

# **Measuring Human Capital**

By

**Noam Angrist**

**Simeon Djankov**

**Pinelopi K. Goldberg**

**Harry A. Patrinos**

**DISCUSSION PAPER NO 783**

**February 2019**

**Any opinions expressed here are those of the authors and not necessarily those of the FMG. The research findings reported in this paper are the result of the independent research of the authors and do not necessarily reflect the views of the LSE.**

# Measuring Human Capital

By NOAM ANGRIST, SIMEON DJANKOV, PINELOPI K. GOLDBERG  
AND HARRY A. PATRINOS\*

*Students around the world are going to school but are not learning – an emerging gap in human capital formation. To understand this gap, we introduce a new dataset measuring learning in 164 countries and territories. The data covers 98 percent of the world’s population from 2000 to 2017. The dataset will be publicly available and updated annually by the World Bank. We present several stylized facts in a first application of the data: (a) while enrollment has increased worldwide, learning has stagnated (b) girls outperform boys on learning – a positive gender gap – in contrast to a negative gender gap observed for schooling (c) learning is associated with growth on a global scale (d) associations with growth are heterogenous (e) human capital accounts for up to a third of cross-country income differences – a middle ground in the recent development accounting literature. These stylized facts demonstrate the potential of our data to reveal new insight into the relationship between human capital and economic development. (JEL I20, O40, O15, H40, H52, J24, P50)*

\* Angrist: University of Oxford, and World Bank, 1818 H Street NW, Washington, DC 20433 (e-mail: noam.angrist@bsg.ox.ac.uk); Djankov: London School of Economics, and World Bank, 1818 H Street NW, Washington, DC 20433 (e-mail: sdjankov@worldbank.org); Goldberg: Yale University, and World Bank, 1818 H Street NW, Washington, DC 20433 (e-mail: pgoldberg@worldbank.org); Patrinos: World Bank 1818 H Street NW, Washington, DC 20433 (e-mail: hpatrinos@worldbank.org). We are grateful to Syedah Aroob Iqbal and Husein Abdul-Hamid for research support. This work builds on co-authored work with Nadir Altinok. Particular thanks to Aart Kraay for detailed comments and contributions to the methodology. Valuable comments were provided by Eva L. Baker, Eduardo Cascallar, Paul Collier, Stefan Dercon, Deon Filmer, Roberta Gatti, Rachel Glennerster, Daniel Koretz, Silvia Montoya, George Psacharopoulos, Heiner Rindermann, Halsey Rogers, Jaime Saavedra, Shwetlena Sabarwal, Justin Sandefur, and Eduardo Velez. This paper benefited from seminars held at the World Bank, World Congress of Cliometrics, American Economic Association, IDB, Oxford, USAID and FHI360. A great number of individuals and organizations supplied us with data. A special thanks to Rebecca Rhodes for access to the microdata for many EGRA learning assessments. The views expressed here are those of the authors and should not be attributed to the World Bank.

The notion of human capital – resources imbedded in people – was alluded to as early as 1776 by Adam Smith and formalized two centuries later by Becker (1962). Ever since, the literature has explored the role of human capital in economic development. For decades, this literature proxied human capital with measures of schooling. Examples include Nelson and Phelps (1966), Mincer (1984), Lucas (1988), Barro (1991), and Mankiw et al. (1992).

However, proxying human capital with schooling assumes that being in school translates into learning. Evidence suggests that this is often not the case. A recent analysis reveals that six out of ten adolescents world-wide cannot meet basic proficiency levels in math and reading (UNESCO 2017). The gap between schooling and learning is particularly acute in developing countries. In Kenya, Tanzania, and Uganda three-quarters of grade 3 students cannot read a basic sentence such as “the name of the dog is Puppy.” In rural India, half of students in grade 3 cannot solve a two-digit subtraction problem such as 46 minus 17 (World Bank 2018).

These stylized facts demonstrate a gap in human capital formation: students are in school but are not learning. Figure 1 demonstrates the contrast between rising enrollment and stagnating learning. Closing this gap presents a significant margin to drive economic development. The literature indicates that when human capital is measured by schooling it fails to deliver the returns predicted by growth models (Krueger and Lindahl 2001; Pritchett 2006; Hanushek and Woessmann 2008). However, when measured by learning, human capital is more strongly associated with growth (Hanushek and Woessmann 2012). This result is robust to improved measures of schooling such as those produced by Cohen and Soto (2007). Thus, when human capital is measured by a central output of schooling – learning – predictions from economic growth models prove more robust.

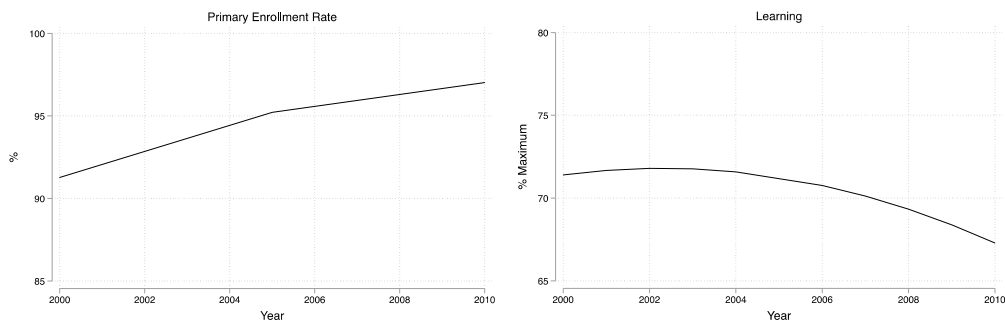


FIGURE 1. ENROLLMENT VERSUS LEARNING (2000-2010)

*Notes:* Primary enrollment rates are averaged for the cohorts 25 years and older. Both enrollment and learning are averaged across countries. Percent maximum is calculated by dividing scores by the score at the upper end of the scale (625) in line with Kraay (2018).

*Source:* Primary enrollment rates are from Lee and Lee (2016); Learning outcomes are from our database.

However, to date much of the analysis of human capital when measured by learning has focused on advanced economies. This is due to the absence of comparable measures of learning in developing countries. This excludes a significant portion of the global distribution, in particular countries with the most potential to gain from human capital accumulation.

In this paper, we bridge this gap. We introduce a database of globally comparable learning outcomes for 164 countries and territories covering 98 percent of the global population from 2000 to 2017. The database will be updated annually and made available for public use. We hope this database will enable tracking of human capital formation, a critical ingredient for development. A large-scale effort to do this is the World Bank's new Human Capital Index. This database is a key ingredient in the index. This database will enable deeper understanding of factors correlated with human capital formation and with potential causal links to economic development.

We present stylized facts in a first application of the database to demonstrate its utility. We show that learning has stagnated while schooling has increased. We further find that gender gaps in learning are large and positive with girls

outperforming boys. This points in the opposite direction of the gender gap for schooling which is negative. In an illustrative growth exercise, we find that human capital measured by learning is associated with growth. This finding reinforces recent literature on a global scale for the first time. The expanded coverage of our database makes possible heterogeneity analysis of previously underexplored dimensions, enabling new insight. We find significant heterogeneity in growth patterns by initial and current income status.

We also contribute to a development accounting literature which accounts for the role of human capital in cross-country income differences. To date the literature has relied on years of schooling data, and in the absence of global learning data, explored an array of techniques to infer school quality. This has resulted in a wide range of accounting estimates. Rather than infer, we provide a direct measure of school quality and human capital. We find that human capital accounts for a third of cross-country income differences – a middle ground in a literature that provides estimates ranging from nearly all (Jones 2014) to potentially zero (Caselli and Ciccone 2018).

We compare our measure of human capital to alternatives. We find that our measure is highly correlated with other learning measures while tripling coverage. Moreover, our measure of human capital has a strong association with growth while schooling and the measure of human capital in the Penn World Tables (Feenstra et al. 2015) are more limited. This indicates that years of schooling and the Penn World Tables measures might understate the role of human capital. However, their use remains standard practice in part since these data have the broadest coverage. By constructing learning data across 164 countries we fill a key gap: broad coverage over nearly two decades and a measure of human capital with strong links to economic development.

The rest of this paper is organized as follows. Section I outlines the methodology. Section II describes the database. Section III showcases descriptive statistics.

Section IV compares our data to various alternatives. Section V presents several stylized facts and Section VI concludes.

## **I. Methodology**

We leverage the growth of international assessments to construct globally comparable learning outcomes. These tests are derived from assessments conducted in the United States since the 1960s such as the SAT and the National Assessment of Educational Progress (NAEP). The tests are psychometrically designed, standardized assessments of cognitive skills. Since the 1990s, international assessments have been conducted by organizations such as the OECD. Two high profile examples are PISA and TIMSS which covered 71 and 65 countries in 2015. These assessments enable credible global comparison of learning across countries and over time.

To date, a series of international assessments have been included in analysis of human capital and development (Hanushek and Kimko 2000; Barro and Lee 2001; Hanushek and Woessmann 2012). These analyses focus on OECD countries and cover few developing countries. For example, Hanushek and Woessmann (2012) include 50 countries, around half of which are from the OECD. This limits the distribution of countries represented and has implications for our understanding of the link between human capital and economic development.

We include developing countries by constructing a learning ‘exchange rate’ between international assessments and their regional counterparts. Regional assessments cover much of sub-Saharan Africa and Latin America. Thus, through construction of a cross-test exchange rate between international and regional assessments we quantify the difference between them, adjust for this difference, and then place learning outcomes from regional assessments on a global scale.

Our approach builds on seminal work by Hanushek and Kimko (2000) and Barro and Lee (2001). Barro and Lee (2001) use regression to obtain different constants between various tests, allowing for test differences. Hanushek and Kimko (2000) create comparable scores over time using the United States as an anchor, since the United States participates in consistent national assessments as well as various international assessments. Altinok, Angrist and Patrinos (2018) use a similar approach across countries and over time.

#### *A. Construction of a Learning Exchange Rate*

The central intuition behind the construction of globally comparable learning outcomes is the production of a learning ‘exchange rate’ between international and regional assessments. This exchange rate can be produced for countries that participate in a given pair of assessments and captures the difference in difficulty between the two assessments. This exchange rate can then be used to place scores for countries that only participate in regional assessments on the international scale. This enables construction of globally comparable learning outcomes.

We present an illustrative example. Lesotho has data from a regional assessment in 2007. We want to place Lesotho’s regional score on an international scale. To do this, we can produce an exchange rate between the regional assessment and an international assessment. We produce this exchange rate first by comparing scores for countries that participate in both assessments in the same subject and adjacent years. In this case, Botswana participated in the same regional assessment as Lesotho as well as an international assessment in 2007. Since in Botswana the same population of students took two different tests at the same time, we can derive an exchange rate. We can then apply this exchange rate to Lesotho’s score on the regional assessment to place it on the international scale.

In this example, Botswana has a mean score of 364 in math in 2007 on an international assessment, TIMSS. Botswana also participated in a regional assessment, SACMEQ, in the same year and subject and scored 520. The exchange rate between SACMEQ and TIMSS in math in 2007 is therefore 0.7. Lesotho's regional score was 477 in math in 2007. Applying the exchange rate of 0.7 to Lesotho's regional score of 477 would produce a score of 334 on the international scale.

More formally, suppose that a population of students takes two different assessments  $X$  and  $Y$ . We assume that any differences in the score distributions on  $X$  and  $Y$  can be attributed to the assessments with constant group ability. In its most straightforward form, we convert a score from regional test  $X$  to an international test  $Y$  as follows:

$$(1) \quad y = x * e$$

where  $e$  is an exchange rate;  $y$  is a score on international test  $Y$  and  $x$  is a score on regional test  $X$ .

The success of this approach hinges on three key assumptions. First, linked tests must capture the same underlying population. This assumption is satisfied by using sample-based assessments representative at the national level where a country participated in both a regional and international assessment. This ensures that the underlying population tested is the same on average and we capture differences between tests.

Second, tests should measure similar proficiencies. To this end, we link within subjects (math, reading and science) and schooling levels (primary and secondary) to ensure overlap.

Third, the exchange rate should capture differences between tests rather than country-specific effects. This assumption is most likely to hold the larger the



number of countries which participate in a given pair of tests being linked. To ensure this last assumption holds, we fix the exchange rate over the entire interval. This increases the sample size used to produce the exchange rate, increasing the likelihood that we capture test-specific rather than country-specific differences. In fixing the exchange rate, we assume that the relationship between tests stays constant across rounds. This assumption is reasonable and is true by design since the mid 1990s when assessments started to use a standardized approach and to link testing rounds with overlapping test items. A related advantage of fixing the exchange rate is that it guarantees that any changes in test scores over this interval are due to realized progress in learning rather than changing exchange rates.

Of note, every update of the database increases the number of countries participating in a given pair of assessments. Thus, each update both expands coverage as well as enhances the reliability of all estimates by enabling construction of a more robust exchange rate.

Below we capture a level of precision needed to satisfy the above assumptions. We produce a fixed exchange  $e_{sl}$  within subjects and schooling levels (primary and secondary) from test X to test Y:

$$(2) \quad e_{sl} = \frac{1}{r} \sum_{r=1}^r \frac{\mu(y_{rsl})}{\mu(x_{rsl})} \quad \text{where}$$

$$\mu(x_{rsl}) = \frac{1}{n_{rsl}} \sum_{i \in Y_{rsl} \cap X_{rsl}} x_{irsl}$$

$$\mu(y_{rsl}) = \frac{1}{n_{rsl}} \sum_{i \in Y_{rsl} \cap X_{rsl}} y_{irsl}$$

where  $i$  is a country in the set of  $n_{rsl}$  countries that participate in both tests X and Y in a given testing round  $r$ , subject  $s$ , and schooling level  $l$  such that  $Y_{rsl} \cap X_{rsl}$ .

The upper value of  $r$  is the number of testing rounds over the period of the fixed exchange rate.

Mean scores,  $\mu(x)$  and  $\mu(y)$ , are calculated across countries for a given round  $r$ , subject  $s$  and schooling level  $l$ . We consider tests to be in the same round if they are five years apart and optimize to have the rounds as tight as possible. Most often this translates to within one to two years. In some cases, this extends to three to five years apart. In a few exceptions, we average adjacent years across one another. This minimizes the likelihood that test differences are a function of time, proficiency, schooling level, or data availability and are an accurate reflection of test difficulty.

We apply this exchange rate to a country  $j$  that participates in test X but not test Y to produce a comparable score (referred to as a Harmonized Learning Outcome (HLO) in the database):

$$(3) \quad y_{jtsl} = x_{jtsl} * e_{sl}$$

where  $t$  is the official year of test X. Means can be calculated for various disaggregated groups, for example, by gender. We can also aggregate scores across levels and subjects by averaging across subjects and then across levels, weighting each equally.

In line with this level of precision, we explore another demonstrative example. We want to know Brazil's primary math score in 2013. While Brazil does not participate in an international assessment at the primary level, it does participate in a regional assessment, LLECE. To place this score on the international scale, we produce an exchange rate for countries that participate in this regional test and an international counterpart at the primary level in math, TIMSS.

In the latest testing round, these countries were Chile and Honduras. Chile's regional math score was 581 and Honduras's score was 480. On the international assessment, they score 462 and 397. Thus, the average regional score is 530 and

the average international score is 429, and the resulting exchange rate in the latest testing round is 0.81. We next examine the prior testing round to see if the exchange rate is stable across rounds and to produce a fixed exchange rate. In this round, the countries participating in the regional and international assessment were Colombia and El Salvador. Colombia scored a 493 and El Salvador scored a 472 on the regional assessment. On the international assessment they scored 356 and 330, respectively. Thus, the exchange rate was 0.76. The two exchange rates of 0.76 and 0.81 are relatively stable across a 15-year period and variation across participating countries. Averaging across both rounds, we produce a fixed exchange rate of 0.79. Applying this exchange rate to Brazil's regional primary math score of 520 in 2013, an internationally comparable score is 411 for math in 2013.

Our methodology enables us to compare learning outcomes on a global scale. On this scale, 625 represents advanced attainment and 300 represents minimum attainment. This interpretation is derived by taking an international benchmark for high performance on the upper end of the distribution and a regional benchmark on the lower end of the distribution. This approach enables us to capture performance across the distribution and accounts for floor and ceiling effects that would be introduced by taking international or regional benchmarks on both ends.

### *B. Measure of Uncertainty*

We include measures of uncertainty to quantify the degree of confidence around our estimates. We capture two sources of uncertainty: scores on the original test and uncertainty in the calculation of exchange rate across tests. Simplifying equation (3), the score for a given country on test  $X$  is:

$$(4) \quad y = x * e_{sl}$$

We calculate the variance by bootstrapping. We assume scores at the country level are asymptotically normally distributed. We take 1,000 draws from the distribution of subject-level average test scores for each testing regime. We create exchange rates and scores from each bootstrapped sample. We take the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the distribution and use this to construct lower and upper bounds of uncertainty. We find small uncertainty intervals overall, with an average of 3.5 points and ranging from 1 to 11 points. This is consistent with original standard errors from each respective testing regime. We find larger uncertainty for our estimates relative to original scores when testing regimes have fewer countries participating in a given pair of tests. By quantifying this uncertainty, we can more reliably bound our estimates.

### *C. Limitations*

A potential limitation of this methodology is that it is sensitive to the scale of each test being similar. We address this limitation by using tests with a mean of 500 and standard deviation of 100. In rare cases where tests do not have initial scores on this scale, we rescale the test using the microdata. This ensures we capture differences in test difficulty rather than differences in scaling.

Another limitation of our approach is the ability to compare *variation* as well as *levels* across assessments. Recent work by Hanushek and Woessmann (2012) aims to address this issue. To do so, the authors express performance in terms of standard deviations. They project the standard deviation of a relatively homogenous and stable group of OECD countries and transform these standard deviations into scores using a standardized scale for one of the well-known international assessments. However, as the authors acknowledge, this approach does not apply for countries far from the OECD distribution of scores. This is a particularly noteworthy limitation for analyses focused on developing countries. While our approach does

not account for variation, itself a limitation, it is also not biased by them. This approach is consistent with earlier approaches by Hanushek and Kimko (2000) and Barro and Lee (2001).

Another potential concern regards data availability. While this is the largest learning outcomes database to date, data is still sparse for some countries. This introduces bias if data availability is correlated with education quality or progress. For example, if countries that perform worse have data only in later years (because they were later to introduce assessments), their average score will be likely biased upwards, as the test scores will reflect more recent testing, not stronger performance. Since we provide year-by-year scores this can be accounted for.

Relatedly, when averaging data across subjects, levels and over time, there is a possibility that averages reflect the availability of data rather than true learning gains. For example, say a country has a score of 500 in 2000 in math and then jumps up to 550 in 2005. If this country added reading in 2005 and scored 450, the average score across subjects in 2005 would be 500, reflecting no learning progress since average scores would be 500 in both years. However, an apples-to-apples comparison in math shows significant learning gains from 500 to 550. To address this issue, we construct disaggregated measures of learning by subject and schooling levels as well as aggregated ones. This enables analyses at each level considering the trade-offs.

## **II. Data**

This section describes the achievement tests we use to construct our database. The first set of assessments consists of international standardized achievement tests; the second are regional standardized achievement tests; and the third is the Early Grade Reading Assessment (EGRA). Each test covers between 10 to 72 countries. By combining these assessments and making them comparable we

include 164 countries and territories covering 98 percent of the global population. Appendix A includes a detailed description of all data included in the database.

The database was produced through a combination of methodological innovation as well as a large-scale effort by the World Bank to identify, collect and collate student assessment data worldwide across eight testing regimes. We include data from the Early Grade Reading Assessment for the first time. This adds 48 countries to the database with at least one data point in the past 10 years, including large developing economies such as Bangladesh, Nigeria and Pakistan.

The final database includes mean scores for all 164 countries and territories from 2000 to 2017. We do not extend the time series prior to 2000 since data quality is low and since this does not significantly affect country coverage, with an addition of just two territories: Zanzibar and Puerto Rico. The database will be publicly available on World Bank EdStats. The codebook is summarized in Table 1.

Scores are disaggregated by schooling level (primary and secondary), subject (reading, math and science) and gender (male and female). We include standard errors for each measure to quantify uncertainty. We include year-by-year data, an indicator for whether the data is nationally representative, as well as the source test from which the scores were derived. Finally, we include the exchange rates used to convert an original score to a globally comparable score. The database will be made publicly available by the World Bank and will be updated annually.

Of note: while this is the largest learning database to date, data is still sparse. This introduces a few potential biases when aggregating data to maximize country coverage as described in the limitations section of the methodology. The inclusion of year-by-year and disaggregated data provides transparency and granularity to enable informed analysis.

TABLE 1 – DATABASE CODEBOOK

Variable	Description
<i>country</i>	Country
<i>cntabb</i>	World Bank country code
<i>test</i>	Original assessment
<i>year</i>	Year in which the original assessment was taken
<i>n_res</i>	Dummy variable indicating if EGRA data is nationally representative
<i>subject</i>	Subject (math, reading or science)
<i>level</i>	Schooling level (primary, secondary)
<i>score</i>	Score from source assessment
<i>se</i>	Standard error from source assessment
<i>score_m</i>	Score from source assessment for males
<i>se_m</i>	Standard error from source assessment for males
<i>score_f</i>	Score from source assessment for females
<i>se_f</i>	Standard error from source assessment for females
<i>HLO</i>	Harmonized Learning Outcome (HLO)
<i>HLO_se</i>	Harmonized Learning Outcome (HLO) standard error
<i>HLO_m</i>	Harmonized Learning Outcome (HLO) for males
<i>HLO_m_se</i>	Harmonized Learning Outcome (HLO) standard error for males
<i>HLO_f</i>	Harmonized Learning Outcome (HLO) for females
<i>HLO_f_se</i>	Harmonized Learning Outcome (HLO) standard error for females
<i>e_index</i>	Exchange rate used to convert original scores to HLO scores

### III. Descriptive Statistics

Several statistics demonstrate the coverage and detail of the database. Table 2 presents coverage for country-year observations by region. The database includes 2083 observations across all countries from 2000-2017. Disaggregation by gender is available for nearly all the data with 2020 country-year observations. Most data come from math scores with 744 country-year observations, followed by reading scores with 692 and lastly by science scores with 647. Almost two-thirds of observations are secondary school scores. Latin America and the Caribbean and sub-Saharan Africa make up nearly a quarter of all available data. This provides the largest representation of developing countries to date in a learning database – the countries that have the most potential to gain from human capital accumulation.

TABLE 2 – COUNTRY-YEAR OBSERVATIONS BY DISAGGREGATION AND REGION

Region	Total	Male	Female	Reading	Math	Science	Primary	Secondary
East Asia & Pacific	336	332	332	102	118	116	103	233
Europe & Central Asia	954	948	948	301	328	325	261	693
Latin America & Caribbean	247	247	247	88	83	76	111	136
Middle East & North Africa	262	256	256	62	102	98	95	167
North America	56	56	56	18	19	19	17	39
South Asia	16	9	9	11	4	1	13	3
Sub-Saharan Africa	212	172	172	110	90	12	189	23
Total	2083	2020	2020	692	744	647	789	1294

Table 3 presents country-subject-level observations by year. The data is spread over time, slightly weighted towards recent years since countries are increasingly participating in assessments. A related feature of the data is a large influx of data in particular testing years. This is more prevalent for developing regions which participate in sporadic assessment.

TABLE 3 – COUNTRY-SUBJECT-LEVEL OBSERVATIONS BY YEAR

Total	Total	Male	Female	Reading	Math	Science	Primary	Secondary
2000	192	157	157	74	75	43	63	129
2001	34	34	34	33	1	0	34	0
2002	4	4	4	2	2	0	4	0
2003	218	218	218	40	89	89	44	174
2004	2	2	2	1	1	0	2	0
2006	277	277	277	123	88	66	107	170
2007	193	193	193	15	96	82	97	96
2008	3	3	3	3	0	0	3	0
2009	225	225	225	79	73	73	6	219
2010	6	5	5	6	0	0	6	0
2011	239	239	239	55	92	92	151	88
2012	202	201	201	74	64	64	10	192
2013	77	54	54	35	27	15	77	0
2014	28	27	27	18	10	0	28	0
2015	324	324	324	75	126	123	98	226
2016	55	54	54	55	0	0	55	0
2017	4	3	3	4	0	0	4	0
Total	2083	2020	2020	692	744	647	789	1294



Figure 2 below presents our measure of human capital for 164 countries and territories from 2000-2017. Figure 2 makes the global coverage of the database immediately apparent with typically excluded regions from international tests such as PISA and TIMSS included in our database. This includes the vast majority of sub-Saharan Africa, Latin America and the Caribbean, and South Asia – economies with significant potential to harness human capital for economic development. A few clear trends emerge when comparing our measure of human capital across regions. Advanced economies far outpace developing economies in terms of human capital; sub-Saharan Africa lags behind all regions; within sub-Saharan Africa, a few countries such as Kenya, Tanzania, and Cameroon lead, on par with many countries in Latin America; within Latin America, a few countries such as Chile are on par with European counterparts; the Middle East performs similarly or worse than Latin America; many Asian countries outperform North American and European counterparts, while a few South Asian countries such as India and Bangladesh perform on par with many sub-Saharan African countries.

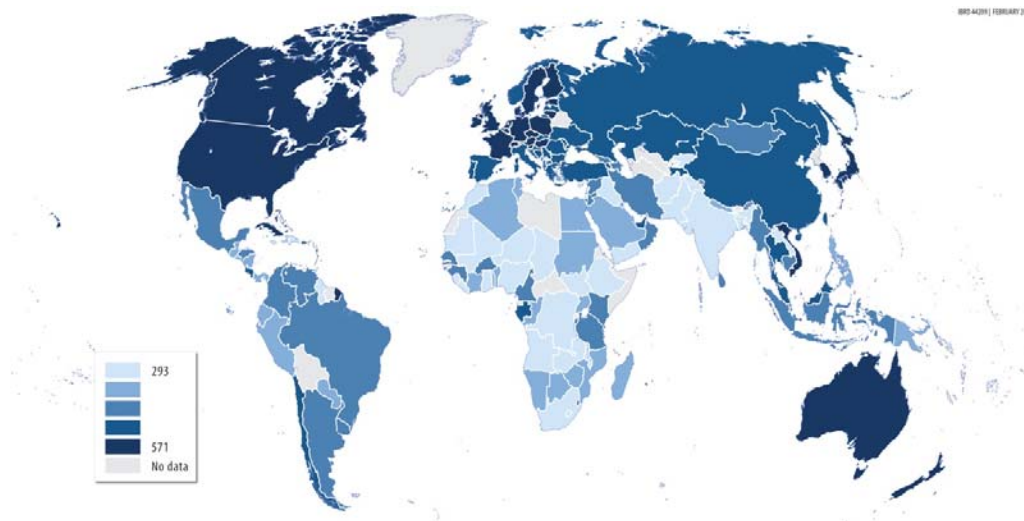


FIGURE 2: AVERAGE HUMAN CAPITAL SCORE (2000-2017)

*Notes:* Average human capital is calculated across subjects and levels over the given period of time. Human capital scores vary on a range of 300 to 625, where 300 is minimum attainment and 625 is advanced attainment.

Figure 3 summarizes human capital for selected countries for the last decade. We observe a few case studies consistent with the trends described above. Russia outperforms the United States – a surprising finding. Chile outperforms Eastern European countries such as Georgia. Saudi Arabia places near the bottom outperforming only African countries. The gap between Morocco and Singapore is nearly double. Singapore and Finland have low variation due to a potential plateau on the upper end of performance. Rwanda has low variation due to limited data. Russia has high variation due to improving human capital, whereas South Africa has high variation due to declining human capital.

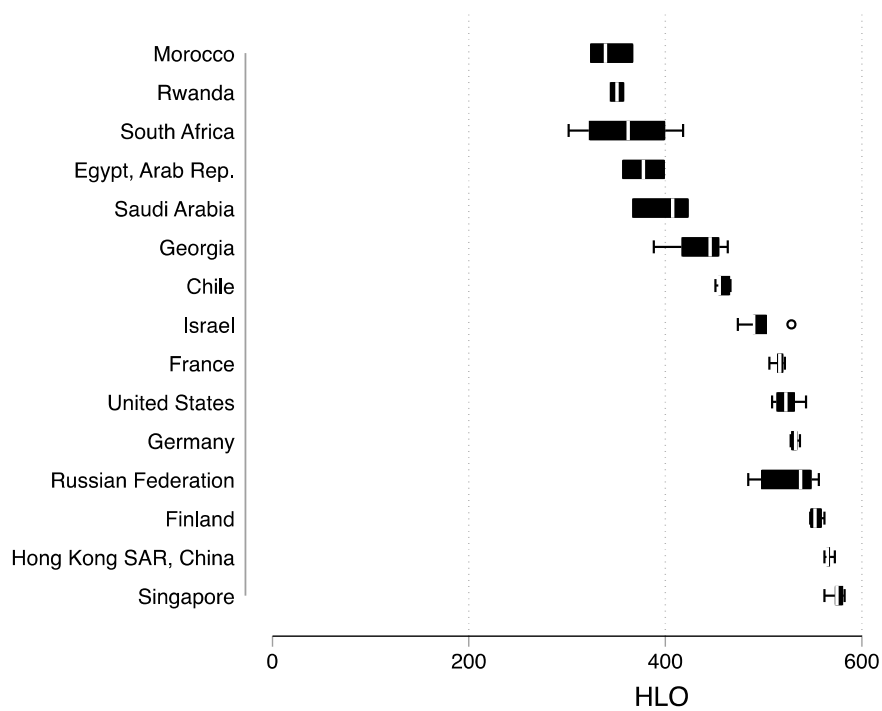


FIGURE 3. AVERAGE HUMAN CAPITAL (2007-2017) - SELECTED COUNTRIES

Notes: Average human capital is calculated across subjects and schooling levels over the given period of time.

Source: Our learning outcomes database.

Figure 4 plots human capital for each country by the log of their GDP per capita. This graph illuminates cases where countries have managed to improve human capital despite a lack of resources, as well as cases where countries have resources to invest in to date unrealized human capital accumulation. Former or current centrally planned economies display better human capital than their income would suggest, such as Singapore, Poland, Bulgaria, Cuba and Vietnam. Countries in the Middle East and Africa reach lower human capital levels than predicted by income, such as Qatar, Kuwait, United Arab Emirates, South Africa, Nigeria and Ghana. We also highlight large developing countries: India, China, Mexico, and Brazil. China outperforms its counterparts, Mexico and Brazil perform slightly below where their income would predict, and India and South Africa trail far behind.

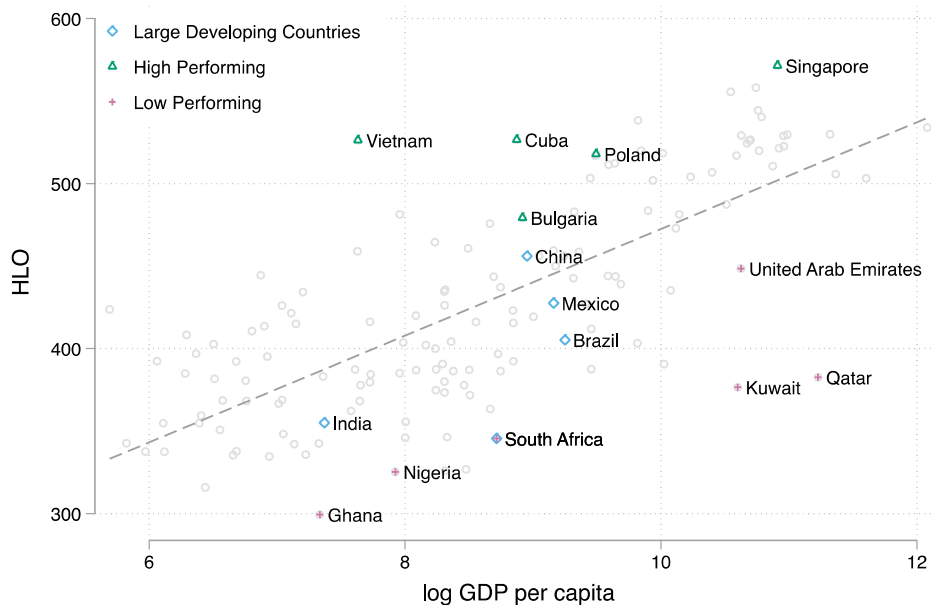


FIGURE 4. HUMAN CAPITAL (2000-2017) VERSUS 2015 GDP PER CAPITA

Notes: Average human capital is calculated across subjects and schooling levels over the given time period.

Source: GDP per capita estimates are from World Bank national accounts data; learning outcomes are from our database.

## IV. Comparison to Alternative Measures

We compare our measure of human capital to a few alternatives. First, we compare our measure to related learning measures of human capital. We find high levels of consistency with other learning measures. Next, we compare our measure to schooling measures of human capital. While coverage is similar, learning is consistently associated with growth, whereas the association with schooling is more limited. Finally, we compare our measure of human capital to a measure included in the recent version of the Penn World Tables. We find that our measure is consistently associated with growth, whereas the Penn World Tables measure, like schooling, is limited. In summary, the learning measure of human capital in this database is highly correlated with other learning measures, while tripling coverage to include 164 countries on a global scale for the first time. Moreover, our measure of human capital has a strong association with growth in contrast with schooling and Penn World Tables measures.

### *A. Comparison to Learning Measures of Human Capital*

Our database complements alternative learning measures of human capital. We correlate our measures with a few notable alternatives. Figure 5 below shows direct comparisons to learning data used in growth regressions by Hanushek and Kimko (2000) and Hanushek and Woessmann (2012). We find correlations of 0.794 and 0.925, indicating high consistency.

In Table 4, we also compare our data to learning outcomes produced using Item Response Theory (IRT) – the technique used by psychometricians to generate scores for each respective international and regional assessment. IRT models the probability a given pupil answers a given test item correctly as a function of pupil and item-specific characteristics (Mislevy et al. 1992; Holland and Dorans 2006). While this methodology is used to construct the underlying tests we use, to use it

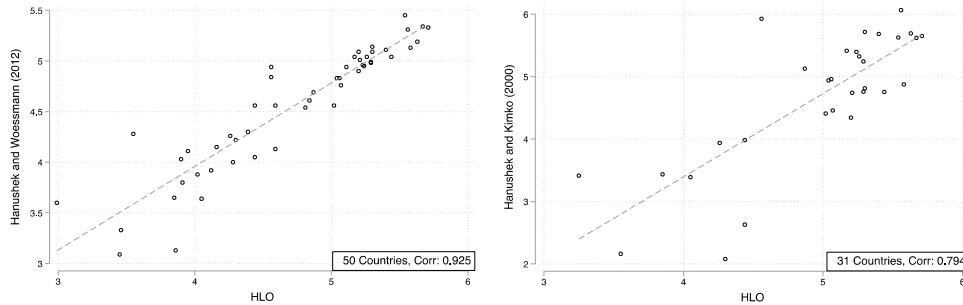


FIGURE 5. COMPARISON TO HANUSHEK AND KIMKO (2000) AND HANUSHEK AND WOESSMANN (2012)

*Notes:* Hanushek and Kimko (2000) provide estimates of comparable learning measures of human capital. Hanushek and Woessmann (2012) improve and expand these estimates. We compare our HLO measure to both estimates for the set of countries included in their growth regressions.

to compare assessments would require enough overlap in the test items. This is not true for a significant enough set of tests and time intervals to create a globally comparable panel dataset.

When this overlap is small, standard maximum likelihood estimates will reflect both true variance and measurement error, overstating the variance in the test score distribution. Das and Zajonc (2010) elaborate on the various challenges of estimating IRT parameters with limited item-specific overlap. Sandefur (2018) shows that IRT produces unreliable results in a context with limited item overlap between an international and regional assessment in East and Southern Africa.

While IRT might not be a reliable approach when there is limited item-by-item overlap, we conduct a few comparisons where overlap is larger. We compare our results to the *Linking International Comparative Student Assessment (LINCS)* project which uses IRT methods and has significant overlap in items for a subset of international studies focused on reading at primary school (Steinmann et al. 2014). We find that our database can produce similar results to IRT methods where there is overlap with a correlation coefficient above 0.92.

TABLE 4 – RELATIONSHIP WITH ALTERNATIVE LEARNING HUMAN CAPITAL MEASURES

	Pearson Coefficient	p-value	Observations
Hanushek and Woessmann (2012)	0.925	<.001	50
Hanushek and Kimko (2000)	0.794	<.001	31
Item Response Theory	0.922	<.001	34

*Notes:* Comparisons are made using average scores across subjects and schooling levels.

*Source:* Hanushek and Woessmann (2012) and Hanushek and Kimko (2000) data are from the respective papers. Item Response Theory data is from the LINC project (Steinmann et al. 2014). We compare these data to the learning outcomes in our database.

Table 4 summarizes the results. In all cases the correlation is roughly .8 or higher. These comparisons indicate that even as we expand coverage from 50 to 164 countries and territories we maintain high levels of consistency with alternative measures where there is overlap.

#### *B. Comparison to Schooling and Penn World Tables Human Capital Measures*

We compare our measure of human capital with a schooling measure of human capital from Barro-Lee (2013) as well as a new measure of human capital in the Penn World Tables. The measure of human capital in the Penn World Tables is a combination of years of schooling from Cohen-Leker (2014) and Barro-Lee (2013) as well as returns to schooling estimates from Psacharopoulos (1994). We include these measures of human capital in growth regressions in Table 5.

Each measure is significant on its own in columns (1)-(3). However, column (4)-(6) show that when we include the measures side-by-side pairwise and all together, the only measure that is significant is learning. Neither schooling from the Barro-Lee dataset nor the Penn World Tables is significantly associated with growth. Column (7)-(8) show this result is robust to multiple specifications, although the effect is slightly reduced.

TABLE 5 – HUMAN CAPITAL AND GROWTH – COMPARING MEASURES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Human Capital - Learning		1.61 (0.487)		1.256 (0.603)		1.233 (0.616)	1.223 (0.614)	1.099 (0.662)
Human Capital - Penn World Tables			1.292 (0.466)		1.012 (1.739)	0.361 (1.747)	-1.233 (2.010)	0.603 (1.860)
Human Capital - Schooling	0.272 (0.100)			0.122 (0.122)	0.0626 (0.374)	0.0502 (0.369)	0.400 (0.424)	0.205 (0.388)
Initial GDP per capita	-0.112 (0.0193)	-0.127 (0.0207)	-0.115 (0.0197)	-0.131 (0.0210)	-0.115 (0.0198)	-0.131 (0.0213)	-0.13 (0.0212)	-1.836 (0.384)
Observations	117	117	117	117	117	117	117	117
R-squared	0.232	0.254	0.234	0.260	0.234	0.261	0.266	0.176

*Notes:* Dependent variable: annual growth rates averaged across 2000-2010. Human Capital - Schooling refers to estimates in 2000, the beginning of the time period. Initial GDP per capita refers to levels at the beginning of the period in the year 2000. Human Capital – Penn World Tables refers to the measure of human capital in the Penn World Tables. Human Capital – Learning refers to the measure of human capital in this database from 2000 onwards. Results exclude countries in civil war, inflation crises and with rents from natural resources above 25 percent. Column (7) includes a measure of Human Capital – Schooling as the average from 2000-2010. Column (8) includes a measure of the log of initial GDP per capita.

*Source:* Schooling data are from Barro-Lee (2013). GDP and human capital data are from PWT 9.0. Learning estimates are from our database.

This illustrative analysis indicates that the measure of human capital introduced in this database is associated with growth and development while schooling measures such as Barro-Lee (2013) are more limited. Thus, while schooling databases have global coverage, the association with growth is most robust when human capital is measured by learning. This finding is consistent with the recent literature across a smaller sample of 50 countries (Hanushek and Woessmann 2012). We demonstrate that this stylized fact holds on a global scale across over 100 countries. This does not mean that schooling is not useful, but that it leads to growth largely through the channel of learning.

This analysis further reveals that the human capital variable in the Penn World Tables might miss important dimensions and underestimates the role of human capital in economic development.

## V. Stylized Facts

We present a series of stylized facts relevant to human capital and economic development. First, we observe that as schooling has increased over the last decade, learning has stagnated or declined. Second, we observe a positive gender gap for learning in contrast to the negative gender gap observed for schooling. Third, we find human capital measured by learning is consistently associated with growth on a global scale, whereas schooling is not. Fourth, we find significant heterogeneity in growth patterns by income status. The wide coverage of our database enables meaningful heterogeneity analysis for previously underexplored levels of granularity. Fifth, we conduct an illustrative development accounting exercise and find that human capital accounts for up to a third of cross-country income differences, roughly in the middle of the recent literature.

### *A. Schooling Is Not Learning*

Figure 6 explores the contrast between changes in schooling and learning over time. We measure schooling using primary school enrollment rates. We compare this to the learning measure of human capital in our database for the years 2000-2010. We use data from 2000-2010 since this is the period with the most overlap of schooling and learning measures.

We see a clear trend towards increased schooling, while learning has stagnated and, in some instances, declined. Part of this decline, for example, in East Asia and Pacific is reflective of low-income countries being included in our database for the first time. In other cases, stagnation reflects a stall in learning. In both scenarios the data highlight an emerging human capital gap: students are increasingly in school but are not learning. These stylized facts reinforce recent emphasis by the World Bank and UNESCO of a ‘learning crisis’ (UNESCO 2017; World Bank 2018).



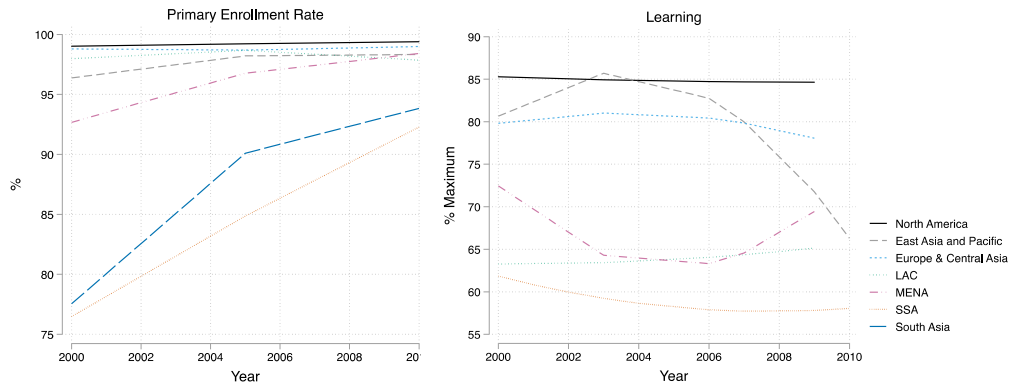


FIGURE 6. ENROLLMENT VERSUS LEARNING (2000-2010), BY REGION

*Notes:* Primary enrollment and learning estimates are averaged within regions. LAC refers to Latin American and the Caribbean; MENA refers to the Middle East and North Africa; and SSA refers to sub-Saharan Africa. Percent maximum learning is calculated by dividing scores by the maximum score at the upper end of the score scale (625) in line with the methodology used in the Human Capital Index (Kraay 2018).

*Source:* Primary enrollment rates are from Lee and Lee (2016). Learning estimates are taken from our database.

### *B. The Gender Gap Is Flipped*

Figure 7 disaggregates learning and enrollment data by gender. We see gender gaps for schooling and learning across most regions. Gender gaps are both larger and have more variance across years and regions for learning than for schooling.

We further observe that the gender gaps move in opposite directions. In terms of enrollment, girls lag while on learning girls outperform boys. This finding is consistent with the literature (Muralidharan and Prakash 2017). This flip in the gender gap might be due to selection. In regions where enrollment is low, only high-achievers might be taking assessments. However, in sub-Saharan Africa, where enrollment is second lowest, the gender gap is null to negative. Moreover, across Europe and Central Asia there is little gender gap in enrollment yet meaningful gender gaps in learning. Together, these stylized facts indicate selection is unlikely the only driver.

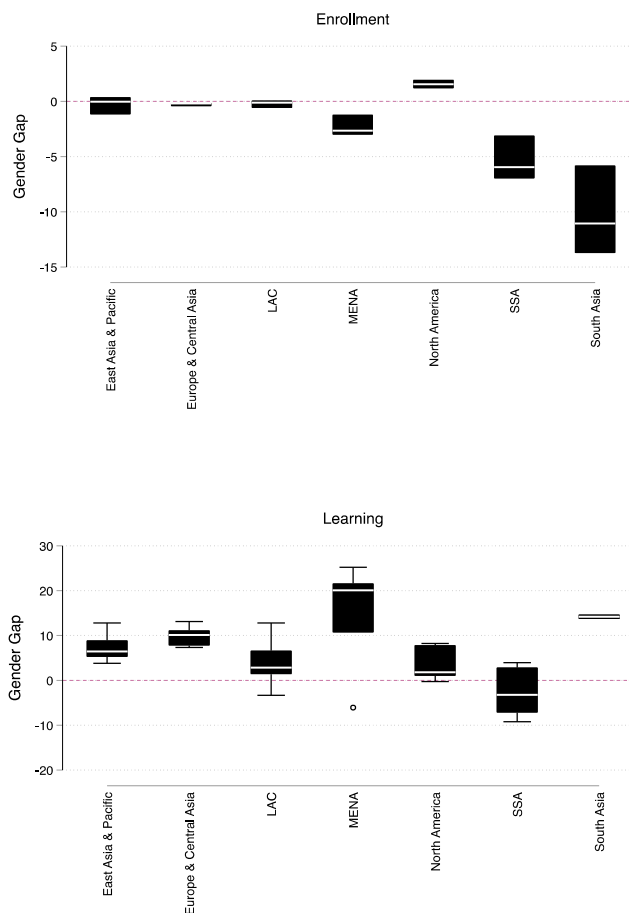


FIGURE 7. GENDER GAP – ENROLLMENT VERSUS LEARNING (2000-2010), BY REGION

*Notes:* LAC refers to Latin American and the Caribbean; MENA refers to the Middle East and North Africa; and SSA refers to sub-Saharan Africa. The gender gap is calculated by taking the difference of female and male enrollment or learning. A positive gender gap indicates females do better and vice-versa.

*Source:* Primary enrollment rates are from Lee and Lee (2016). Learning estimates are taken from our database.

### *C. Human Capital when Measured by Learning is Associated with Growth*

We examine the relationship between human capital and economic growth. We include a global distribution of countries in an analysis where human capital is measured by learning for the first time. A prior analysis by Hanushek and Woessmann (2012) included 50 countries, about half of which were from the OECD. We analyze more than double this number of countries and include the

largest number of developing countries to date – the countries with the most potential to gain from human capital accumulation.

We focus on the 2000-2010 interval which has the largest overlap in data on human capital as measured by both schooling and learning. This ensures that when making comparisons between these two types of measures of human capital we capture differences in the explanatory power of the human capital variable rather than time effects.

The full sample of countries for which we have growth, years of schooling and learning data is 132. We exclude two countries, Syria and Yemen, due to recent civil wars as well as Zimbabwe which had a historic inflation crisis during this time period. We further exclude 12 countries for which more than 25 percent of their GDP is comprised of rents from natural resources. This cutoff is arbitrary, and the findings are robust to various cutoffs, including 15 and 20 percent. These economies have highly volatile growth rates and are driven by factors orthogonal to human capital, for example, shocks to oil and mineral prices. This is consistent with the approach taken in standard cross-country growth regressions (Mankiw et al. 1992; Hanushek and Woessmann 2012) and is applied systematically using data from the World Bank on rents from natural resources as a share of GDP (Lange et al. 2018). It is also typical to exclude former communist countries. We do not show results with this exclusion. When this exclusion is made the associations are slightly stronger. The final sample includes 117 countries. Table 6 presents the results and Figure 8 presents an associated added variable plot.

We find that learning is highly associated with growth, even when we control for initial GDP and years of schooling. The association is large and significant varying between the 1 percent and 10 percent level across nearly all specifications. This result is consistent with the recent literature (Hanushek and Woessmann 2012).

TABLE 6 – HUMAN CAPITAL AND GROWTH FOR A GLOBAL DISTRIBUTION OF COUNTRIES

	<i>Annual Growth Rates</i>				<i>GDP Per Capita</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Human Capital - Learning	1.610 (0.487)	1.256 (0.603)	1.164 (0.605)	1.131 (0.652)	3.729 (0.525)	3.547 (0.652)	3.472 (0.655)	4.871 (1.715)
Years of Schooling		Yes	Yes	Yes		Yes	Yes	Yes
Initial GDP Per Capita	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	117	117	117	117	117	117	117	117
R-squared	0.254	0.260	0.264	0.176	0.970	0.970	0.970	0.804

*Notes:* Dependent variable: annual growth rates and GDP per capita are averaged across 2000-2010. Years of Schooling refers to estimates in 2000, the beginning of the time period. Initial GDP per capita refers to levels at the beginning of the period in the year 2000. These results exclude countries in civil war, inflation crises, and with rents from natural resources above 25 percent. Human capital measured by learning refers to average estimates from our database from 2000 onwards. Column (3) and (7) refers to Years of Schooling as the average over the 2000-2010 time period. Columns (4) and (8) use the log of initial GDP per capita.

*Source:* Years of schooling data are from Barro-Lee (2013). GDP data are from PWT 9.0. Learning estimates are from our database.

This similarity in findings is noteworthy on two fronts. First, with regards to the time period. We find broadly consistent results with prior analyses from 1960 to 2000. Second, with regards to the distribution of countries. We include 117 countries, including a significant number of developing countries. This indicates that prior associations between learning and growth are unlikely idiosyncratic - they hold even when a global distribution of countries is considered.

These findings are robust to multiple specifications. We include the average of years of schooling across the time period. We find that results persist when controlling for the log of initial GDP at the 10 percent level. This allows for conditional convergence as well as variation in competing growth theories with advocates of the neoclassical model preferring this specification (Lucas 1988; Romer 1990; Mankiw et al. 1992). We find robust findings regardless of the model specified.

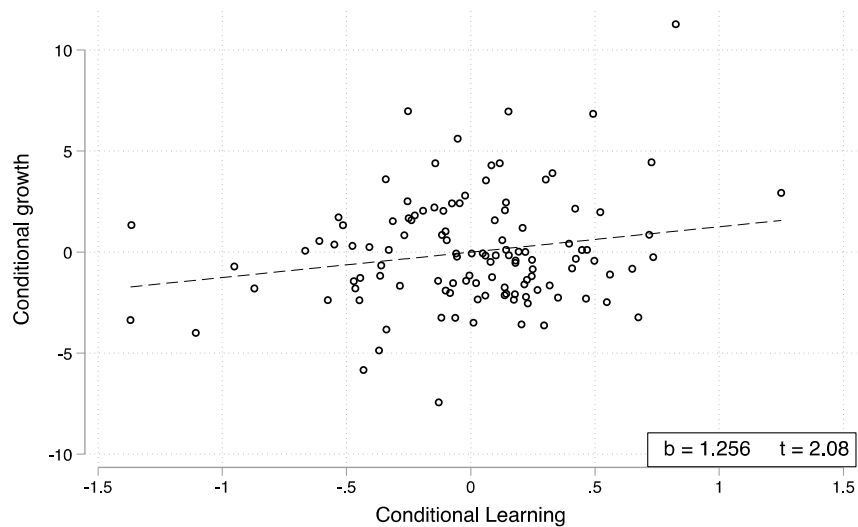


FIGURE 8. LEARNING AND GROWTH (2000-2010)

*Notes:* 117 countries and territories are included in the added-variable plot above. This added variable plot is derived from a regression of the average annual rate of growth of real GDP per capita from 2000-2010 on learning outcomes from 2000 onwards conditional on years of schooling in 2000 and initial GDP in 2000.

*Source:* Years of schooling data are from Barro-Lee (2013). GDP data are from PWT 9.0. Learning data is from our database.

#### *D. Heterogeneity in Growth*

A significant contribution of our database is the ability to conduct heterogeneity analysis since we include learning data for 164 countries and territories. Whereas Hanushek and Woessmann (2012) have one country in the low-income sub-group, we are now able to include 20. Our database enables growth analysis for sub-Saharan Africa, Europe and Central Asia, and South Asia, as well as substantially improves coverage in Latin America and the Caribbean and the Middle East and North Africa. To this end, our database enables us to uncover precise trends within sub-groups.

Table 7 presents results by income status as classified by the World Bank as well as by initial quartile of income at the start of the growth period in 2000. These results are robust to multiple specifications similar to Table 6. We present only the main specification in the table.

TABLE 7 – HUMAN CAPITAL AND GROWTH BY CURRENT AND INITIAL INCOME STATUS

	<i>Current Income Status</i>				<i>Initial Income Quartile</i>			
	<i>Low Income</i>	<i>Lower Middle Income</i>	<i>Upper Middle Income</i>	<i>High Income</i>	<i>Quartile 1</i>	<i>Quartile 2</i>	<i>Quartile 3</i>	<i>Quartile 4</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Human Capital - Learning	0.709 (1.592)	1.819 (1.435)	3.892 (0.853)	1.270 (0.903)	2.001 (1.323)	4.002 (1.487)	0.558 (0.908)	3.135 (1.464)
Years of Schooling	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial GDP per Capita	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17	25	29	46	30	29	29	29
R-squared	0.608	0.154	0.635	0.418	0.151	0.373	0.389	0.467

*Notes:* Dependent variable: annual growth rates are averaged across 2000-2010. Years of Schooling refers to estimates in 2000. Initial GDP per capita refers to levels at the beginning of the period in the year 2000. These results exclude countries in civil war, inflation crises, and with rents from natural resources above 25 percent. Human capital measured by learning refers to average estimates from our database from 2000 onwards.

*Source:* Years of schooling data are from Barro-Lee (2013). GDP data are from PWT 9.0. Learning data is from our database.

We find a positive correlation between learning and growth in all cases. However, results are only significant for current upper middle-income countries. For upper middle-income countries, the association is over triple the average and significant at the 1 percent level. The R-squared is also triple the average at 0.635. When exploring results by initial income levels at the start of the period, we find the largest and most significant associations are for the second quartile with more than triple the average association between human capital and growth.

These results together indicate a potential human capital ‘sweet spot’ where investment in human capital is most associated with growth as countries move from the bottom to upper end of middle-income status. Given weak effects at the lowest end of the distribution, this has meaningful implications for development. This stylized fact indicates human capital is potentially a necessary but not sufficient condition for growth. It must be accompanied, for example, by institutions and labor markets that value skills and can harness human capital into productive assets.

By expanding coverage to 164 countries our database enables heterogeneity analysis at previously underexplored level of disaggregation. Such analyses have the potential to shed new insights into the relationship between human capital and economic development.

### *E. Development Accounting*

The development accounting literature studies the relative contribution of human capital in cross-country income differences. However, this question remains unsettled, largely due to difficulties in measuring human capital. While direct measures of years of schooling exist, the quality of schooling has been inferred.

Several approaches have emerged to estimate the quality of schooling, including cross-country differences in Mincerian returns (Hall and Jones 1999; Caselli 2005), immigrant returns (Schoellman 2011), and cross-country skill premia (Caselli and Coleman 2006). However, these approaches are fraught with challenges such as the substitutability between skilled and unskilled workers (Jones 2014; Caselli and Ciccone 2018; Jones 2018). As a result of the measurement problem, estimates of the role of human capital in accounting for cross-country income differences vary wildly, from one fifth to four fifths (Hendricks and Schoellman 2017).

In this paper, we provide a direct and reliable measure of the quality of schooling – learning outcomes – on a global scale. In a motivating application, we construct human capital stocks by combining years of schooling with the quality of schooling using our direct measure of learning outcomes to produce learning-adjusted years of schooling.

A year of schooling produces different amounts of learning across countries. For example, in South Africa students learn half of what students in Singapore learn despite being in school for a similar number of years. We adjust the years of schooling by the amount of learning in a given country relative to a high-

performance benchmark. In the case of South Africa, which has just over 8 years of average schooling from 2000-2010 (Barro-Lee 2013), the learning-adjusted years of schooling estimate is 4.59. In contrast, Singapore, a top performer, has 9.35 years of average schooling and 8.55 learning-adjusted years of schooling. This is the approach used in the World Bank’s Human Capital Index, following Filmer et al. (2018) and Kraay (2018).

We apply these estimates in a development accounting framework. Table 8 shows our results in comparison to the literature. The first line is the ratio of human capital in the 90th to 10th percentiles. For Hall and C. Jones (1999) and Hendricks (2002) the number is around 2. For Schoellman (2011), who also constructs a measure of quality-adjusted years of schooling, it is much higher at 4.7. We find a similar, and slightly higher estimate to Schoellman (2011), of 5.3. This indicates that learning adjustments introduce significant cross-country variation in human capital. We provide two plausible estimates in our baseline accounting results. The second line compares the human capital ratio of the 90th and 10th percentiles to the output per worker ratio of the 90th and 10th percentiles. The third line compares the variance of log human capital per worker to the variance of log output per worker.

TABLE 8 – BASELINE ACCOUNTING RESULTS AND COMPARISON TO LITERATURE

	This Paper	Literature					
	Learning-adjusted Years of Schooling	Hall and C. Jones (1999)	Hendricks (2002)	Schoellman (2011)	B. Jones (2014)	Hendricks and Shoellman (2017)	Caselli and Ciccone (2018)
$h_{90}/h_{10}$	5.3	2	2.1	4.7			
$\frac{h_{90}/h_{10}}{y_{90}/y_{10}}$	0.25	0.09	0.22	0.21	<i>Nearly All</i>	0.62	<i>Potentially None</i>
$\frac{\text{var}[\log(h)]}{\text{var}[\log(y)]}$	0.33	0.06	0.07	0.26			

*Notes:* Dependent variable: real output per worker from 2000-2010. Schooling and learning estimates are also averaged from 2000-2010.

*Source:* Schooling data are from Barro-Lee (2013). GDP data are from PWT 9.0. Learning estimates are from our database. Literature estimates are derived from the referenced papers.



We find that our measure of human capital accounts for 25–33 percent of output per worker differences. This estimate sits roughly in the middle of the literature: it is larger than most estimates relying on years of schooling, and in between estimates using a variety of schooling quality measures.

This development accounting exercise is designed to be suggestive. Our central contribution is in providing a direct measure of schooling quality. This provides an analogy to direct years of schooling estimates, and introduces the possibility of better school quality estimation.

## **VI. Conclusions**

To understand and track human capital formation, a critical ingredient for development, there is need for globally comparable measurement of learning. The growth of international standardized achievement tests is a significant step in this direction. However, many of the countries that participate in these tests are often already rich. This limits the ability to track, compare or understand education patterns in developing countries – the countries that often have the most potential to gain from human capital formation.

We build on the literature on comparable learning outcomes to construct a globally comparable database of 164 countries and territories from 2000-2017, representing more than 98 percent of the global population. Approximately two-thirds of these countries are developing economies.

We contribute to the literature in several ways. This is the most current and expansive cross-country dataset on learning, including the most developing countries. We hope this dataset can be used to reveal important descriptive trends in human capital formation across developed and developing countries. Descriptive trends suggest learning is stagnating and that there is a female learning premium.

We also hope this database enables analysis of factors correlated with, and that have plausible causal links to, the formation of human capital and economic growth. In a first application on a global scale, we find that human capital is consistently associated with growth, reinforcing the recent literature. We also find significant heterogeneity, indicating scope for generation of new insights about the relationship between human capital and development. We further find that this database provides a measure of human capital that is more explanatory than schooling measures and the current measure included in the Penn World Tables 9.0.

We also contribute to a development accounting literature. To date the literature has explored an array of techniques to infer school quality in the absence of globally comparable data. This has resulted in a wide range of estimates of the role human capital plays in cross-country income differences, from nearly all to zero. We provide a direct measure of school quality to build human capital stocks. We find that human capital accounts for a third of cross-country income differences – a middle ground in the literature. We hope our measure of human capital can further inform this debate.

This database comes at a moment when a series of global efforts have been launched to measure and track learning on a global scale. A notable effort to do this is the World Bank's Human Capital Index which compares countries' levels of human capital around the world (Kim 2018a; Kraay 2018; World Bank 2019). The Human Capital Index includes learning outcomes from this database as one of its core ingredients in addition to health (Kim 2018b). The database introduced in this paper will be updated annually, will be made publicly available, and feature in future large-scale efforts to understand and track human capital and economic development.

Future iterations of this database will expand coverage across countries and time as countries join existing assessments and by including additional assessments such as early grade reading and mathematics assessments. Moreover, we aim to build a

dataset that enables over-time isolation of value-added learning by including variables which can account for various selection effects. This includes adding measures of enrollment and retention across schooling levels. We also aim to enable further identification of the link between human capital and economic growth by including variables such as comparable estimates on the returns to education. This dataset and future iterations will enable a deeper understanding of mechanisms driving human capital formation and the link to economic development.

## APPENDIX A - DATA DESCRIPTION

### *A. International Standardized Achievement Tests (ISATs)*

In the mid-1990s, there was an emergence of standardized, psychometrically-robust and relatively consistent ISATs. Below we describe the major ISATs we use in this database.

*TIMSS*. The Trends in International Mathematics and Science Study (TIMSS) is conducted by the IEA. Five TIMSS rounds have been held to date in Math and Science subjects covering grades 4 and 8. The first, conducted in 1995, covered 45 national educational systems and three groups of students.<sup>2</sup> The second round covered 38 educational systems in 1999, examining pupils from secondary education (grade 8). The third round covered 50 educational systems in 2003, focusing on both primary and secondary education (grades 4 and 8). In 2007, the fourth survey covered grades 4 and 8 and more than 66 educational systems. In 2011, the survey covered 77 educational systems across grades 4 and 8. The last round was performed in 2015 and covered 63 countries/areas. The precise content of the questionnaires varies but remains systematic across countries.

*PIRLS*. The Progress in International Reading Literacy Study (PIRLS) survey is also conducted by the IEA. The PIRLS tests pupils in primary schools in grade 4 in reading proficiency. Four rounds of PIRLS have been held to date in 2001, 2006, 2011 and 2016.

<sup>2</sup> IEA assessments define populations relative to specific grades, while PISA assessments focus on the age of pupils. In IEA studies, three different group of pupils were generally assessed: pupils from grade 4, grade 8 and from the last grade of secondary education. In 1995, two adjacent grades were tested in both primary (3-4) and secondary schools (7-8). To obtain comparable trends, we restricted the sample to grades 4 and 8. Some Canadian provinces and states in the United States of America have occasionally taken part in the IEA surveys.

In 2006, PIRLS included 41 countries/areas, two of which were African countries (Morocco and South Africa), 4 lower-middle-income countries (Georgia, Indonesia, Moldova, Morocco) and 8 upper-middle-income countries (Bulgaria, Islamic Republic of Iran, Lithuania, Macedonia, Federal Yugoslavian Republic, Romania, Russian Federation, South Africa). The 2011 round of PIRLS was carried out alongside TIMSS and included 60 countries/areas. The newest round of PIRLS in 2016 includes 50 countries.

*PISA*. The Organization for Economic Co-operation and Development (OECD) launched the Programme for International Student Assessment (PISA) in 1997 to provide comparable data on student performance. Since 2000, PISA has assessed the skills of 15-year-old pupils every three years. PISA concentrates on three subjects: mathematics, science and literacy. The framework for evaluation remains the same across time to ensure comparability.<sup>3</sup> In 2009, 75 countries/areas participated; in 2012, 65 countries/areas participated and in 2015, 72 countries/areas participated. An important distinction between PISA and IEA surveys is that PISA assesses 15-year-old pupils, regardless of grade level, while IEA assessments assess grade 4 and 8.

<sup>3</sup>As explained in the PISA 2006 technical report, this is only the case for reading between 2000-2009, for mathematics between 2003 and 2009 and for science between 2006 and 2009. See OECD (2010) for more details.

### *B. Regional Standardized Achievement Tests (RSATs)*

In addition to the above international assessments, a series of regional assessments have been conducted in Africa and Latin America and the Caribbean.

*SACMEQ*. The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ). SACMEQ is a psychometrically designed, standardized test which generally assesses math, reading and English in grade 6 pupils. The first SACMEQ round took place between 1995 and 1999. SACMEQ I covered seven different countries and assessed performance only in reading. The participating countries were Kenya, Malawi, Mauritius, Namibia, United Republic of Tanzania (Zanzibar), Zambia and Zimbabwe. The studies shared common features (research issues, instruments, target populations, sampling and analytical procedures). SACMEQ II surveyed pupils from 2000-2004 in 14 countries: Botswana, Kenya, Lesotho, Mauritius, Malawi, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania (Mainland), Tanzania (Zanzibar), Uganda, and Zambia. Notably, SACMEQ II also collected information on pupils' socioeconomic status as well as educational inputs, the educational environment and issues relating to equitable allocation of human and material resources. SACMEQ II also included overlapping items with a series of other surveys for international comparison, namely the *Indicators of the Quality of Education* (Zimbabwe) study, TIMSS and the 1985-94 IEA *Reading Literacy Study*. The third SACMEQ round (SACMEQ III) spans 2006-2011 and covers the same countries as SACMEQ II plus Zimbabwe. SACMEQ collected its latest round of data in 15 countries in East and Southern Africa from 2012-2014. These include: Botswana, Kenya, Lesotho, Mauritius, Malawi, Mozambique, Namibia, Seychelles, South Africa, Tanzania, Uganda, Zambia, Zanzibar and Zimbabwe. SACMEQ was designed and scaled to be comparable to past rounds. We include estimates from reports for this latest round of SACMEQ since the microdata is pending.

*PASEC*. The “Programme d’Analyse des Systèmes Éducatifs” (*PASEC*, or “Programme of Analysis of Education Systems”) was launched by the Conference of Ministers of Education of French-Speaking Countries (*CONFEMEN*). These surveys are conducted in French-speaking countries in Sub-Saharan Africa in primary school (grade 2 and 5) in math and French. Each round includes 10 countries. *PASEC I* occurred from 1996 to 2003; *PASEC II* from 2004 to 2010 and *PASEC III* was conducted in 2014. Of note, *PASEC* has not always been conducted simultaneously across countries and participation has varied considerably since 1994.<sup>4</sup> The most recent *PASEC* in 2014 uses Item Response Theory (*IRT*) and is a high-quality test. Ten countries participated, including Benin, Burkina Faso, Burundi, Cameroon, Chad, Congo, Cote d’Ivoire, Niger, Senegal and Togo. We include these countries using available microdata. Madagascar also participated in 2015 and was scaled to the *PASEC* 2014 round. We include Madagascar in our database using estimates from reports. To this end, inclusion of the recent *PASEC* data adds countries as well as enhances the quality of data from sub-Saharan Africa. This marks significant improvement over past datasets. To provide a link to past *PASEC* rounds, which use classical test theory and have substantially different test items, we create an inter-temporal comparison using Togo, which participated in all rounds of *PASEC*. However, given that *PASEC* did not conduct an intertemporal scaling calibration as is typical for item-response-theory *ISATs* and *RSATs*, priority should be given to analyzing the most recent *PASEC* 2014 data. Any intertemporal comparisons should be conducted with caution.

<sup>4</sup> The following is a list of participating countries in chronological order: Djibouti (1994), Congo (1994), Mali (1995), Central African Republic (1995), Senegal (1996), Burkina Faso (1996), Cameroon (1996), Côte d’Ivoire (1996), Madagascar (1997), Guinea (2000), Togo (2001), Mali (2001), Niger (2001), Chad (2004), Mauritania (2004), Guinea (2004), Benin (2005), Cameroon (2005), Madagascar (2006), Mauritius (2006), Republic of Congo (2007), Senegal (2007), Burkina Faso (2007), Burundi (2009), Côte d’Ivoire (2009), Comoros (2009), Lebanon (2009), Togo (2010), Democratic Republic of Congo (2010), and Chad (2010). Additional countries took a slightly different test between 2010 and 2011 (Lao PDR, Mali, Cambodia and Vietnam).

*LLECE*. The Latin American Laboratory for Assessment of the Quality of Education (LLECE) was formed in 1994 and is coordinated by the UNESCO Regional Bureau for Education in Latin America and the Caribbean. Assessments conducted by the LLECE focus on achievement in reading and mathematics in primary school. The first round was conducted in 1998 across grades 3 and 4 in 13 countries (Casassus et al. 1998, 2002). These countries include: Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Honduras, Mexico, Paraguay, Peru and the República Bolivariana de Venezuela (Casassus et al. 1998). The second round of the LLECE survey was initiated in 2006 in the same countries as LLECE I. In round two, called the Second Regional Comparative and Explanatory Study (SERCE), pupils were tested in grade 3 and grade 6. The Third Regional Comparative and Explanatory Study (TERCE), was done in 2013 across grades 3 and 6 and included 15 Latin American and Caribbean countries. Our analysis includes both SERCE and TERCE results, since these assessments are most similar and cover comparable grades.

*MLA*. A joint UNESCO and UNICEF project, *Monitoring Learning Achievement* (MLA), covers more than 72 countries and ranges from early childhood, basic and secondary education to non-formal adult literacy (Chinapah 2003). In our database, we use data across two subjects, mathematics and reading/literacy. We use data from reports for MLA I across 11 African countries of interest (Botswana, Madagascar, Malawi, Mali, Morocco, Mauritius, Niger, Senegal, Tunisia, Uganda and Zambia; see UNESCO 2000; Chinapah 2003). Of note, in the methodology section we explain that MLA data is used only if no other data is available since it is old, less reliable, and has not been repeated since 2000.



### C. The Early Grade Reading Assessment (EGRA)

The Early Grade Reading Assessment (EGRA) is a basic literacy assessment conducted in early grades. The assessment is conducted most often in grades 2-4. Since 2006, EGRA has been conducted in over 65 countries.

We analyze the reading comprehension indicator in EGRA, which is available in nearly all EGRA datasets, is less sensitive to differences in context and implementation, and has a strong conceptual link to RSATs and ISATs (Abadzi 2008; Dubeck and Gove 2015). To ensure robustness to language effects, we only include data when students took the test in their language of instruction. We use data for grades 2-4. The inclusion of EGRA adds 48 countries to the database with at least one data point in the past 10 years, nearly all of which are developing economies. This inclusion paves the way for inclusion of non-traditional assessments through careful choice of testing regime, indicators, language considerations, grade-level participation, and score scaling.

TABLE 9 – REVIEW OF STUDENT ACHIEVEMENT TESTS

Organization	Abbr.	Year	Subject	Countries/ Areas	Grade/Age
IEA	TIMSS	Every four years since 2003 (latest round is 2015)	M,S	38, 26, 48, 66, 65	4,8
UNESCO	MLA	2000	M,S,R	72	6,8
UNESCO	LLECE	2006, 2013	M,S,R	13, 16 (only 6 for science)	3,6
UNESCO	SACMEQ	2000, 2003, 2007, 2013	M,R	7, 15, 16	6 6
CONFEMEN	PASEC	2006, 2014	M,R	22 (before 2014), 10	Until 2014: 2,5 After 2014: 3, 6
IEA	PIRLS	Every five years since 2001 (latest round is 2016)	R	35, 41, 55	4
OECD	PISA	Every three years since 2000 (latest round is 2015)	M,S,R	43, 41, 57, 74, 65, 71	Age 15
RTI/USAID	EGRA	2007-2017	R	65	2,3,4

Notes: When denoting subjects, M=math; S=science; and R=reading.

## REFERENCES

- Abadzi, Helen.** 2008. "Efficient Learning for the Poor: New insights into Literacy Acquisition for Children." *International Review of Education* 54 (5-6): 581-604.
- Altinok, Nadir, Noam Angrist, and Harry A. Patrinos.** 2018. "Global Dataset on Education Quality 1965-2015." World Bank Policy Research Working Paper No. 8314.
- Barro, Robert J.** 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics* 106 (2): 407-443.
- Barro, Robert J. and Jong-Wha Lee.** 2001. "International Data on Educational Attainment: Updates and Implications." *Oxford Economic Papers* 53 (3): 541-563.
- Barro, Robert J. and Jong-Wha Lee.** 2013. "A New Data Set of Educational Attainment in the World, 1950–2010." *Journal of Development Economics* 104: 184-198.
- Becker, Gary S.** 1962. "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy* 70 (5): 9-49.
- Caselli, Francesco.** 2005. "Accounting for Cross-country Income Differences." *Handbook of Economic Growth, 1*, pp.679-741.
- Caselli, Francesco and Antonio Ciccone.** 2018. "The Human Capital Stock: A Generalized Approach. Comment." *American Economic Review*.
- Caselli, Francesco and John Coleman.** 2006. "The World Technology Frontier." *American Economic Review* 96.3: 499-522.
- Cohen, Daniel and Laura Leker.** 2014. "Health and Education: Another Look with the Proper Data", Manuscript.
- Cohen, Daniel and Marcelo Soto.** 2007. "Growth and Human Capital: Good Data, Good Results." *Journal of Economic Growth* 12 (1): 51-76.
- Das, Jishnu and Tristan Zajonc.** 2010. "India Shining and Bharat Drowning: Comparing Two Indian States to the Worldwide Distribution in Mathematics Achievement." *Journal of Development Economics* 92 (2): 175–187
- Dubeck, Margaret M. and Amber Gove.** 2015. "The Early Grade Reading Assessment (EGRA): Its Theoretical Foundation, Purpose, and Limitations." *International Journal of Educational Development* 40: 315-322.

- Feenstra, Robert. C., Robert Inklaar and Marcel P. Timmer.** 2015. "The next generation of the Penn World Table." *American Economic Review* 105(10): 3150-82.
- Filmer, Deon, Halsey Rogers, Noam Angrist, and Shwetlena Sabarwal.** 2018. "Learning-adjusted Years of Schooling (LAYS): Defining a New Macro Measure of Education." World Bank Policy Research Working Paper 8591. Washington, DC.
- Hall, Robert E. and Charles I. Jones.** 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 114(1), pp.83-116.
- Hanushek, Eric A. and Dennis D. Kimko.** 2000. "Schooling, Labor-force Quality, and the Growth of Nations." *American Economic Review* 90 (5): 1184-1208.
- Hanushek, Eric A. and Ludger Woessmann.** 2008. "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature* 46.3: 607-68.
- Hanushek, Eric A., and Ludger Woessmann.** 2012. "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation." *Journal of Economic Growth* 17 (4): 267-321.
- Hendricks, Lutz.** 2002. "How Important is Human Capital for Development? Evidence from Immigrant Earnings." *American Economic Review*, 92(1), pp.198-219.
- Hendricks, Lutz. and Todd Schoellman.** 2017. "Human Capital and Development Accounting: New Evidence from Wage Gains at Migration." *Quarterly Journal of Economics*, 133(2), pp.665-700.
- Holland, Paul W. and Neil J. Dorans.** 2006. "Linking and Equating." *Educational Measurement* 4: 187-220.
- Jones, Benjamin F.** 2014. "The Human Capital Stock: A Generalized Approach." *American Economic Review* 104.11: 3752-77.
- Jones, Benjamin F.** 2018. "The Human Capital Stock: A Generalized Approach. Reply." *American Economic Review*.
- Kim, Jim Yong.** 2018. "The Human Capital Gap: Getting Governments to Invest in People." *Foreign Affairs* 97: 92
- Kim, Jim Yong.** 2018. "Eliminating Poverty in the 21st Century: The Role of Health and Human Capital." *JAMA* 320.14: 1427-1428.
- Kraay, Aart.** 2018. "Methodology for a World Bank Human Capital Index." World Bank Policy Research Working Paper No. 8593.

- Krueger, Alan B. Mikael Lindahl.** 2001. "Education for Growth: Why and For Whom?" *Journal of Economic Literature* 39 (4): 1101-1136.
- Lange, Glenn-Marie, Quentin Wodon and Kevin Carey.** 2018. *The Changing Wealth of Nations 2018: Building a Sustainable Future*. The World Bank.
- Lee, Jong-Wha, and Hanol Lee.** 2016. "Human Capital in the Long Run." *Journal of Development Economics* 122: 147-169.
- Lucas, Robert E.** 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (1): 3-42
- Mankiw, N. Gregory, David Romer and David N. Weil.** 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107 (2): 407-437.
- Mincer, Jacob.** 1984. "Human Capital and Economic Growth." *Economics of Education Review* 3 (3): 195-205.
- Mislevy, Robert J., Albert E. Beaton, Bruce Kaplan and Kathleen M. Sheehan.** 1992. "Estimating Population Characteristics from Sparse Matrix Samples of Item Responses." *Journal of Educational Measurement* 29 (2): 133-161.
- Muralidharan, Karthik, and Nishith Prakash.** 2017. "Cycling to School: Increasing Secondary School Enrollment for Girls in India." *American Economic Journal: Applied Economics*, 9 (3): 321-50.
- Nelson, Richard R. and Edmund S. Phelps.** 1966. "Investment in Humans, Technological Diffusion, and Economic Growth." *American Economic Review* 56 (1/2): 69-75.
- OECD.** 2010. *PISA 2009 Results: Overcoming Social Background: Equity in Learning Opportunities and Outcomes (Volume II)*. OECD Publishing.
- Psacharopoulos, G.** 1994. "Returns to Investment in Education: A Global Update." *World Development*, 22(9), pp.1325-1343.
- Pritchett, Lant.** 2006. "Does Learning to Add Up Add Up? The Returns to Schooling in Aggregate Data." *Handbook of the Economics of Education* 1: 635-695.
- Romer, Paul M.** 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98 (5, Part 2), pp.S71-S102.
- Sandefur, Justin.** 2018. "Internationally Comparable Mathematics Scores for Fourteen African Countries." *Economics of Education Review* 62: 267-286.
- Schoellman, Todd.** 2011. "Education Quality and Development Accounting." *The Review of Economic Studies*, 79(1), pp.388-417.

- Smith, Adam.** 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations (Volume One)*. London: printed for W. Strahan; and T. Cadell, 1776.
- Steinmann, Isa, Rolf Strietholt and Wilfried Bos.** 2014. *Linking International Comparative Student Assessment*. LINC Technical Report.
- UNESCO.** 2017. "More Than One-half of Children and Adolescents are not Learning Worldwide." UIS Fact Sheet No. 46.
- World Bank.** 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC.
- World Bank.** 2019. *World Development Report 2019: The Changing Nature of Work*. Washington, DC.