Cross-Country Differences in the Quality of Human Capital*

Nicolai Kaarsen†

September 2011

Abstract

This paper constructs a cross-country measure of the quality of human capital using a novel approach based on international test scores data. The first main finding is that there are large differences in the quality of human capital - one year of schooling in the U.S. is equivalent to three or more years of schooling in a number of low-income countries. I incorporate the estimated series for the quality of human capital in an accounting framework calibrated using evidence on Mincerian returns. This leads to the second important finding, which is that the fraction of income differences explained by the model rise substantially when one includes the quality of human capital; the increase is around 25 percentage points.

*I would like to thank Carl-Johan Dalgaard, Oded Galor, Michael Svarer, David Weil, Asger Moll Wingender and seminar participants at Brown University, University of Copenhagen and the 25th Annual Congress of the European Economic Association for helpful comments and suggestions.

†Contact: University of Copenhagen, Øster Farimagsgade 5B, Building 26, DK-1353 Copenhagen K, Denmark. Nicolai.kaarsen@econ.ku.dk.

1 Introduction

How important is human capital in determining income per capita? The literature on development accounting seems to agree that human capital is an important determinant of income, but that the lion's share of the gap in income between poor and rich countries is not attributable to differences in human capital or physical capital endowments[1]. Instead, the main cause of the world income differences lies in differences in a residual productivity term which is unexplained. However, recent work by Manuelli and Seshadri (2006), Hanushek and Woessman (2009) and Schoellman (2011) suggest that the role of human capital may be underappreciated. The central charge is that the literature hitherto has has ignored differences in human capital quality.

In the development accounting literature, the human capital stock is usually computed using average years of schooling as the only input. This approach implicitly assumes that one year of schooling in Ghana is equal to one year of schooling in the U.S. If, however, one year of schooling in high-income countries is more productive relative to one year of schooling in low-income countries, human capital may be able to account for a larger share of income differences than previously thought.

In this paper, I estimate differences in the quality of human capital defined as the increase in cognitive skills obtained from an additional year of schooling. This measure can be directly incorporated into a development accounting framework. I find that there are large differences in the quality of human capital across countries. One year of schooling in the U.S. corresponds to three or even four years of schooling in many developing countries. Moreover, these quality differences are able to account for a considerable share of the variation in income across countries. I find that including the quality of human capital

[1] See e.g. Klenow and Rodriguez-Clare (1997), Hall and Jones (1999), Bils and Klenow (2002) and Caselli (2005).

increases the log-variance of human capital as a fraction of the log-variance of income from 0.04 to 0.26.

How is the quality of human capital estimated? I use an international test scores data set, which has the important feature that the same test was given to two different grades. This allows me to identify differences in the test scores gradient in years of schooling for a cross-section of countries. I define this gradient as the quality of years of schooling. It measures the effectiveness of one year of schooling in country i relative to one year of schooling in the U.S., which I choose to be the numeraire country. The measure can be seen as a conversion factor which adjusts years of schooling to be measured in U.S.-equivalent years of schooling.

This series is then used to evaluate the role of human capital quality in accounting for income differences. To do this I modify a standard accounting framework to include the estimated series of human capital quality. More specifically, I generalize the human capital production function of Bils and Klenow (2002) and calibrate the parameters such that the model is consistent with micro-evidence on Mincerian returns.

We would like the estimated measure of the quality of human capital to reflect the quality of an average worker in the labour force. However, the quality of human capital is estimated based on test scores of students which are not necessarily representative of the labour force. In particular, the earliest test scores data used in this paper is from 1995. Hence, in principle the estimated quality of human capital only reflects the quality of younger cohorts.

How could this bias the main results? If test scores have decreased over time, the cognitive abilities of young workers are low relative to those of the average worker in the labour force. In this case, the quality of human capital will be underestimated. Hence, if test scores have decreased over time in low-income countries relative to test scores in high income countries, the variance of quality of human capital will be biased upwards.

The evidence presented in this paper suggests that this is not an issue. In particular, I show that: 1. Differences in test scores across countries are considerably smaller than differences over time. 2. Changes in test scores over time are not correlated with income per capita.

A limitation of using test scores data is that it is only available for 65 countries and most of these are high-income countries. Hence, the sample of countries for which I can estimate the quality of human capital is not representative of the countries of the world. To deal with this problem I follow Weil (2007) in extending the data using a number of variables which are highly correlated to the quality data, and which are available for 174 countries. These variables are used to predict the quality of human capital for the countries where tests score data is not available. I find the same main conclusion using the small sample consisting only of countries where the quality of human capital is estimated as I do using the extended sample.

This paper is related to a string of contributions which seek to quantify the impact of human capital quality on growth. Within this literature two distinct approaches to the issue at hand can be identified.

The first approach seeks to identify an aggregate effect of human capital quality. This effect could go through many channels; some prominent examples are technology adoption (see e.g. Nelson and Phelps (1966)) and fertility (see e.g. Galor and Weil (2000)). Examples of papers seeking to estimate the aggregate effect are Hanushek and Kimko (2000) and Hanushek and Woessmann (2009).

The second approach focuses on identifying the effect running through individual productivity. Workers with better cognitive skills accomplish more complicated tasks faster and hence produce more. This paper belongs to that literature as does the papers of Hendricks (2002), Schoellman (2011) and Caselli (2005).

Hendricks (2002) uses the wages of immigrants in the U.S. to estimate the quality of human capital in their country of birth. To do this he compares wages levels of immigrants holding constant the level of education. His findings are quantitatively very similar to the standard findings of the literature as seen in e.g. Hall and Jones (1999) and Klenow and Rodriguez-Clare (1997), that is, human capital accounts for a relatively small fraction of income differences.

Schoellman (2011) also uses U.S. immigrant data to estimate differences in the quality of human capital, but he uses a different methodology than Hendricks (2002), and he reaches a different conclusion. Schoellman (2011) estimates a separate Mincer regression for each country of origin using the U.S. wage data and interprets the slope estimates as the quality of education pertaining to the respective countries. He includes these estimates in a development accounting framework similar to the framework used in the present paper, and finds that this increases the fraction of income explained by human capital considerably. Quantitatively, his results are very close to mine.

An issue with using immigrant data to infer the quality of human capital of source countries is that of selection. It is clear that immigrants are not selected randomly out of the population of the source country, however, it is unclear how the selection occurs and what the consequences are for the estimates of the quality of human capital. Another issue, pointed out by Friedberg (2000), is that education is not perfectly transferable across borders. Knowledge obtained in foreign countries may be valued less because of specificities in e.g. norms and institutions in the schooling system. Hence, the Mincerian return for a Dane in the U.S. might differ from the return of a Dane in Denmark. If such barriers are relatively higher for migrants coming from developing countries, the variance of the quality of human capital could be overestimated.

In the test scores data used in this paper, these issues are less likely to be a problem. The participating students are selected to be representative of the entire student population, and around 5000 students are tested in each country. Furthermore, to mitigate barriers of language and culture all test questions are back-translated and based on an international curriculum representative of all participating countries.

Caselli (2005) uses micro-evidence on the wage return to test scores as well as international test scores data to account for the effect quality of human capital on development. He finds that quality differences are relatively unimportant in explaining cross-country income differences. A drawback with this methodology is that the test scores data used to estimate the returns and the cross-country data on test scores are not comparable. They are based on different tests and different samples. Hence, it is very difficult to directly translate differences in international test scores into differences in productivity across countries.

The paper proceeds as follows. The next section estimates the quality of human capital. In section 3, these estimates are used in a development accounting exercise. Section 4 investigates the role of potential biases arising from using the test scores of younger cohorts to estimate the average quality of human capital. The final section concludes.

2 Estimating the quality of human capital

This section falls in four subsections. The first subsection presents the data, the second the methodology. The third subsection contains the main estimation results, and the final subsection expands the estimated human capital quality data to a larger number of countries.

2.1 The data

Trends in Math and Science Study (TIMSS) is a series of science and math tests conducted in schools in a number of countries in the years 1995 to 2007 by the International Association for the Evaluation of Educational Achievement (IEA). In each of the years 1995, 1999, 2003 and 2007, four different tests - a science test in primary school, a math test in primary school, math test in secondary school and a science test in secondary school - were administered in an unbalanced panel of countries. In the following, these four classifications will be denoted as *test types*.

The tests were assigned to a large number of students in each country (usually over 5000 students per test). Furthermore, great care was taken in constructing the tests so that they matched an international curriculum, and not just the curriculum of one country such as the U.S. In 1995, the same test was given to different grades which, as illuded to above, is invaluable to the identification of the quality of human capital.

In all of the TIMSS programs, each student is given a multiple choice test where the answers are ranked according to correctness. The grading of the tests is done separately for each test type, and is based on item response theory (IRT) which is a method used to convert answers into a test score. This conversion method is designed such that the resulting test scores are placed on a certain predetermined metric. In TIMSS, it is decided that the pooled sample of test scores from students of all countries in 1995 should have a mean of 500 and a standard deviation of 100. A detailed description of the method is given in Chapter 11 of TIMSS (2007).

Even though different tests were constructed from year to year some of the questions were repeated, which allows the IEA to temporally link the scaling of test scores such that all of the scores are placed on the 1995 metric. Thus, test scores are comparable over time.

Table 1 provides an overview of the availability of the data. All of the test scores data used in this paper are country averages. The maximum number of participating countries in one type of test is 46 (secondary math test, 2007), and the total number of countries which participated in at least one test is 65.

Table 1: Availability of TIMSS test scores data

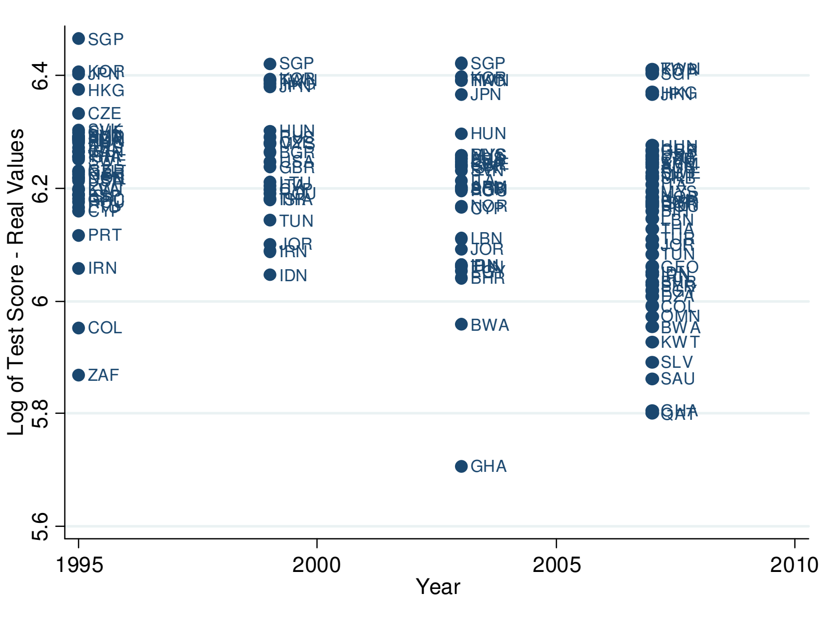
Test type:		Math primary		Science primary		Math secondary		Science secondary	
Grade:		3rd	4th	3rd	4th	7th	8th	7th	8th
Year	Publication								
1995	TIMSS (1997a,b)	25	25	25	25	38	38	38	38
1995	TIMSS (2008)	0	25	0	25	0	38	0	38
1999	TIMSS (2008)	0	0	0	0	0	22	0	21
2003	TIMSS (2008)	0	20	0	20	0	31	0	30
2007	TIMSS (2008)	0	35	0	32	0	46	0	43

Notes: The table shows the number of participating countries in the TIMSS round for a given year and grade.

The main source of test scores data is TIMSS (2008). This publication includes all the data shown in Table 1 with the exception of the data for 3rd and 7th graders from the 1995 round. TIMSS (1997a,b) includes the data for 3rd and 7th graders, but the test scores from this publication are placed on a slightly different metric. Fortunately, since both publications contain data on 4th and 8th graders from the 1995-round, it is possible to rescale the results for the 3rd and 7th graders from the TIMSS (1997a,b) metric to the TIMSS (2008) metric. I use linear regression to do this for each of the four test types; the details are in the appendix.

The final data set consists of 1170 observations of four different test scores from 65 countries from four different years. The correlation between test types is high. In 1995, the year with the most observations, the average correlation between two test types is 0.82. Figure 1 shows the rescaled test scores over time for mathematics tests of 8th graders.

Figure 1: Log of TIMSS test scores for mathematics in 8th grade



Notes: The TIMSS test scores have been rescaled to the metric of TIMSS (1997a,b), see the main text for details.

At the high end of the spectrum we find East Asian countries such as Korea, Japan and Singapore. The Western industrialized countries lie in the middle, while low-income countries such as Colombia, South Africa and Ghana are amongst the countries with the lowest test scores.

2.2 The model and the empirical specification

To estimate the quality of human capital I assume that the test score $T_{k,i,s,t}$ of test type k, in country i, at grade s and in year t is determined by the following production function:

$$T_{k,i,s,t} = f\left(s \times q_i, \Gamma_k\right), \tag{1}$$

where q_i is the quality of years of schooling, and Γ_k is a vector of parameters. For example, $T_{msec,US,8,2003} = f(8 \times q_{US}, \Gamma_{msec})$ is the 8th grade math test score of the U.S. in 2003. It is produced with 8 years of schooling and the quality of U.S. education as input and the production function parameters are those of the math secondary school test type.

The goal is to use a specific functional form for (1) to estimate q_i and Γ_k jointly using the test scores data. This produces an estimated series for q_i for the 65 different countries participating in the TIMSS.

The set of parameters Γ_k are varying across test types. Thus, in effect, (1) reflects four different test scores production functions, one for each test type. Since the test scores are not comparable across test types, that is, one point on the math primary scale does not correspond to one point of the science secondary scale, we cannot use the same production function for all test types.

In all four production functions, however, I let the same q_i enter as input. This is because we wish to estimate the general level of human capital quality in the country. This seems like a reasonable assumption given that the test scores are closely correlated across test types, as was shown above. Furthermore, q_i is assumed constant over time, since we wish to estimate the overall quality of human capital ignoring idiosyncratic changes in test scores over shorter periods. As will be evident in the robustness section, this assumption appears to be plausible since the variation of test scores over time is substantially smaller than the variation across countries.

(1) assumes that there are no differences in the quality of human capital evaluated at zero years of schooling. Such differences could be captured by adding a country-specific constant term to (1). Differences in e.g. parental education or health could generate cross-country variation in the cognitive ability measured at zero years of schooling. Unfortunately, given that we only have data for two years of schooling within each test type, it is not possible to add a constant term capturing such effects. Furthermore, to estimate

the quality of human capital at zero years of schooling one would have to include cognitive tests of persons who did not go to school, and such data is difficult to find for a large cross-section of countries.

Which functional form should we choose for (1)? Since the physical capacities of the brain places an upper bar on the stock of knowledge which can be accumulated, it seems plausible that the test scores production function exhibits decreasing returns. Hence, I will use the following functional form which potentially satisfies this assumption:

$$T_{k,i,s,t} = \beta_k \left(s \times q_i\right)^{\gamma_k}, \tag{2}$$

where β_k and γ_k are a production function parameters, which, as noted, are allowed to vary across test types. If γ_k is estimated to be below one, (2) exhibits decreasing returns to the input $s \times q_i$.

I estimate (2) using the following empirical specification:[2]

$$\ln T_{k,i,s,t} = \ln \beta_k + \gamma_k \left[\ln s + \sum_{j=1}^{65} D_{ji} \ln q_i\right] + e_{k,i,s,t}, \tag{3}$$

where D_{ij} is a country dummy which is one if $j = i$, and $e_{k,i,s,t}$ is an error term. (3) is estimated using non-linear least squares. Even though the number of participating countries is 65 and each country potentially could have had up to 18 different test scores, the total number of observations is only 544 since some of countries participated in only one or two years.

[2] (2) could also be estimated using OLS as:

$$\ln T_{k,i,s,t} = \ln \beta_k + \gamma_k \ln s + \sum_{j=1}^{65} D_{ji} x_{k,i} + e_{k,i,s,t},$$

where $x_{k,i} = \gamma_k \ln q_i$. The estimates of q_i could then be inferred from the estimates of $x_{k,i}$ and γ_k. However, this method would not give us the standard deviations of the estimated q_i's directly as estimation output.

Note, that it is only possible to estimate (3) since we have test scores data for the same test type at different grades. If we only had data for one year there would be no variation in s. In this case, since β_k, γ_k and q_i are parameters, there would be no variation in the right-hand side of (3).

To identify β_k and the q_i's separately it is necessary to fix one of the q_i's. I choose to set the q_i of U.S. to one ($q_{US} \equiv 1$). Thus, q_{US} acts as a numeraire allowing us to interpret $1/q_i$ as the years of education it takes for the average student in country i to learn as much as the average student in the U.S. learns in one year. $s \times q_i$ is then denoted quality-adjusted years of schooling or U.S.-equivalent years of schooling. The multiplication sign is included to underline that $s \times q_i$ is the product of two separate variables, but will be suppressed from now on as will the index i.

2.3 Results

Estimating (3) results in a data set of q's spanning 65 countries. The estimated q's and their standard errors are shown in the appendix. The mean standard error of the estimated q's is 0.04 and it does not exceed 0.07 for any one country. Table 2 shows the estimated parameters for the test scores production functions.

Table 2: Estimated parameters of the test scores production functions

k (test type):	Primary math	Primary science	Secondary math	Secondary science
Parameter				
β (scaling parameter)	5.63 (0.04)	5.63 (0.04)	5.31 (0.06)	5.43 (0.05)
γ (elasticity)	0.49 (0.03)	0.49 (0.03)	0.47 (0.03)	0.41 (0.02)

Summary: This table shows the estimated parameters of the four test scores production functions. For all four test types, the elasticity is below 1 indicating that there is decreasing returns to scale in the production of test scores. Furthermore all parameters are estimated with a fairly high degree of precision.

Notes: The table shows the output of the estimated parameters in the test score production function. All parameters are significant at a 1%-level. Standard errors are in paranthesis.

The standard errors of all parameters are low compared to the point estimates. Furthermore, the estimated β_k and γ_k are relatively constant over test types. γ_k is around 0.5 and significantly different from 1 for all test types. This confirms our prior that the returns to test scores are decreasing in inputs.

Figure 2 below shows a scatter plot of log GDP per worker and the quality of human capital for the 41 countries for which both data series were available.

Figure 2: Log of GDP per worker in 2007 and quality of human capital.

Notes: The quality of human capital is estimated using test scores data, see the main text for details on how this is done.

The figure shows that q varies considerably across countries going as low as 0.25-0.30 in Yemen and Ghana. In a country where $q = 0.5$ the average student achieves in two years of schooling what the average student in the U.S. achieves in one year. There is a strong positive relationship between q and log income per worker. Regressing on log income on q yields a slope of 1.74 with an R^2 of 0.25.

An interesting finding is that for many Eastern Asian countries the quality of human capital is relatively high compared to the level of GDP per capita. The q of e.g. Singapore, Japan, South Korea and Japan is around 1.2 implying that one year of schooling generates 20% more knowledge as in the U.S.

Another interesting group of outliers consists of the oil-producing countries Saudi Arabia, Quatar, Kuwait and Oman. They all have relatively a high GDP per worker but a low quality of human capital.

2.4 Extending the data set of quality-adjusted years of schooling

The data set created in the previous section consists of only 65 countries with an over-representation of high-income countries. Therefore, this section extends the sample of q's. The method used to do so is taken from Weil (2007).

First, the quality of education estimated in the previous section is regressed on a set of variables. Second, the predicted values from this regression are then used as q's for 107 countries which do not have estimated data for q.

I use the following three variables which are strongly correlated to the quality of human capital: GDP per worker (in 2007, from Penn World Table 7.0), average years of schooling (in 2005, from Barro and Lee (2010), and the population density (population per 100 m^2 in 2000, extracted online from UNdata[3]). I also include region dummies[4].

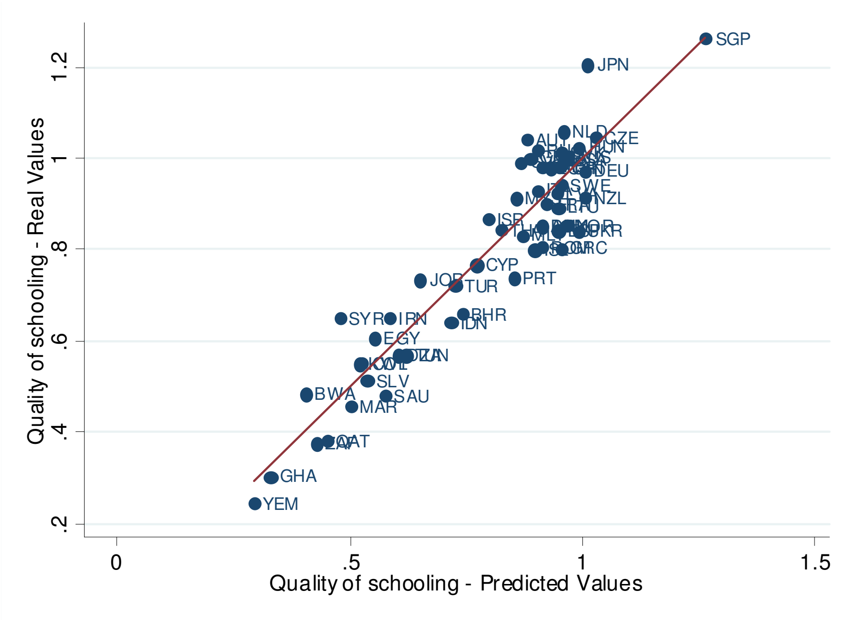
Regressing the estimated q on these variables for the 56 countries for which all data is available yields an R^2 of 0.87. The details of the regression are shown in the appendix.

[3] http://data.un.org/

[4] I use the region dummies provided by Barro and Lee (2010).

Figure 3 plots the actual values against the predicted values. It shows that the regression line provides a reasonable fit, also in the case of countries with lower q.

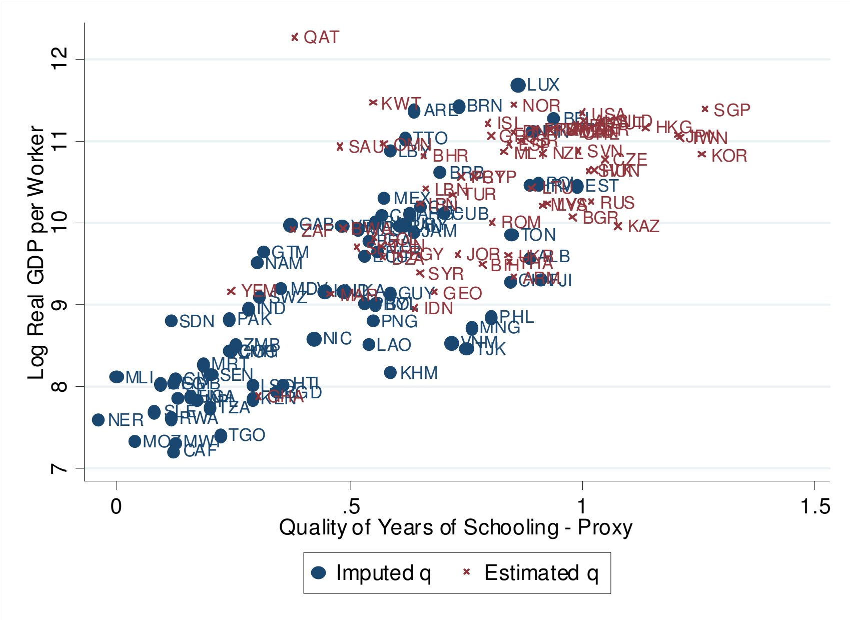
Figure 3: Predicted and actual values of quality of schooling.



Notes: The actual values of the quality of schooling are estimated using test scores. Predicted values are based on regressing actual values on a number of proxy variables. See the main text for more details.

The above regression results are now used to predict q's for 74 countries resulting in a data set containing q for 139 countries shown in Table A1 in the appendix. Figure 4 plots log income per worker against q for the full data set.

Figure 4: Log of GDP per worker, 2007 and quality of years of schooling - full sample.



Notes: The quality of schooling is estimated using international test scores data. For countries where the test scores data is not available the quality of human capital has been imputed. See main text for details

As in Figure 2, there is a strong positive relationship between log GDP per worker and q and a lot of dispersion in the quality of human capital. For countries with a q below 0.25 (the q of Yemen), data is extrapolated and should only be used for further analysis with caution[5]. As is evident from the figure this is the case for quite a lot of developing countries. Hence, all decomposition results below are shown for both the sample of extrapolated q's and the baseline sample of estimated q's. The main results are similar for the two samples.

[5] For Mali and Nigeria the imputed q was slightly below zero. For both countries, I replace q by zero.

3 Quality-Adjusted Years of Schooling and World Income Differences

The estimated differences in the quality of human capital across countries of the world are substantial. However, can they account for the large income differences observed? To answer this question we need to incorporate the quality measure in a development accounting framework. I rely on the standard framework used by e.g. Klenow and Rodriguez-Clare (1997), Hall and Jones (1999), Bils and Klenow (2002) and Caselli (2005).

The first step is to generate human-capital stocks. I generate two series, one which is based on the assumption that there are no differences in the quality of human capital. This series is based entirely on average years of schooling and thus follows the usual methodology from the literature. The second series is computed incorporating the quality data estimated above.

The second step is to use the two series in an accounting framework. I use two tools which are commonplace in the literature of development accounting. The first is decomposition of the log-variance of income; the second is decomposition of differences in income percentiles. Using either of these tools gives the same result which is that incorporating differences in the quality of human capital substantially increases the fraction of income differences explained by the model.

3.1 Construction of human-capital stocks

To generate human capital stocks I follow Schoellman (2011) who generalize the framework of Bils and Klenow (2002) to include quality of human capital. The backbone of the model is the Cobb-Douglas production function[6]:

$$y = Ak^{\alpha}h^{1-\alpha}, \tag{4}$$

where y is output per worker, A is total-factor productivity, k is physical capital per worker, and h is human capital per worker. To use (4) to account for differences in output we need data for y, k and h. A is computed as a residual under the assumption that $\alpha = 1/3$. For y and k I use data from 2008 from PWT 7.0[7]. The computation of the cross-country series for h proceeds as follows.

I assume that the human capital production function is:

$$h = e^{\phi(s,q)}, \tag{5}$$

where s is years of schooling, q is quality of human capital, and $\phi(s,q)$ is given by

$$\phi(s,q) = \frac{\theta}{1-\psi}(sq)^{1-\psi}. \tag{6}$$

For $q = 1$ this collapses into the production function used by Bils and Klenow (2002). Data for s is taken from Barro and Lee (2010) and data for q is taken from above. I calibrate θ and ψ using evidence on the return of log wages to years of schooling. The basic methodology is similar to that of Bils and Klenow (2002) and Schoellman (2011).

The standard reference for data on Mincerian returns is Psacharopoulos (1994) who surveys a list of micro studies. In this data, the returns to years of schooling is consistently higher for developing countries. However, a newer survey from Banerjee and Duflo (2005)

[6]This analysis does not take into account the effect of human capital on income levels through externalities as e.g. technology. Altough such effects potentially could be important, there is no relliable estimates of the magnitude of such effects, so they are left out of the analysis.

[7]k is computed by using the perpetual inventory method described in Caselli (2005).

concludes that, on average, the return to years of schooling is around 0.1 for both developing and developed economies. Hence, I assume that the return to years of schooling is constant at 0.1.

To calibrate θ and ψ first assume that markets are competitive implying that the wage of the individual worker is given by his human capital times the return on human capital in the country. Hence, the Mincerian returns MR are given by

$$MR = \frac{\partial \ln h}{\partial s} = \theta s^{-\psi} q^{1-\psi}, \tag{7}$$

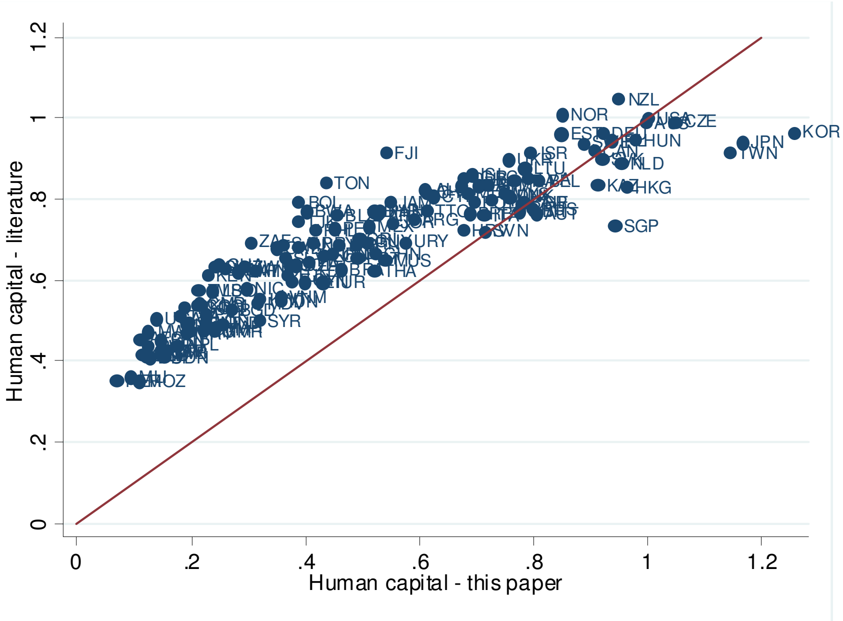
Now insert $MR = 0.1$, take logs and rearrange to obtain:

$$\ln s = \frac{\ln \theta - \ln 0.1}{\psi} + \frac{1-\psi}{\psi} \ln q. \tag{8}$$

Estimates of θ and ψ can be backed out from the estimated constant term and slope one gets from regressing $\ln s$ on $\ln q$. To do this, I take data for q from above and data for s from Barro and Lee (2010). I use the sample of 65 estimated q's, however, using the extended sample of q's does change not calibration results markedly. 58 of these countries have data for s. For this sample I regress $\ln s$ on $\ln q$ yielding an estimated constant and slope of 2.36 and 0.51, respectively. The corresponding standard errors are 0.027 and 0.062, respectively, and R^2 is 0.55. Backing out the implied parameters of the human capital production function yields $\psi = 0.66$ and $\theta = 0.48$.

With these calibrated values for ψ and θ we can use (8) to compute a series for quality-adjusted human capital. As a benchmark, I also computed a human capital series under the assumption that $q = 1$, implying that $h = e^{0.1s}$. Figure 5 below plots the two different measures of human capital against each other.

Figure 5: Human capital as estimated in the literature vs. human capital adjusted for quality differences.



Notes: Both measures are relative to human capital in the U.S. See the main text to get a description of how the human capital variables are constructed.

In the figure, I have normalized the human capital measures such that U.S. human capital is equal to one. The figure also shows a 45^o line. Accounting for the quality differences increases the variance of human capital. Human capital computed using only years of schooling varies from 0.4 to around 1 - a factor 2.5. Quality-adjusted human capital varies from around 0.1 to 1.2 - a factor 12.

3.2 Accounting results

How large a share of income differences can the generated human capital stocks account for? To find out first take logs and variance of (4) yielding

$$var\left[\ln y\right] = var\left[\ln A\right] + var\left[\ln X\right] + 2Cov\left[\ln A, \ln X\right], \tag{9}$$

where $X = k^{\alpha}h^{1-\alpha}$ is GDP per capita predicted by the pure input-factors model. This formulae forms the basis of the variance decomposition exercise found in e.g. Caselli (2005)[8]. With data for y, k and h, A can be computed as a residual using (4). Hence, we have all the data needed to compute the terms in (9).

An alternative way to evaluate the role of various factors in explaining income differences is to compare income ratios at different percentiles. Define V_f as the fth fractile of V. For instance, y_{90}/y_{10} is the ratio of income at the 0.9 to income at the 0.1 fractile. $(h_{90}/h_{10})\,/\,(y_{90}/y_{10})$ and $(x_{90}/x_{10})\,/\,(y_{90}/y_{10})$ measure, respectively, the fraction of the income ratio explained by human capital alone and the fraction of the income ratio explained by the model.

Table 3 below shows the results of the decomposition of log-variance of income and income ratios.

[8] Another way to decompose income differences, based on Hall and Jones (1999) and Klenow and Rodriguez-Claire (1997), is to write (4) in terms of capital-output ratio instead of capital-labour ratio. This tends to put more weight on human capital differences and less weight on physical capital differences. Hence, the increase in the fraction income differences explained by the model is even larger. See the appendix for results.

Table 3: Decomposition of income differences

Sample:	Only estimated q's (N = 47)		Imputed and estimated q's (N = 116)	
Includes quality of human capital:	No	Yes	No	Yes
	(1)	(2)	(3)	(4)
Var(lny)	0.64	0.64	1.65	1.65
Var(lnh)	0.04	0.19	0.06	0.48
Var(lnx)	0.15	0.30	0.38	0.82
Var(lnh)/Var(lny)	0.06	0.29	0.04	0.29
Var(lnx)/Var(lny)	0.23	0.46	0.23	0.50
(Var(lnx)+Cov(lnA,lnx))/Var(lny)	0.46	0.64	0.46	0.67
y_{90}/y_{10}	5.62	5.62	26.93	26.93
h_{90}/h_{10}	1.61	2.80	1.96	7.24
x_{90}/x_{10}	2.74	3.99	5.10	11.70
$(h_{90}/h_{10})/(y_{90}/y_{10})$	0.29	0.50	0.07	0.27
$(x_{90}/x_{10})/(y_{90}/y_{10})$	0.49	0.71	0.19	0.43

Summary: This table decomposes income differences into different components. It shows that accounting for the quality of human capital increases substantially the fraction of income differences explained by the model.

Notes: y is GDP per capita, h is human capital, x is GDP per capita predicted by the factors-only model. Subscript 90 indicates the 90% percentile and subscript 10 indicates the 10% percentaile. The first two columns show the results for the small sample of countries where data for q is estimated from test scores data. The third and fourth columns show the results for the large sample of countries, where q is imputed. See section on details of how data on q is estimated and imputed. In the first and third columns, human capital is computed using only years of schooling. In the second and fourth columns human capital is computed using quality of human capital and years of schooling.

The two first columns show the results for the small sample consisting only of countries for which estimated q's are available. Columns 3 and 4 show the results for the extended sample consisting of countries for which q is either estimated or imputed. Columns 1 and 3 show the results for the model where there are no differences in q, whereas columns 2 and 4 show the results for the model where differences in q are taken into account.

First of all, the table confirms what Figure 5 illustrated: accounting for quality differences increases the variance of human capital. Moreover, this increase appears to be substantial. For both samples $var(\ln h)/var(\ln y)$ increases substantially when the quality of human capital is taken into account. In the full sample, this figure increases from 0.04 to 0.26. Schoellman (2011) who uses immigrant wage data to compute quality-adjusted human capital finds that $var(\ln h)/var(\ln y) = 0.26$, a number which is identical to mine. Although he uses a different methodology and different data to estimate the quality of human capital, he reaches the same conclusion as this paper.

Turning to the fifth row, $var(\ln x)/var(\ln y)$, which is Caselli's (2005) preferred measure of model success, increases substantially when human capital quality is included. For both samples $var(\ln x)/var(\ln y)$ increases from around 24% to 47%[9].

Another thing to note from this row is that, in the baseline model with no differences in the quality of human capital, $var(\ln x)/var(\ln y) = 0.24$. This is substantially smaller than the 0.39 reported by Caselli (2005). Since the model and method of decomposition are the same, these differences can be attributed to the differences in the data used[10]. To check that the choice of data does not make a difference to the main results I redid the calibration and variance decomposition using Caselli's (2005) data set. Using this data does not change the main conclusion. The results are given in the appendix.

As shown in the sixth row, adding a covariance term increases the fraction of income differences explained by both models substantially. However, the main result that the quality of human capital explains a large share of income differences holds through.

Using income ratios as done in the bottomn 5 rows produces similar conclusions as those obtained by looking at variance decomposition. For both samples, the fraction of income differences explained by the model increases substantially when the quality of human capital is included.

For the large sample, $(h_{90}/h_{10})/(y_{90}/y_{10})$ is equal to, respectively, 0.07 and 0.23 in the model without and with differences in the quality of human capital. Hence, my finding for

[9] Unfortunately, it is not possible to compare these numbers to the findings of Schoellman (2011) since he does not report results involving income predicted by the model x.

[10] Caselli (2005) uses earlier versions of Penn World Table and the Barro-Lee years of schooling data. He also uses data from an earlier year (1995) and the sample is different. Although the precise reasons for this reduction in the fraction of the variance explained by the baseline model are interesting, it is beyond the scope of this paper to investigate this matter further.

the model with no differences in human capital matches exactly those of Hall and Jones (1999) who also find that $(h_{90}/h_{10})/(y_{90}/y_{10}) = 0.07$. As for the model where quality differences are accounted for, my findings are again very close to those of Schoellman (2011) who finds that $(h_{90}/h_{10})/(y_{90}/y_{10}) = 0.21$.

To sum up the main findings, accounting for quality differences seems to increase the explanatory power of the model substantially. This conclusion holds through for different samples and different methods of decomposition.

4 Trends in test scores and representativeness of students

As noted above, I use student test scores from 1995 and later to infer the quality of human capital of the entire labour force in 2007. However, the cognitive abilities of older cohorts could differ from those of younger cohorts. If, for instance, test scores have increased over time, the quality of human capital will be underestimated.

If, for instance, test scores historically have increasted faster in poor countries than in rich countries, the estimated differences in human capital quality across the world may be underestimated. Accordingly, in this case the accounting results should be viewed as lower bound estimates of the contribution from human capital quality. Of course, if quality has risen faster in ex ante rich countries, human capital quality differences may be overestimated.

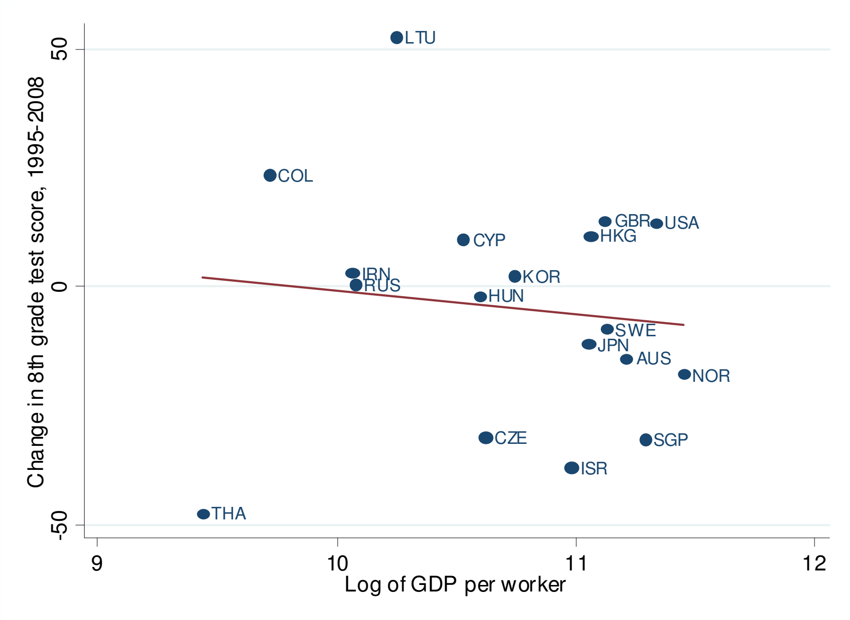
As another example, suppose human capital quality has risen worldwide at a fairly uniform speed. Then human capital quality may be overestimated in richer countries where the population is more mature.

Hence, it is useful to examine whether test scores tend to change over time, and, if so, at a differential speed in rich and poor countries. The following empirical evidence suggests that: 1. The variation in test scores over time is low relative to the variation across countries. 2. Changes in test scores over time are not systematically related to income per capita.

To show the first point I focus on test scores of 8th graders since these are available for most countries and most years. I first compute the average math and science test score for the 8th grade. To find the across-time standard deviation I then compute, for each country, the standard deviation of the test score in 1995 and 2007. The average across-time standard deviation across the 19 countries where data is available is 11. To find the cross-country standard deviation, I first compute the average test score over the years from 1995-2007 ignoring missing observations. The standard deviation of the average test scores across the 62 countries which have at least one observation from 1995-2008 is 61. The standard deviation across the 19 countries which has data for both 1995 and 2007 is 48. It is clear that although the test scores are not completely constant over time the cross-country variation dwarfs the cross-time variation.

To show the second point I look at the change in the average math and science test score for the 8th grade from 1995 to 2007. Figure 1 below plots this change against GDP per capita in 2007.

Figure 6: Log GDP per worker and change in test scores.



Notes: The 8th grade test scores is computed as the average between a math test score and a science test score. The change measured in absolute terms, that is, in points on a scale from 100 to 500.

The slope coefficient is -4.87 and has a t-value of -0.47. Thus, it does not seem that the change in test scores over time is correlated to the level of economic development. If anything, the correlation is negative. In this case, we would expect the quality of human capital to be overestimated in low-income countries, which implies that the analysis above underestimates the importance of human capital quality in accounting for income differences. But the evidence presented in this section suggests that the biases resulting from using students of recent cohorts to infer the average quality of human capital of the labour force should be relatively small and not change the main findings.

5 Conclusion

With a few exceptions, the development accounting literature has so far implicitly assumed that education is equally effective across countries. That is, one year of education in the U.S. correspond to one year of education in Ghana. This paper challenges this assumption.

I use test scores data to estimate the differences in human capital quality, and find that there are large differences in the quality of education throughout the world. In many developing countries the average student needs two years of schooling or more to gain knowledge corresponding to what the average student in the U.S. learns in one year.

Including the quality of human capital in a development accounting exercise increases the fraction of the variance of income explained by the model by around 0.25. The fraction of income differences explained by human capital alone is around 0.26, which is very close to the findings of Schoellman (2011), who uses immigrant data to estimate the quality of human capital.

While the methodology used produces a measure of the overall quality of human capital across countries it has one limitation. It does not illuminate the fundamental causes of quality differences. Future work should concentrate on quantifying the relative importance of parental input, teacher quality, health etc. in explaining cross-country differences in the quality of human capital.

Appendix

A Rescaling of Test Scores

First, four separate linear regressions are run, one for each test type, with the TIMSS (1997a,b) version of the 1995 test scores as left-hand side variables and the TIMSS (2008) version of the 1995 test scores as right-hand side variables:

$$T^{97}_{k,i,95} = B_k + b_k T^{08}_{k,i,95} + \varepsilon_i, \tag{10}$$

where $k = mpri$, $spri$, $msec$, $ssec$ indexes test types and i indexes countries. In all four regressions, the correlation is strong with R^2's in the range of $0.85 - 0.93$. The next step is to use the regression results to convert the TIMSS (2008) version of the test scores from 1999, 2003 and 2007 into the scale of TIMSS (1997a,b). This is done by replacing the TIMSS (2008) values of test scores from 1999, 2003 and 2007 with the predicted values from the regressions, i.e.

$$\hat{T}^{97}_{k,i,t} = \hat{B}_k + \hat{b}_k T^{08}_{k,i,t}. \tag{11}$$

B Table of the Quality of Human Capital for a Cross Section of Countries

Table B: The quality of human capital by country

Country	Iso	q	Std. dev.	q, imputed
Afghanistan	AFG			0.09
Albania	ALB			0.88
Algeria	DZA	0.57	0.03	0.57
Argentina	ARG			0.63
Armenia	ARM	0.85	0.03	0.85
Australia	AUS	1.00	0.03	1.00
Austria	AUT	1.04	0.04	1.04
Bahrain	BHR	0.66	0.04	0.66
Bangladesh	BGD			0.34
Barbados	BRB			0.69
Belgium	BEL			0.94
Belize	BLZ			0.55
Benin	BEN			0.13
Bolivia	BOL			0.55
Bosnia and Herzegovina	BIH	0.79	0.06	0.79
Botswana	BWA	0.48	0.03	0.48
Brazil	BRA			0.54
Brunei Darussalam	BRN			0.73
Bulgaria	BGR	0.98	0.04	0.98
Burundi	BDI			0.09
Cambodia	KHM			0.59
Cameroon	CMR			0.24
Canada	CAN	0.98	0.04	0.98
Central African Republic	CAF			0.12
Chile	CHL			0.65
China	CHN			0.84
Colombia	COL	0.55	0.03	0.55
Congo	COG			0.24
Costa Rica	CRI			0.57
Croatia	HRV			0.88
Cuba	CUB			0.70
Cyprus	CYP	0.77	0.03	0.77
Czech Republic	CZE	1.05	0.03	1.05
Côte d'Ivoire	CIV			0.13
Denmark	DNK	0.85	0.04	0.85
Ecuador	ECU			0.53
Egypt	EGY	0.61	0.04	0.61
El Salvador	SLV	0.51	0.03	0.51
Estonia	EST			0.99
Fiji	FJI			0.91
Finland	FIN			0.89
France	FRA	0.90	0.05	0.90

Country	Iso	q	Std. dev.	q, imputed
Gabon	GAB			0.37
Gambia	GMB			0.12
Georgia	GEO	0.68	0.04	0.68
Germany	DEU	0.97	0.04	0.97
Ghana	GHA	0.30	0.03	0.30
Greece	GRC	0.80	0.03	0.80
Guatemala	GTM			0.31
Guyana	GUY			0.58
Haiti	HTI			0.36
Honduras	HND			0.44
Hong Kong	HKG	1.14	0.04	1.14
Hungary	HUN	1.02	0.03	1.02
Iceland	ISL	0.80	0.03	0.80
India	IND			0.28
Indonesia	IDN	0.64	0.03	0.64
Iran, Islamic Republic of	IRN	0.65	0.02	0.65
Ireland	IRL	0.99	0.04	0.99
Israel	ISR	0.87	0.03	0.87
Italy	ITA	0.93	0.04	0.93
Jamaica	JAM			0.64
Japan	JPN	1.20	0.04	1.20
Jordan	JOR	0.73	0.04	0.73
Kazakhstan	KAZ	1.08	0.07	1.08
Kenya	KEN			0.29
Korea, republic of	KOR	1.26	0.04	1.26
Kuwait	KWT	0.55	0.03	0.55
Lao People's Democratic Republic	LAO			0.54
Latvia	LVA	0.92	0.03	0.92
Lebanon	LBN	0.66	0.04	0.66
Lesotho	LSO			0.29
Libyan Arab Jamahiriya	LBY			0.58
Lithuania	LTU	0.89	0.03	0.89
Luxembourg	LUX			0.86
Malawi	MWI			0.13
Malaysia	MYS	0.91	0.04	0.91
Maldives	MDV			0.35
Malta	MLT	0.83	0.06	0.83
Mauritania	MRT			0.18
Mauritius	MUS			0.51
Mexico	MEX			0.57
Mongolia	MNG			0.76
Morocco	MAR	0.46	0.03	0.46
Mozambique	MOZ			0.04
Namibia	NAM			0.30
Nepal	NPL			0.17
Netherlands	NLD	1.06	0.04	1.06
New Zealand	NZL	0.92	0.03	0.92
Nicaragua	NIC			0.42
Norway	NOR	0.85	0.03	0.85
Oman	OMN	0.57	0.04	0.57
Pakistan	PAK			0.24
Panama	PAN			0.61
Papua New Guinea	PNG			0.55

Table B, continued: The quality of human capital by country

Country	Iso	q	Std. dev.	q, imputed
Paraguay	PRY			0.53
Peru	PER			0.56
Philippines	PHL			0.80
Poland	POL			0.90
Portugal	PRT	0.74	0.03	0.74
Qatar	QAT	0.38	0.03	0.38
Romania	ROM	0.81	0.03	0.81
Russian Federation	RUS	1.02	0.03	1.02
Rwanda	RWA			0.11
Saudi Arabia	SAU	0.48	0.04	0.48
Senegal	SEN			0.20
Serbia	SER	0.84	0.05	0.84
Sierra Leone	SLE			0.08
Singapore	SGP	1.26	0.04	1.26
Slovakia	SVK	1.01	0.04	1.01
Slovenia	SVN	0.99	0.03	0.99
South Africa	ZAF	0.38	0.03	0.38
Spain	ESP	0.84	0.05	0.84
Sri Lanka	LKA			0.49
Sudan	SDN			0.11
Swaziland	SWZ			0.31
Sweden	SWE	0.94	0.04	0.94
Switzerland	CHE	0.98	0.05	0.98
Syria	SYR	0.65	0.05	0.65
Taiwan	TWN	1.21	0.05	1.21
Tajikistan	TJK			0.75
Tanzania, United Republic of	TZA			0.20
Thailand	THA	0.84	0.03	0.84
Togo	TGO			0.22
Tonga	TON			0.85
Trinidad and Tobago	TTO			0.62
Tunisia	TUN	0.57	0.03	0.57
Turkey	TUR	0.72	0.05	0.72
Uganda	UGA			0.16
Ukraine	UKR	0.84	0.04	0.84
United Arab Emirates	ARE			0.64
United Kingdom	GBR	1.00	0.03	1.00
United States	USA	1.00		1.00
Uruguay	URY			0.62
Venezuela, Bolivarian Republic of	VEN			0.48
Viet Nam	VNM			0.72
Yemen	YEM	0.25	0.02	0.25
Zambia	ZMB			0.25
Zimbabwe	ZWE			0.35

C Regression output for sample extension

Table C: Results of proxy regression

	Dependent variable: estimated quality of human capital
GDP per worker	-0.088
	(0.069)
Years of schooling	0.497
	(0.075)
Population density	0.242
	(0.064)
Region dummies	Yes
N	56
R2	0.870

Summary: This table shows the results of the regression used to extend the sample of estimated q's.

Notes: GDP per worker is from Penn World Table v 7.0, average years of schooling of the population above 15 years is from Barro and Lee (2010), Population density is from (www.). Coefficients are standardized. Standard errors are in paranthesis.

D Robustness to choice of data

A surprising result found in Table 3 of the main section is that the baseline model without differences in the quality of human capital is only able to account for a fraction 0.23 of income differences. This number is considerably higher in the literature e.g. in Caselli (2005) who gets 0.4. Given that the Barro-Lee data set as well as the Penn World Table data has been revised several times it is perhaps not so surprising that differences occur. However, it does leave us with the question of whether the main results are robust to using an earlier version of the data.

To show that the main results are not affected by the choice of data I redo the construction of human capital stocks and the development accounting using Caselli's (2005) data for years of schooling, GDP per capita and the capital stock. The accounting results are shown in Table D below. The table shows that the main results persist. For both samples the explanatory power increases considerably when human capital quality is added to the model.

Table D: Decomposition of income differences - using Caselli's (2005) data

Sample:	Only estimated q's (N = 41)		Imputed and estimated q's (N = 89)	
Includes quality of human capital:	No	Yes	No	Yes
	(1)	(2)	(3)	(4)
Var(lny)	0.42	0.42	1.13	1.13
Var(lnh)	0.07	0.17	0.08	0.26
Var(lnx)	0.22	0.34	0.45	0.67
Var(lnh)/Var(lny)	0.16	0.41	0.07	0.23
Var(lnx)/Var(lny)	0.53	0.80	0.40	0.59
(Var(lnx)+Cov(lnA,lnx))/Var(lny)	0.88	1.08	0.73	0.84
y_{90}/y_{10}	3.79	3.79	18.28	18.28
h_{90}/h_{10}	2.04	2.76	2.17	4.15
x_{90}/x_{10}	2.98	3.99	6.64	10.20
$(h_{90}/h_{10})/(y_{90}/y_{10})$	0.54	0.73	0.12	0.23
$(x_{90}/x_{10})/(y_{90}/y_{10})$	0.79	1.05	0.36	0.56

Summary: This table decomposes income differences into different components. It shows that the main results are robust to using the data set of Caselli (2005) for years of schooling, GDP and the physical capital stock.

Notes: y is GDP per capita, h is human capital, x is GDP per capita predicted by the factors-only model. Subscript 90 indicates the 90% percentile and subscript 10 indicates the 10% percentaile. The first two columns show the results for the small sample of countries where data for q is estimated from test scores data. The third and fourth columns show the results for the large sample of countries, where q is imputed. See section on details of how data on q is estimated and imputed. In the first and third columns, human capital is computed using only years of schooling. In the second and fourth columns human capital is computed using quality of human capital and years of schooling.

E Robustness to method of decomposition

In this appendix, I show that qualitatively the main results persist if one uses the alternative method of decomposition based on Hall and Jones (1999) and Klenow and Rodriguez-Clare (1997). In this case, the per capita production function is written in terms of the capital-output ratio instead of the capital-labour ratio.

The main reason for doing this is that in a neoclassical growth model, larger TFP or more human capital per worker increases the capital labour ratio in steady state. Hence, differences in TFP or human capital might be wrongly attributed to differences in capital per worker. To deal with this issue we can write GDP per capita in terms of the capital-output ratio, which is unaffected by changes in TFP in a neoclassical growth model.

In order to do this, first rewrite (4) to get

$$y = A^{\frac{1}{1-\alpha}} \left(\frac{k}{y}\right)^{\frac{\alpha}{1-\alpha}} h. \tag{12}$$

In this case, I define $Z \equiv (k/y)^{\alpha/(1-\alpha)}\, h$ to be the contribution from input factors and $B \equiv A^{\frac{1}{1-\alpha}}$ to be the contribution from TFP. To decompose the variance take logs and variance to get:

$$var\left[\ln y\right] = var\left[\ln A\right] + var\left[\ln Z\right] + 2Cov\left[\ln B, \ln Z\right] \tag{13}$$

This yields an expression identical to (9) with Z replacing X.

The results of the variance decomposition based on (13) are shown below. Table D shows that the results are robust. For both samples the explanatory power increases considerably when human capital quality is added to the model.

Table E: Variance decomposition - using capital-output ratio

Sample:	Only estimated q's (N = 47)		Imputed and estimated q's (N = 116)	
Includes quality of human capital:	No	Yes	No	Yes
	(1)	(2)	(3)	(4)
Var(lny)	0.64	0.64	1.65	1.65
Var(lnz)	0.05	0.21	0.13	0.60
Var(lnz)/Var(lny)	0.08	0.33	0.08	0.37
(Var(lnx)+Cov(lnB,lnz))/Var(lnz)	0.19	0.47	0.19	0.51

Summary: This table decomposes income differences into different components. It shows that accounting for the quality of human capital increases substantially the fraction of income differences explained by the model.

Notes: y is GDP per capita, h is human capital, z is GDP per capita predicted by the factors-only model using the capital-output ratio. B is TFP computed as a residual. The first two columns show the results for the small sample of countries where data for q is estimated from test scores data. The third and fourth columns show the results for the large sample of countries, where q is imputed. See section on details of how data on q is estimated and imputed. In the first and third columns, human capital is computed using only years of schooling. In the second and fourth columns human capital is computed using quality of human capital and years of schooling.

References

Banerjee, A. V., and Duflo, E. (2005). "Growth Theory Through the Lens of Development Economics." Handbook of Economic Growth, Volume 1A. Edited by P. Aghion and S. N. Durlauf. Elsevier Science, North-Holland Publishers.

Barro, R., and Lee, J., (2010). "A New Data Set of Educational Attainment in the World, 1950-2010". NBER Working Paper No. 15902.

Bils M. and P.J. Klenow (2000). "Does Schooling Cause Growth?" The American Economic Review 90(5), 1160-1183.

Caselli, F. (2005). "Accounting for Cross-Country Income Differences." Handbook of Economic Growth, Volume 1A. Edited by P. Aghion and S. N. Durlauf. Elsevier Science, North-Holland Publishers.

Card, D. (1999). "The Causal Impact of Education on Earnings." Handbook of Labour Economics, Volume 3. Edited by O. Ashenfelter and D. Card.

Friedberg, R. M. (2000). "You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital". Journal of Labor Economics 18(2), 221-251.

Galor, O. and Weil, D. N. (2000). "Population, Technology and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." The American Economic Review 90 (4), 806-828.

Gollin, D., Lagakos, D. and Waugh, M. (2011). "The Agricultural Productivity Gap in Developing Countries." Working paper available at

https://sites.google.com/site/davidlagakos/home/research.

Hall, R.E., Jones, C.I. (1999). "Why do some countries produce so much more output per worker than others?" The Quarterly Journal of Economics 114 (1), 83–116.

Hanushek, E.A., Kimko, D.D. (2000). "Labor-Force Quality and the Growth of Nations." The American Economic Review 90(5), 1184-1208.

Hanushek, E.A, Woessman, L. (2009). "Do Better Schools Lead to more Growth? Cognitive Skills, Economic Outcomes and Causation." NBER Working Paper 14633.

Hendricks, L. (2002). "How Important Is Human Capital for Development? Evidence from Immigrant Earnings." The American Economic Review 92 (1), 198-219.

Klenow, P. J. and Rodriguez-Clare, A. (1997). "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?". NBER Macroeconomics Annual 12(1997), 71-103.

Manuelli, R.E. and Seshadri, A. (2006) "Human Capital and the Wealth of Nations." Presented at USC FBE Dept. Macroeconomics & International Finance Workshop.

Nelson, R. R. and Phelps, E. S. (1966). "Investments in Humans, Technology Diffusion and Economic Growth." The American Economic Review 56(1), 69-75.

Psacharopoulos, G. (1994). "Returns to Investment in Education: A Global Update." World Development 22(9), 1325–1343.

Schoellman, T. (2011) "Education Quality and Development Accounting". Forthcoming in Review of Economic Studies.

TIMSS (1997a), "TIMSS Highlights from the Primary Grades." Third International Mathematics and Science Study. Boston College.

TIMSS (1997b), "Highlights of Results from TIMSS." Third International Mathematics and Science Study. Boston College.

TIMSS (2008), "Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth and Eighth-Grade Students in an International Context." U.S. Department of Education and National Center for Education Statistics.

UNESCO (2006) "Education for All Global Monitoring Report - Literacy for Life." UNESCO publishing.

Weil, N. D.(2007). "Accounting for the Effect of Health on Economic Growth." The Quarterly Journal of Economics 122(3), 1265-1306.