate measures of the stock of human capital of the *current* workforce used in the development accounting literature. Typically, the objective in this literature is to develop a measure of the aggregate human capital stock that can be used in an aggregate production function to document the contribution of human capital to aggregate output.[3] A common approach in this literature is to simply assume that there is a representative agent equipped with the levels of education and health corresponding to the average outcomes observed in the current workforce. Inserting these average outcomes into equation (1) then results in a measure of aggregate human capital per worker that can be used in an aggregate per worker production function.

In contrast, the HCI measures the expected *future* human capital of a child born today, given *current* education and health outcomes for the young. This forward-looking emphasis reflects the objective of the HCI to draw attention to the consequences for future productivity of shortfalls of current investment in the human capital of children. Measures of the human capital of the existing workforce are less effective in this respect – particularly for education. This is because the educational attainment of the current workforce primarily reflects the educational opportunities that were available to current workers in the past when they were school-aged children, and now are largely beyond the influence of current and future policy interventions. Instead, the HCI measures how current health and education outcomes – which are amenable to improvement through current and future policy efforts – shape the likely future human capital of children born today.

This measure of expected future human capital (with subscript NG representing the future or "next generation" of workers) is defined as:

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

where s_{NG} and z_{NG} represent expected future education and health; and p is the probability that a child born today will survive.[4] Multiplying by p captures the loss in future productivity per child born today due to premature mortality, since the human capital as a future worker of a child who dies is not realized. Expected future education and health are measured based on the assumption that current education and health outcomes persist into the future. For example, expected future education is measured as the number of learning-adjusted years of school a child progressing through the education system is likely to obtain by her 18^{th} birthday, given prevailing enrollment rates and test scores as a measure of learning. Similarly, expected future health is proxied using current health outcomes, under the assumption that current health conditions will prevail into the future.

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 $s^* = 14$ learning-adjusted years of school, corresponding to the maximum possible number of years of school achieved by age 18 by a child who starts school at age 4. The benchmark corresponding to full health is z^* and is discussed later in this paper.

With this notation, the HCI is:

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

The HCI is the product of three easily interpretable components, each measuring productivity relative to the benchmark of complete education and full health. 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 forgone future productivity due to children completing less than a full 14 learning-adjusted 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.

Empirically implementing this expression requires data on child survival rates (p), expected years of school adjusted for learning (s_{NG}) and expected health (z_{NG}), as well as values for the returns to health and education parameters ($\phi,\ \gamma$). Data on child survival rates is readily available from the UN Interagency Group on Child Mortality Estimation. Obtaining data on education and health, together with estimates of the corresponding returns, is more challenging and is the subject of the next two sections.

Measuring the Education Component of the HCI

Quantity of Education

This section discusses the education component of the HCI, consisting of the learning-adjusted number of years of school that a child who starts school at age 4 would attain by her 18$^{\text{th}}$ birthday. Calculating expected years of school is conceptually straightforward: It is simply the sum of age-specific enrollment rates over the age range of interest (see, e.g., UNESCO et al. (2014), section 2.2 and annex 2.2). Age-specific enrollment rates can be estimated from household survey data, or they can be obtained from administrative records produced by ministries of education.

A practical consideration, however, is that age-specific enrollment rates from either source are not systematically available for a broad cross-section of countries. Instead, more readily available and frequently updated data on enrollment rates by level of school can be used to approximate enrollment rates in different age brackets. Specifically, preprimary enrollment rates are used to 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. Cross-country differences 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.

Enrollment rates by level of school are curated and reported by the UNESCO Institute for Statistics (UIS). 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 UIS 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. If TNER data are not available, ANER, NER, or GER are used instead to approximate age-specific enrollment rates, in that order of preference. For most countries, levels, and years, repetition rates are also available, and enrollment rates are adjusted for repetition where possible to avoid counting a repeated year of school as an additional year of education.

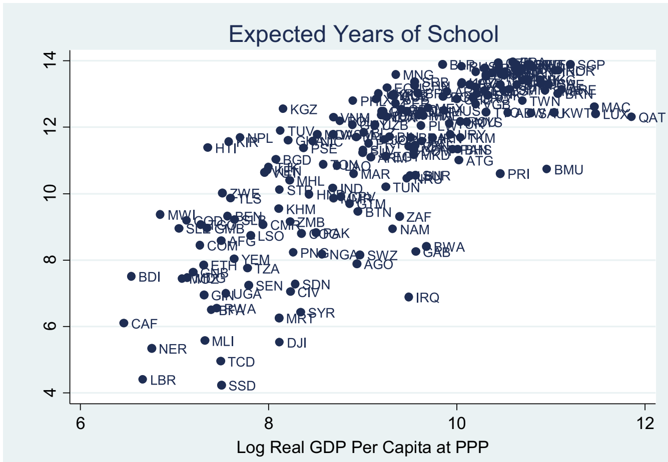
The resulting measure of expected school years approximates the number of years of school that a child can expect to attain by her 18th birthday if she starts school at age 4, for a maximum of 14 years. Data on this measure of expected years of school is reported in the top panel of figure 1. Conceptually this calculation corresponds to the measure of "school life expectancy" (SLE) calculated and reported by UIS. However, the implementation here differs in that UIS uses GERs to calculate school life expectancy, whereas in the HCI, TNER is used wherever possible. This choice involves 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. Moreover, the TNER is by construction between 0 and 1, while in the UIS data GERs often exceed 1, sometimes by a substantial margin. This clouds the interpretation of expected years of school as calculated by UIS, which in about one-quarter of countries exceeds the statutory duration of school due to the high GERs used in the calculation.

Quality of Education

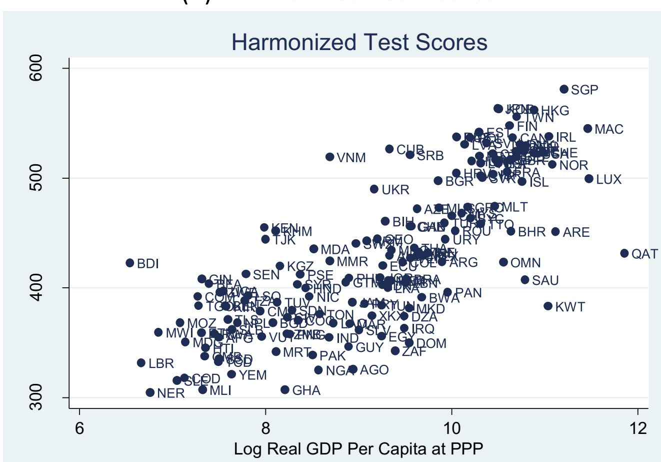
The school quality adjustment is based on a new large-scale effort to harmonize international student achievement tests from several multicountry testing programs, described in Patrinos and Angrist (2018). This project assembles data from (a) three major international testing programs (Trends in International Maths and Science Study

Figure 1. Quantity and Quality of Education Data

(A) Expected Years of School By Age 18



(B) Harmonized Test Scores



Source: Author's calculations as described in note below and in main text.

Note: 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; that is, 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, Inklaar, and Timmer 2015), with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for all countries.

(TIMSS), Progress in International Reading Literacy Study (PIRLS), and Programme for International Student Assessment (PISA)); (b) 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)); and (c) the Early Grade Reading Assessments (EGRA) program covering primarily low-income countries. The expanded dataset covers over 160 countries.

Test scores from these different testing programs are harmonized into common units, with TIMSS (for math and science) and PISA and PIRLS (for reading, at the secondary and primary level, respectively) serving as the numeraires. The 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 average 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. Patrinos and Angrist (2018) refer to the resulting test scores in common units as "harmonized learning outcomes (HLOs)." Harmonization is done at the subject x grade level. The country-level test scores used in the HCI average across subjects and grades.

Finally, the country-level test scores from different testing programs are combined into a single time series for each country in a way that balances coverage and comparability over time. Data from TIMSS and PIRLS, both of which are conducted by the same parent organization, the International Association for the Evaluation of Educational Achievement (IEA) and use common units, are first combined by shifting PIRLS rounds to the nearest TIMSS round and averaging. These are combined with PISA assessments by taking whichever of the two are available in a given country-year, or the average of the two in 2003 and 2015 when PISA and TIMSS rounds coincided. These major international assessments contribute 521 of the 657 country-year observations in the HCI dataset. Next, scores the three regional testing programs (SACMEQ, PASEC, and LLECE) are used in years where data from international testing programs are not available, contributing a further 82 country-year observations. Finally, EGRAs provide information for a further 54 country-year observations, including in 27 countries for which this is the only available test. The bottom panel of figure 1 reports the most recently available harmonized learning outcome for the set of 157 countries included in the HCI. Harmonized test scores range from around 300 in the worst-performing countries to near 600 in the best-performing countries.

It should be noted that the approach to test score harmonization in Patrinos and Angrist (2018), based on ratios of average scores across testing programs in the common set of countries participating in both programs, is an approximation. Fully psychometrically robust linking of testing programs requires that different testing programs share a common subset of questions (or "linking items"). Comparison of student performance on the linking items across tests, and across linking items and other questions in the same test, would provide a more rigorous basis for test harmonization. Unfortunately, however, such linking items do not yet exist across all of the testing programs used by Patrinos and Angrist (2018), and in order to maximize cross-country coverage, approximations such as harmonization based on the ratio of average test scores are unavoidable.

Adjusting Expected Years of School for Quality

The harmonized international test score data described in the previous subsection is used to adjust expected years of school for quality. The approach taken here follows World Bank (2017), and is developed further in Filmer et al. (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. They explain how this scaling factor can be justified by the assumption that grade-learning trajectories are linear through the origin; that is, the stock of learning as measured by a test administered in a given grade reflects the linear accumulation of knowledge from a base of zero at the beginning of school.

A natural benchmark value for top performance is a TIMSS score of 625, which corresponds to the TIMSS "advanced" international benchmark. With this benchmark, learning-adjusted expected years of school is defined as $s_{NG} = EYS \times (\frac{HLO}{625})$ where EYS is expected future years of school described below and HLO is the harmonized learning outcome described below. Recall that country-level average test scores range from around 300 in the worst-performing countries to 600 in the best-performing countries. At the bottom end of this range, the quality-adjustment factor is $300/625 = 0.48$. 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.

This is not the only approach to incorporating test scores into measures of human capital. Some parts of the literature have treated test scores as a measure of cognitive ability that contributes to human capital directly and independently of the level of education (see, e.g., Caselli (2005, 2014); Vogl (2014); Hanushek, Ruhose, and Woessman (2015), and the summary in Rossi (2018)). Other parts of this literature have proposed a variety of other ways to infer the equivalence between "quantity" and "quality" of school. For example, some studies have exploited the fact that the same test may be administered to students of the same age but at different grade levels. This provides suggestive evidence of how an additional year of school maps into higher test scores (see, e.g., OECD (2016) using PISA scores, and

Kaarsen (2014) using TIMSS data). A drawback of this approach is selection: students in a higher grade-for-age may have higher abilities than their peers in the grade below. Another approach is to compare performance on tests applied to adults with different levels of educational attainment as children, thereby avoiding the selection problem (e.g., Hanushek and Zhang (2009), Evans and Yuan (2017)). Yet another approach is to use calibrated theory to infer quality from optimal choices by individuals to invest in education given their knowledge of its quality (e.g., Schoellman (2012), Manuelli and Seshadri (2014)).

Unfortunately, estimates of the years-of-school-equivalent between test scores of 300 and 600 vary widely across these different approaches, and the literature has not yet reached a clear consensus. Absent such a consensus, the HCI uses the mapping from test scores to quality differences suggested by Filmer et al. (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 may however be questionable in countries where there is substantial 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 et al. 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 U.S. 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 comparisons across children in the same schools.

Returns to Education

Values for the return to education parameter ϕ can be anchored in the vast empirical literature that estimates Mincer (1958)-style regressions 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.

This value can be situated in the context of the literature on measuring returns to education. For example, 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 that are most relevant for the HCI gives a value of 8.9 percent. 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 $\frac{1}{3} \times 0.10 + \frac{2}{3} \times 0.07 = 0.08$, for an overall average return of 10 percent.

The choice of $\phi = 0.08$ at the low end of this range is deliberate. The vast majority of estimates of the return to education 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 for health (or the factors determining health) that likely are positively correlated with education, the resulting estimated returns 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 education in a similar setting.

Measuring the Health Component of the HCI

Implementing the health component of the HCI is complicated by the fact that there is no single widely accepted and directly measured scalar metric of health status that is comparable to learning-adjusted years of school as a summary measure of education, either at the individual or at the country level. The most basic indicator of the health environment, mortality, tends to be fairly directly and regularly measured in the majority of countries that have well-functioning vital registries to record deaths. Beyond this, however, systematic direct measurement of nonfatal health outcomes is much more sparse, and typically focuses on individual diseases or conditions, often measured infrequently and only in certain countries.

The HCI follows the development accounting literature in interpreting health measure z in the HCI as a scalar index of "latent" health that summarizes the aspects of health that matter for worker productivity (Weil (2007), Ashraf, Lester, and Weil (2009)). Specifically, the HCI follows the strategy in Weil (2007), which recognizes that wages as well as observable proxies of health status such as adult height 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.[5] 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.[6] Poor *in utero* and early childhood nutrition and health lead to stunting among children, which in turn is reflected in reduced 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 expression for human capital in equation (1). The interpretation 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, that is, height serves as an observable proxy for unobservable latent health.

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. A further benefit is that ASR can be more meaningfully compared across countries than adult height. 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. Again, however, the correct interpretation is not that there is a labor 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.

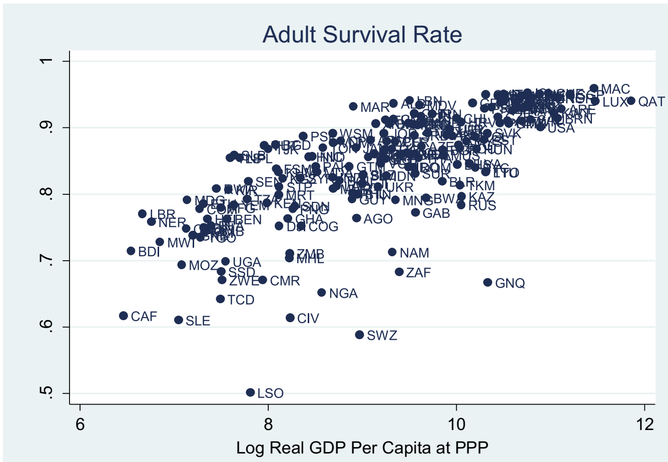
A complementary strategy to address the problem of data scarcity for adult height is to instead use data on stunting in childhood, which has become increasingly available, particularly in low-income countries where stunting is common and is recognized as an important marker of poor early childhood development outcomes. Although country coverage of stunting data is less complete than for ASR, a benefit of using 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 provides 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}.

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; that is, $\gamma \times z$ 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 does not make a systematic difference for the calculation of the contribution of health to productivity.

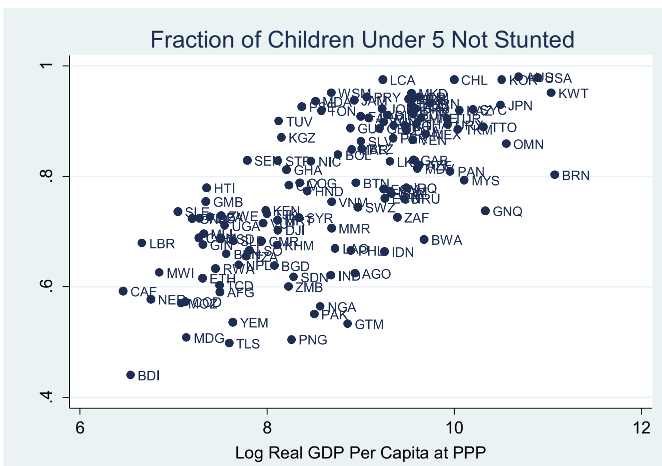
The cross-country variation in these two proxies for health is displayed in figure 2, which plots adult survival rates (top panel) and the share of children under five who are not stunted (bottom panel) against log GDP per capita, using the most recently available data for each country. Adult survival rates range from lows of 60 percent in countries with the worst health outcomes to around 95 percent in countries with the best outcomes. The proportion of children not stunted also exhibits considerable variation, ranging from around 50 percent in the countries with the worst child health outcomes to close to 100 percent in the countries with the best outcomes.

Figure 2. Health Indicators

(A) Adult Survival Rate



(B) Fraction of Children Under 5 Not Stunted



Source: Author's calculations as described in note below and in main text.

Note: 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, Inklaar, and Timmer 2015), with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for all countries.

The rest of this section details the specific steps needed to transform this cross-country variation in health proxies into estimates of the effects of latent health on worker productivity. It first discusses microeconometric estimates of the return to height, γ_{HEIGHT}. It then discusses estimates of the relationship between adult height and adult survival rates, $\beta_{HEIGHT,ASR}$. Finally, it discusses how to calibrate estimates of the change in adult height associated with a given reduction in stunting rates, $\beta_{HEIGHT,STUNTING}$. While in principle both the returns to height and the relationship between height and other observable proxies for health could vary across countries, the HCI uses the same values of these parameters for all countries. As noted above for education, this ensures the variation in the HCI across countries reflects only variation in its component measures of health and education outcomes.

Estimates of the Return to Height

Weil (2007) uses a baseline estimate of $\gamma_{HEIGHT} = 0.034$; that is, 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 birth weight 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, given their particular rich country settings. To assess this concern, it is useful to briefly consider 3 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 value of 3.4 percent preferred by Weil (2007) 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 five studies with estimates greater than 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). Although this range of estimates is uncomfortably wide, the variation in estimates is also not all that surprising as they come from an array of different studies in different settings and countries and with different control variables and identification strategies. The consequences for the HCI of allowing the return to height to vary over this range is discussed below.

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 twentieth 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 demographic and health surveys (DHS) covering 65 developing countries between 1991 and 2014.[7] 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 countries other than Sub-Saharan African, 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 $\gamma_{ASR} = \gamma_{HEIGHT} \times \beta_{HEIGHT,ASR} = 0.034 \times 19.2 = 0.65$.

The Relationship Between Childhood Stunting and Adult Height

The stunting rate is a fairly widely available anthropometric measure of childhood health that serves as an alternative proxy for latent health in the HCI. 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 labor market returns to height creates a link from stunting in childhood to worker productivity in adulthood.

This link between childhood stunting and adult height can be created in a variety of ways. The simplest way to do so is described 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. That is, a reduction in stunting by 10 percentage points raises adult height by 0.6 cm.

This estimate is, however, 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 the overall health environment that reduce the proportion of children who are stunted. These wider effects can be captured by calibrating how the mean of the distribution of adult height shifts when childhood stunting falls. This calibration is described in detail in the working paper version of this article (Kraay (2018), appendix A3.4). The basic idea however is simple. Assume that adult height is normally distributed and that the fraction of adults below normal height-for-age is the same as the corresponding fraction when these adults were children; that is, the childhood stunting rate. The mean and standard deviation of the distribution of adult height can be retrieved from two moment conditions: (a) the 6 cm average height difference between adults who were, and were not, stunted in childhood; and (b) the childhood stunting rate. These moment conditions can be solved for the mean of the distribution of adult height as a function of the childhood stunting rate. This calibration suggests that a reduction in the stunting rate by 10 percentage points raises adult height by 1.02 cm.

A third 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 $\gamma_{STUNTING} = \gamma_{HEIGHT} \times \beta_{HEIGHT,STUNTING} = -0.034 \times -10.2 = 0.35$.

The Human Capital Index

Putting the Pieces Together

Drawing together the discussion in the previous sections, the overall HCI is the product of three components:

$$HCI = Survival \times School \times Health \tag{4}$$

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

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

$$School \equiv e^{\phi(s_{NG}-s^*)} = e^{\phi\left(Expected\ Years\ of\ School \times \frac{Harmonized\ Test\ Score}{625} - 14\right)} \tag{6}$$

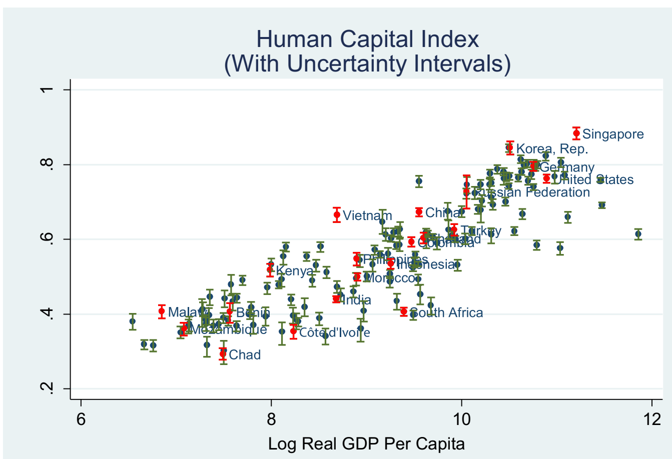
$$Health \equiv e^{\gamma(z_{NG}-z^*)} = e^{\left(\gamma_{ASR} \times (Adult\ Survival\ Rate - 1) + \gamma_{Stunting} \times (Not\ Stunted\ Rate - 1)\right)/2} \tag{7}$$

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. 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 $e^{\phi(3-14)} = e^{0.08(-11)} = 0.4$; that is, 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 high-quality 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 $e^{\gamma_{ASR}(0.6-1)} = e^{0.65(-0.4)} = 0.77$; that is, productivity of a future worker using the ASR-based measure of health is only 77 percent 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 $e^{\gamma_{Stunting}(0.6-1)} = e^{0.35(-0.4)} = 0.87$; that is, productivity of a future worker using the stunting-based measure of health is only 87 percent of what it would be under the benchmark of full health.

The overall HCI is shown in figure 3, 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. 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 five, 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 calculated by using the gender-disaggregated components to evaluate the overall HCI, while assuming that the returns to health and education are the same for both males and females. 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.

Figure 3. The Human Capital Index



Source: Author's calculations as described in note below and in main text.
Note: 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, Inklaar, and Timmer 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 labeled for illustrative purposes.

Uncertainty Intervals

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. Formal measures of imprecision are available for each of the components of the HCI, with the exception of expected years of school. In the case of child and adult survival rates, measures of uncertainty are a feature of the data modelling process that generates such estimates. In the case of stunting, the measures are survey-based, and for most countries uncertainty intervals around the estimate of stunting are reported in the UNICEF-WHO-World Bank compilation of stunting rates used in the HCI. For test scores, Patrinos and Angrist (2018) report bootstrapped uncertainty intervals around country-average harmonized scores, which reflect not just sampling variation around the mean for each country, but also variation in the calculation of the conversion factors between testing programs. Further details on these measures of uncertainty for the HCI components are in the working paper version of this article (Kraay (2018), appendix A4.4).

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 overestimates of some components would be offset by underestimates of other components. If, by contrast, measurement error were highly positively (negatively) correlated across components, then uncertainty intervals for the overall HCI would be larger (smaller) than for the individual components, as overestimates on one component would be compounded (offset) by over- (under-) estimates on other components.

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 as ranges around the HCI values for each country in figure 3. 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 caution against the tendency to overinterpret small differences between countries. While the uncertainty intervals constructed here do not have a rigorous statistical interpretation, they do signal that if they overlap substantially for two countries, 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 toward more useful discussion around the level of the HCI itself and what it implies for the productivity as future workers of children born today.

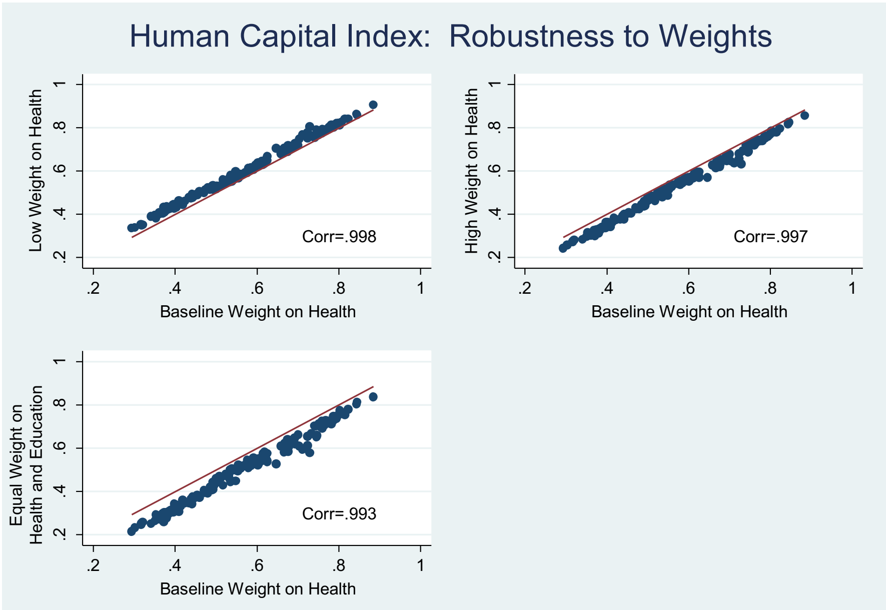
Robustness to Alternative Weights

The calibrated returns to education and health, that is, 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 the returns to education and health are, 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 4, which shows the correlation between the baseline HCI and three alternative versions corresponding to three alternative estimates of the return to height (which in turn feed into the calibrated returns to the two proxies for health, 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 the previous section 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 4. 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 are simply assumed to have equally large effects on worker

Figure 4. Effect of Changing the Weight on Health in the Human Capital Index



Source: Author's calculations as described in note below and in main text.
Note: 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. 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, Inklaar, and Timmer 2015), with missing countries filled using data from World Bank estimates of GDP at PPP. Graph shows the most recent data for all countries.

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 largest 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, that is, $\frac{\gamma}{\phi} = 22$. 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 4. 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.

Connecting the HCI to Future Income Levels and Growth

The HCI can be connected to future aggregate income levels and growth following the logic of the development accounting literature. A basic insight from this literature is that in the long run (that is, in the steady state of a neoclassical growth model), GDP per worker is proportional to human capital per worker, with the factor of proportionality depending on the steady-state capital/output ratio and the level of total factor productivity. This insight can be used to analyze counterfactual questions, such as *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?* Under the assumption that the level of total factor productivity and the steady-state capital/output ratio is unchanged, the answer is that output per worker increases equiproportionately to human capital per worker; that is, 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 HCI requires a few further steps to bridge the gap between the stock of human capital that enters the production function, and the forward-looking measure of human capital of the next generation of workers measured by the HCI. First, 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. Second, the human capital of the next generation and the human capital stock that enters the production function need to be linked. This can be done by considering different scenarios. Imagine first a "status quo" scenario in which the expected years of quality-adjusted 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 years of quality-adjusted school and level of health captured in the current human capital index; that is, $k_{H,NG} = e^{\phi s_{NG} + \gamma z_{NG}}$ is the future human capital stock in this baseline scenario, where s_{NG} represents the number of quality-adjusted school years of the next generation of workers, and γz_{NG} is short-hand notation for the contribution of the two health indicators to productivity in the HCI. Contrast this with a scenario in which the entire future workforce benefits from complete education and enjoys full health, resulting in a higher human capital stock $k_H^* = e^{\phi s^* + \gamma z^*}$, where s^* represents the benchmark of 14 years of high-quality schooling, and z^* represents the benchmark of full health.

Assuming that levels of total factor productivity (TFP) and the physical capital-to-output ratio are the same in the two scenarios, the ratio of the eventual steady-state GDP per worker levels in the two scenarios is just $\frac{y}{y^*} = \frac{k_{H,NG}}{k_H^*} = e^{\phi(s_{NG}-s^*)+\gamma(z_{NG}-z^*)}$. This expression is the same as the human capital index, except for the term corresponding to survival to age five (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 future income levels and growth rates. Setting aside the contribution of the survival probability to the HCI, this expression 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 an HCI value of around 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 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.

Comparison with Other Measures of Human Capital and Development

This section briefly highlights key methodological differences between the HCI and three other closely related cross-country measures of human capital or human development: (1) the Human Development Index (HDI) constructed by the United Nations Development Program, (2) the human capital index recently constructed by the Institute for Health Metrics and Evaluation (IHME) as described in Lim et al. (2018), and (3) the measure of the monetary value of human capital in the World Bank's Changing Wealth of Nations (CWON) report, described in Lange, Wodon, and Carey (2018).

The HDI and IHMEI are similar to the HCI in the sense that they are composite indicators that combine measures of health and education into a single summary index. The HDI is a geometric average of (a) the logarithm of Gross National Income per capita, (b) life expectancy at birth, and (c) the average of expected future years of school and average years of educational attainment of the adult population. Each of the components is expressed as a fraction of a corresponding benchmark value, and so the overall HDI can be interpreted as measuring average distance from these benchmark values. The IHMEI is a measure of expected years lived between age 20 and 64, adjusted for educational attainment, learning, and functional health status proxied using estimates of the prevalence of seven different health conditions. The adjustment factors are expressed as ratios of the education, learning, and health measures relative to the corresponding best possible values, and then life expectancy is discounted by the product of these ratios.

Since measures of health and education tend to increase strongly with per capita income, it is not very surprising that the HCI, HDI, and IHMEI are positively correlated with each other and with per capita income. In the set of 153 countries where all three measures are available, the correlations of the HCI, HDI, and IHMEI with log real GDP per capita are 0.86, 0.95, and 0.87 respectively, while the pairwise correlation between (a) the HCI and HDI is 0.94, (b) the HCI and IHMEI is 0.95, and (c) the HDI and IHMEI is 0.94. These very high positive correlations, however, obscure conceptual differences in their components, how they are measured, and how they are combined into the overall index.

The components of the indices are well described in the documentation of the respective measures. Differences between them can be briefly summarized as follows. The main differences between the HDI and HCI components are that (a) the HDI includes per capita income as one of its components while the HCI does not; (b) the HDI includes a measure of average educational attainment of the entire workforce while the HCI includes only a forward-looking measure of expected years of school; (c) the HCI adjusts educational attainment for quality using data from international testing programs while the HDI does not; and (d) the HCI includes stunting and adult survival rates as proxies for health, while the HDI uses life expectancy at birth. Distinctions between the components of the IHMEI and the HCI are less pronounced. Both indexes contain measures of survival, the quantity and quality of education, and health. The main difference in the components is that the IHMEI contains a larger number of health indicators.

The more interesting differences between the HCI, HDI, and IHMEI concern how the components are measured and how they are aggregated. On the measurement front, a distinguishing feature of the IHMEI is its use of heavily imputed health and education data, taken from the Global Burden of Disease project and related work at IHME. This project takes data from existing surveys and studies that report information on specific health and education outcomes for a particular age/sex/location/period. It then uses statistical modeling techniques to extrapolate to other age/sex/location/period cells for which direct measurement is not available.

The details of the extrapolation methodology vary from indicator to indicator, but typically extrapolations are based on (a) trends in the actual data, (b) estimated relationships with a limited number of more widely available covariates of the indicator of interest, such as per capita income, and (c) outcomes in geographically nearby locations. Large-scale imputation allows the IHMEI to cover 195 countries with annual data between 1990 and 2016. However, gaps in the underlying directly measured data are substantial. For example, the IHMEI data on learning is based on largely the same set of international testing data used in Patrinos and Angrist (2018). This implies that only about 20 percent of country-year observations on learning in IHMEI correspond to a year in which a test was actually taken, and for 65 countries in the IHMEI no directly measured testing data at all are used. Instead, the missing data are extrapolated across countries and within countries over time using estimates of educational attainment as a covariate.

While data extrapolation is a valuable tool to address data gaps, the use of extrapolated data has implications for analysis and policy dialogue. On the analytical front, one concern is the extent to which patterns in imputed data are driven by patterns in the underlying covariates used in the imputation model rather than actual relationships in the unobserved missing data. For example, some of the health indicators used in the IHMEI are imputed across countries and over time using models that include income as a covariate. In this case, the strong positive correlations between changes in log per capita income and changes in the IHMEI reported in Lim et al. (2018) may be difficult to interpret since they reflect a combination of the actual relationship between changes in human capital and growth, and relationships driven by the fact that some of the components of IHMEI are imputed using income and other covariates that move with income.[8]

The use of imputed data may also have implications for policy engagement. At a basic level, the apparent pervasiveness of imputed data on outcomes of interest could dull incentives to invest in the hard work of direct measurement of the actual outcomes. Imputation may also create perverse incentives for policy effort. For example, in a growing economy, policymakers can spuriously take credit for improvements in health and education outcomes that are imputed based on covariates that improve with per capita income, even if there are no actual improvements in the outcomes themselves. Similarly, knowledge of the imputation technology can distort incentives towards improving the covariates in the imputation model rather than the outcome itself. For example, as noted above, data on the quality of education, that is, test scores, in IHMEI are imputed using data on the quantity of education; that is, average attainment. With such imputed data, improvements in quantity will be rewarded as improvements in quality even if the latter did not happen, and this in turn may undermine incentives to improve quality. Until major investments in direct measurement of health and education quality are made, striking an appropriate balance between the obvious appeals of imputed data and these potential shortcomings is a likely to continue to be a challenge for efforts to measure human capital across countries.

Turning to aggregation methods, the HCI is anchored in the development accounting literature and is measured in units of productivity that are calibrated based on microeconometric evidence on the returns to education and health. The HDI is intended to summarize average achievement across its three dimensions, which it interprets as measures of the "capabilities" of members of society. However, as Ravallion (2012) notes that the use of a geometric average implies tradeoffs across components of the HDI that are difficult to rationalize from a welfare standpoint. The IHMEI is described as a measure of human capital. However, unlike the HCI, the IHMEI aggregation method is arbitrary and does not reflect available evidence on the magnitude of contributions of health and education to worker productivity, clouding the interpretation of the units of the IHMEI. Consider for example functional health, which in the IHMEI is measured using the first principal component of the prevalence of seven health conditions, each rescaled to run from zero to one. The resulting health indicator runs from 0.36 in the least healthy country to 0.90 in the most healthy country. Since this components enters the IHMEI multiplicatively, this means that human capital as measured by IHMEI is $\frac{0.90}{0.36} = 2.5$ times higher in the country with the best health outcomes, holding constant the other components of the index. The magnitude of this difference is difficult to reconcile with the more modest contribution of cross-country differences in health to human capital differences across countries in the HCI, which imply that better health improves productivity by a factor of only about 20 percent, consistent with available microeconometric evidence. In fact, to rationalize the 2.5-fold differences in human capital associated with health differences across countries in the IHMEI requires an assumed return to health that is four times as large as is suggested by microeconometric evidence discussed above.[9]

Finally, the HCI is conceptually closely related to measures of the monetary value of human capital based on the present value of future earnings of individuals such as those in CWON. This is because of the close link between the aggregation method of the HCI and the microeconometric literature on the labor market returns to education and health. To see this, suppose that log wages of individual i at some future time t are given by a health-augmented Mincer equation like $ln\, w_{it} = \phi s_i + \gamma z_i + g_i t$, 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). Measures of human capital along these lines have a long history (see, e.g., Jorgenson and Fraumeni (1989)), and are extensively discussed in the context of satellite national accounts in UN (2016). In addition to CWON, 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).

One incremental difficulty in constructing these measures 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 into the denominator of this measure, small changes in assumed growth rates are magnified into large changes in measured human capital. For example, if the discount rate is 5 percent, changing the assumed growth rate from 3 to 4 percent per year has the effect of doubling measured human capital. In practice, CWON calculates the financial value of human capital by taking aggregate labor income from the national accounts, and applying a common difference between the discount rate and the growth rate of $\delta - g_i = 0.015$ for all countries. This means that measured human capital roughly[10] is a fixed multiple of aggregate labor income from the national accounts. Since the variation in labor income shares in GDP is small relative to the variation in GDP per capita across countries, the CWON measure is strongly correlated with GDP per capita across countries.

8 Conclusions and Caveats

This paper has provided a guide to the methodology for a new World Bank Human Capital Index that was launched in October 2018. The HCI measures the amount of human capital that a child born today can expect to achieve by her 18th birthday, given currently prevailing health and education outcomes. The HCI is grounded in the development accounting literature, and uses microeconometric estimates of the returns to education and health to combine education and health outcomes into a measure of expected future worker productivity relative to the benchmark of complete education and full health. With its grounding in the development accounting literature, it is straightforward to connect the HCI to expected future income differences attributable to differences in human capital.

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. Noncognitive skills are not explicitly captured, although they may contribute directly and indirectly to human capital formation (see, e.g., Lundberg (2017)). Child and adult survival rates are imprecisely estimated in countries where vital registries are incomplete or nonexistent.

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.

Notes

Aart Kraay is a member of the World Bank Development Research Group. His email is akraay@worldbank.org. An earlier version of this paper was written 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 go to Rachel Glennerster (United Kingdom Department for International Development), Bill Maloney, Mamta Murthi, and Martin Raiser for peer review comments; to the WBRO editorial board and three anonymous referees for helpful comments; to Chika Hayashi United Nations International Children's Emergency Fund (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 United Nations Educational, Scientific and Cultural Organization (UNESCO), and Said Ould Ahmedou Voffal (UNESCO) for help with enrollment data; to Patrick Eozenou and Adam Wagstaff for help with DHS data; to Kathryn Gilman Andrews for help with analyzing the IHMEI data; and to Krycia Cowling Institute for Health Metrics and Evaluation (IHME), Nicola Dehnen, and Ritika D'Souza for tireless research assistance. Valuable comments were provided by Sudhir Anand (Oxford), George Alleyne Pan-American Health Organization (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 Interamerican Development Bank (IDB). This paper has also benefited 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 on June 11, 2018. The data used in this paper have benefited 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 Human Capital Index (HCI). The HCI and a summary of its methodology

was published in the 2019 World Development Report and the accompanying special report on human capital. Data on the HCI and its components are available at www.worldbank.org/humancapitalproject. *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. See World Bank (2018) for further discussion of the rationale for a human capital index.

2. Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999) are early examples of the development accounting approach, and Caselli (2005); Hsieh and Klenow (2010), and section 4 of Rossi (2018) provide surveys. See also Caselli (2014) for an application of this methodology to Latin America. 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.

3. The development accounting literature has for the most part focused on accounting for cross-sectional differences in GDP per worker at a given point in time. The closely related and much older growth accounting literature has focused on the role of changes over time in factors of production and productivity in accounting for changes over time in GDP per worker over time. Seminal contributions include Solow (1957) and Denison (1962), Barro (1999) provides a survey.

4. Writing expected future human capital as a function of expected future education and health requires a few technical assumptions. Formally, let h_{t+k} represent human capital at some future date $t + k$. Expected future human capital is given by $E_t[h_{t+k}] = pE_t[e^{\phi s_{t+k}}]E_t[e^{\gamma z_{t+k}}] \geq pe^{\phi E_t[s_{t+s}]}e^{\gamma E_t[z_{t+s}]}$, 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 $E_t[e^{\phi s_{t+k}}]$ and $E_t[e^{\gamma z_{t+k}}]$ 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 that would be required to calculate $E_t[e^{\phi s_{t+k}}]$ and $E_t[e^{\gamma z_{t+k}}]$, 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.

5. In fact, Case and Paxson (2008) study data on test scores, height, and earnings in the United States and the UK and conclude that all of the effect of height on earnings operates through cognition in their sample.

6. 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.

7. 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.

8. This problem of extrapolated data analysis uncovering patterns attributable to the extrapolation model rather than actual patterns in the data is not inevitable. There is a well-developed statistical literature on the analysis of imputed data that proposes techniques to avoid these difficulties (see, e.g., Rubin 1996). The basic idea is to generate multiple versions of the imputed dataset, and then average the estimates of relationships of interest across these multiple datasets, including variation across datasets in the measures of uncertainty of these estimates.

9. To see this, consider the contribution of health to productivity, using adult survival rates (ASR) as the proxy for health. In the HCI this is calibrated as $e^{\gamma_{ASR} \times (ASR-1)}$ where γ_{ASR} measures the increase in productivity associated with an improvement in health that raises adult survival rates. ASR ranges from 0.60 to 0.95 across countries in the HCI. To convert this difference into a productivity difference of 2.5 requires $\gamma_{ASR} = \frac{\ln(2.5)}{0.35} = 2.6$ which is four times larger than the value used in the HCI.

10. The CWON calculation also takes into account differences in life expectancy across countries when calculating the present value of labor income.

References

Ashraf, Q., A. Lester, and D. Weil. 2009. "When Does Improving Health Raise GDP?" In *NBER Macroeconomics Annual, Volume 23*, edited by Acemoglu Daron, Rogoff Kenneth and Woodford Michael, 157–204. Chicago: University of Chicago Journals.

Barro, R. 1999. "Notes on Growth Accounting." *Journal of Economic Growth* 4 (2): 119–37.

Bleakley, H. 2010. "Health, Human Capital, and Development." *Annual Review of Economics* 2: 283–310.

Case, A., and C. Paxson. 2008. "Stature and Status: Height, Ability, and Labour Market Outcomes." *Journal of Political Economy* 116 (3): 499–532.

Caselli, F. 2005. "Development Accounting." In *Handbook of Economic Growth, Volume 1*, edited by Philippe Aghion and Steven N. Durlauf, 679–741. Amsterdam: North-Holland.

———. 2014. "The Latin American Efficiency Gap." Manuscript, London School of Economics, London, UK.

Collin, M., and D. 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, World Bank, Washington, DC.

Denison, E. 1962. *The Sources of Economic Growth in the United States and the Alternatives Before US*. Washington, DC: Committee for Economic Development.

De Philippis, M., and F. Rossi. 2017. "Parents, Schools, and Human Capital Differences Across Countries." Manuscript, Department of Economics, Johns Hopkins University, Baltimore, MD.

Evans, D., and F. 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, R. C., R. Inklaar, and M. P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105 (10): 3150–82.

Filmer, D., H. Rogers, N. Angrist, and S. Sabarwal. 2018. "Learning-Adjusted Years of Schooling (LAYS): Defining a New Macro Measure of Education." World Bank Policy Research Working Paper No. 8591, World Bank, Washington, DC.

Galasso, E., and A. Wagstaff. 2016. "The Economic Costs of Stunting and How to Reduce Them." World Bank Policy Research Note 17/5, World Bank, Washington, DC.

Hall, R., and C. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114 (1): 83–116.

Hanushek, E., J. Ruhose, and L. Woessmann. 2015. "Human Capital Quality and Aggregate Income Differences: Development Accounting for US States." Hoover Institutions Economics Working Paper No. 15112, Stanford, CA.

Hanushek, E., and L. Zhang. 2009. "Quality-Consistent Estimates of International Schooling and Skill Gradients." *Journal of Human Capital* 3 (2): 107–43.

Horton, S., and R. Steckel. 2011. "Malnutrition, Copenhagen Consensus Assessment Paper," Copenhagen.

Hsieh, C.-T., and P. Klenow. 2010. "Development Accounting." *American Economic Journal (Macroeconomics)* 2 (1): 207–23.

Jorgenson, D. W., and B. M. Fraumeni. 1989. “The Accumulation of Human and Nonhuman Capital, 1948–1984.” In *The Measurement of Savings, Investment and Wealth Studies in Income and Wealth, Volume 52*, edited by R. E. Lipsey and H. S. Tice, 273—331. Chicago: University of Chicago Press.

Kaarsen, N. 2014. “Cross-Country Differences in the Quality of Schooling.” *Journal of Development Economics* 107: 215–24.

Klenow, P., and A. Rodriguez-Clare. 1997. “The Neoclassical Revolution in Growth Economics: Has It Gone Too Far?” *In NBER Macroeconomics Annual 1997, Volume 12*. edited by Ben S. Bernanke and Julio Rotemberg, 73—114. Cambridge, MA: MIT Press.

Kraay, A. 2018. “Methodology for a World Bank Human Capital Index.” World Bank Policy Research Working Paper No. 8593, World Bank, Washington, DC.

Lange, G.-M., Q. Wodon, and K. Carey. 2018. *The Changing Wealth of Nations 2018: Building a Sustainable Future*. Washington: The World Bank.

Levitt, S., J. List, S. Neckerman, and S. Nadoff. 2016. “The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance.” *American Economic Journal: Economic Policy* 8 (4): 183–219.

Lim, S. S., R. L. Updike, A. S. Kaldjian, R. M. Barber, K. Cowling, H. York, J. Friedman, R. Xu, J. L. Whisnant, H. J. Taylor, A. T. Leever, Y. Roman, M. F. Bryant, J. Dieleman, E. Gakidou, and C. J. L. Murray. 2018. “Measuring Human Capital: A Systematic Analysis of 195 Countries and Territories 1990—2016.” *Lancet* 392 (10154): 1217–34.

Lundberg, S. 2017. “Non-Cognitive Skills as Human Capital.” Manuscript, Department of Economics, University of California Santa Barbara.

Manuelli, R., and A. Seshadri. 2014. “Human Capital and the Wealth of Nations.” *American Economic Review* 104 (9): 2736–62.

Mincer, J. 1958. “Investment in Human Capital and Personal Income Distribution.” *Journal of Political Economy* 66 (4): 281–302.

Montenegro, C., and H. Patrinos. 2014. “Comparable Estimates of Returns to Schooling Around the World.” World Bank Policy Research Working Paper No. 7020, World Bank, Washington, DC.

Organisation for Economic Co-operation and Development (OECD). 2016. *PISA 2015 Results*, vol. 1, *Excellence and Equity in Education*. Paris, OECD. http://dx.doi.org/10.1787/9789264266490-en.

Patrinos, H. A., and N. Angrist. 2018. “A Global Dataset on Education Quality: A Review and an Update (1965—2018).” World Bank Policy Research Working Paper No. 8592, World Bank, Washington, DC.

Ravallion, M. 2012. “Troubling Tradeoffs in the Human Development Index.” *Journal of Development Economics* 99 (2): 201–209.

Rossi, F. 2018. “Human Capital and Macroeconomic Development: A Review of the Evidence.” Manuscript, Johns Hopkins School of Advanced International Studies, Johns Hopkins University, Baltimore, MD.

Rubin, D. 1996. “Multiple Imputation After 18+ Years.” *Journal of the American Statistical Association* 91 (434): 473–89.

Schoellmann, T. 2012. “Education Quality and Development Accounting.” *Review of Economic Studies* 79: 388–417.

Solow, R. M. 1957. “Technical Change and the Aggregate Production Function.” *Review of Economics and Statistics* 39: 312–20.

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

Vogl, T. 2014. “Height, Skills, and Labor Market Outcomes in Mexico.” *Journal of Development Economics* 107: 84–96.

Weil, D. 2007. “Accounting for the Effect of Health on Economic Growth.” *Quarterly Journal of Economics* 122 (3): 1265–1306.

———. 2014. “Health and Economic Growth.” In *Handbook of Economic Growth, Volume 2*, edited by P. Aghion and S. Durlauf, 623–82. Elsevier: North Holland.

World Bank. 2017. *Learning to Realize Education's Promise*. Washington: World Bank.

———. 2018. *The Human Capital Project*. Washington: The World Bank. https://openknowledge.worldbank.org/handle/10986/30498.