Bureau d'économie théorique et appliquée (BETA)
UMR 7522

Documents de travail

« Does one size fit all? The impact of cognitive skills on economic growth »

Auteurs

Nadir Altinok, Abdurrahman Aydemir

Document de Travail n° 2016 – 34

Juin 2016

Faculté des sciences économiques et de gestion
Pôle européen de gestion et d'économie (PEGE)
61 avenue de la Forêt Noire
F-67085 Strasbourg Cedex

Secrétariat du BETA
Géraldine Del Fabbro
Tél. : (33) 03 68 85 20 69
Fax : (33) 03 68 85 20 70
g.delfabbro @unistra.fr
www.beta-umr7522.fr

UNIVERSITÉ DE LORRAINE
UNIVERSITÉ DE STRASBOURG

Does one size fit all?
The impact of cognitive skills on economic growth

Nadir Altinok and Abdurrahman Aydemir[1],[2]

This version: 03/06/2016

Abstract. Our paper reassesses the question of the impact of cognitive skills on economic growth using new indicators for cognitive skills. These data extend measures of cognitive skills substantively. In particular, our data extends the coverage of less developed countries, among them adding 27 countries of Sub-Saharan Africa, a continent that was largely missing from the analysis of the effects of learning outcomes on economic growth. Using this extended dataset and employing several identification strategies, cognitive skills are found to have a positive impact on economic growth. We address the heterogeneity in the causal effect of cognitive skills on growth and show that the effect of skills on growth differs across regions and by the economic level of countries. Our results indicate that high-income countries should focus on increasing the number of high skilled pupils, while countries from Sub-Saharan Africa would benefit more by investing in the development of basic skills.

Keywords: Education Quality, Cognitive Skills, PISA, Human Capital, Growth, Development.

J.E.L. Classification: H5, I2, O4.

[1] Nadir Altinok: BETA, CNRS & University de Lorraine (France), UFR Droit, Sciences Economiques et Gestion, 13 Place Carnot, 54035 Nancy Cedex, France. nadir.altinok@univ-lorraine.fr
Abdurrahman Aydemir: Sabanci University, Faculty of Arts and Social Sciences, Orhanlı, Tuzla 34956, Istanbul, IZA, CReAM-UCL, aaydemir@sabanciuniv.edu
[2] We are grateful to Erik Hanushek, Harry Pritchett, Geraint Johnes, and Jean-Luc de Meulemeester for their comments. We also benefited from the comments of the participants of the 19th EBES Conference and the International Symposium on Education and Training organized by the CSEFRS. Usual disclaimers apply.

1. Introduction

The question of which factors determine economic growth has been a major topic in economic research. Levine and Renelt (1992) reported initial income level and the share of investment in GDP as the only variables that consistently had statistically significant impacts on economic growth across a wide array of specifications. They find that neither primary nor secondary enrolment rates consistently had positive, statistically significant effects on economic growth. The importance of human capital for economic growth has been called into question by a large number of studies that failed to find a positive relationship between the quantity of education and economic growth in cross-country analysis. Recent studies, however, have pointed out the importance of school quality as opposed to quantity (Barro, 1991, Hanushek and Kimko, 2000, Hanushek and Woessmann, 2012a) and have provided evidence of the positive effect of school quality on the rate of economic growth.

This paper extends this recent literature in a number of substantive dimensions. We use a recent dataset that covers many countries, particularly less developed ones, that could not be included in growth regressions by previous studies. For example, among the newly added countries, our database includes 27 countries of Sub-Saharan Africa, a continent that was largely missing from the analysis of the effects of learning outcomes on economic growth. The study also updates the period of analysis by including the most recent data on schooling quality (between 1965 and 2012). Using these data, we pay particular attention to the heterogeneity in the estimated impact of the relationship between education and growth according to level of development. We also assess the importance of minimum and advanced skills on growth experiences of countries with different levels of development. In addition to reporting cross-sectional regression results, we also present results from recently proposed alternative estimation methods for identifying the causal effect of education quality on growth using a panel data framework and new set of instruments. This allows to test the robustness of the estimated impact of cognitive skills on economic growth based on different estimation strategies and subsamples and provides several novel results.

Many previous studies have analyzed the impact of education on economic growth (see Glewwe *et al.*, 2014, Durlauf, Johnson, and Temple, 2005 for literature reviews).[3] In spite of the large number of recent empirical studies carried out on data involving international comparisons, the assumption of a clear and positive relationship between the investment in human capital and economic growth has been largely called into question. While some international comparative studies have shown that many educational variables are among factors that determine the growth of per capita GDP in a country (Barro, 1991; Mankiw *et al.*, 1992), available education indicators such as schooling rates or the average number of school years in these studies have been criticized as vague measures of human capital. Barro (1991) was the first study to emphasize school quality along with other measures of education. Barro presented evidence that school quality matters; higher primary and secondary pupil-teacher ratios in 1960 have marginally significant negative impacts.[4] In 2001, Lant Pritchett underlined the controversies surrounding the relationship between education and growth (Pritchett, 2001). Among explanations put forward of why the most robust econometric analyses do not make it possible to prove a stable and positive relationship between human capital and economic growth, Pritchett highlighted the importance of the quality of education and argued that if the quality of education is so low it may not produce the necessary skills to lead to economic growth.

An important number of research papers analyzed the education-growth relationship. In this paper, we only focus on studies that have included a qualitative dimension to education[5]. Following Barro's analysis that underlined the importance of school quality, the work of Hanushek and Kimko (2000) was the first to include measures of educational quality using data from international student achievement tests (hereafter ISATs). Using data from ISATs that administered similar mathematics and science tests to students in 31 countries, they construct normalized test scores and include these as measures of the human capital stock (called "Labor Force Quality") in their analysis of economic growth (from 1960 to 1990). They find that the coefficient estimate for this variable is positive and statistically significant, and its inclusion in

[3] The question raised by Levine and Renelt (1992) was revisited by Sala-I-Martin, Doppelhofer, and Miller (2004) who ranked variables by their robustness in growth regressions and found that the 1960 primary school enrolment rate is the second most robust variable. Durlauf *et al.* (2005) on the other hand highlight in their review that perhaps the high standard set by Levine and Renelt (1992) may be too strict.

[4] Including a school quality variable can also be interpreted as correcting for measurement error, since years of schooling or enrolment rates may measure human capital with error.

[5] For further details, see the comprehensive review by Durlauf, Johnson, and Temple (2005).

the analysis causes the years of schooling variable to lose statistical significance. Their estimated impact is large, suggesting that a one standard deviation increase in schooling quality (measured by test performance) increases the rate of economic growth by 1.4 percentage points. An important shortcoming of this analysis is the inclusion of the learning outcomes variable for only one point in time, hence making it impossible to analyze economic growth rates as a function of changes in learning outcomes.

A recent paper by Hanushek and Woessmann (2012a) (hereafter HW) aimed at improving the work of Hanushek and Kimko (2000). Using the Cohen and Soto (2007) years of education instead of the Barro and Lee dataset, they update test score data that include more developing countries (50 countries, of which 27 are developing countries) and the period analyzed is extended to cover the 1960-2000 period. Another important change is related to the methodology used: HW used an approach that assumes stability over time of the variance of quality of student achievement in a restricted number of OECD countries, which serves as a standardization benchmark for performance variation over time. The authors call this benchmark group the "OECD Standardization Group" (OSG) of countries.[6] HW assume that cross-country variation among the OSG countries has not changed substantially since 1964, and using this assumption, they build new indicators of student achievement and educational quality. Their main measure of cognitive skills is a simple average of all standardized math and science test scores from the ISATs in which a country participated. Their database includes a combined measure for the 50 countries that have ever participated in any of the math and science tests. Confirming results from Hanushek and Kimko (2000), HW find that years of schooling have no impact on economic growth when the test score measure is included. Moreover, the test score variable is highly significant and adding it increases the adjusted R squared from 0.25 to 0.73. Overall, a one standard deviation increase in school quality is associated with a 1.3-2.0 percentage point higher rate of economic growth.

Following the pioneering analysis of Hanushek and Kimko (2000), we start from the idea that one year of education in country *i* does not confer the same rate of return as one year of education

[6] The OSG countries are: Austria, Belgium, Canada, Denmark, France, Germany, Iceland, Japan, Norway, Sweden, Switzerland, the United Kingdom, and the United States. The authors suggest two criteria for determining the countries to be included in this benchmark group. First, the countries have to be member states of the relatively homogenous and economically advanced group of OECD countries in the whole period of ISATs observations. Second, the countries should already have had a substantial enrollment in secondary education in 1964.

in country *j*. For instance, a one-year increase in education in Japan may not give the same output as a one-year increase yields in Botswana. Thus, the studies which take into account only quantitative indicators of education will be biased, owing to the fact that they regard human capital as a homogeneous factor of production. Using the recent dataset in this paper which is an updated version of Altinok *et al.*, 2014, we investigate the relationship between cognitive skills and the average annual growth rate of the economy between 1960 and 2010, a positive correlation is evident between the two (Figure 1). Our paper aims at improving and extending the literature in a number of ways. Firstly, we use a larger dataset for our test score variable, by including more developing countries: our dataset provides an increase in the number of countries of about 60%. A recent database (Altinok *et al.*, 2014) compiled the results of countries in the international assessments of pupils for each skill (mathematics, sciences and reading) between 1960 and 2007. While this original dataset provides data until 2007, we replicate the same methodology in order to update it until 2012. Compared to the Altinok et al. (2014) dataset, while our updated dataset includes only 3 additional countries, by including more recent testing, it achieves a substantive extension of 30% in the number of observations overall. This increase permits us to improve the precision of our learning outcomes indicators, since we have more observations for each country. The resulting updated database includes comparable cognitive skills for 125 countries, as compared to HW who take into account 77 countries. For instance, our database includes 27 countries of Sub-Saharan Africa which were not captured by previous research.

Secondly, the methodology used for our test score dataset is exclusively based on datasets available at the student level, which helps to reduce errors in measures of school quality. For instance, we improved the measure of test scores between 1960 and 1980 by adjusting the test scores of pioneering achievement tests with original raw data provided by the IEA (International Association for the Evaluation of Educational Achievement)[7]. In most previous studies, only mean scores published in international reports were used. In order to reduce measurement error

[7] A major improvement of the Altinok et al. (2014) dataset deals with the information relative to trends between the first waves of IEA studies. Indeed, the trends between FIMS and SIMS and between FISS and SISS are only available for a common number of items. The list of these items and the anchoring process is described in IEA Reports (see Keeves, 1992; Robitaille and Garden, 1989). This additional information permitted Altinok et al. (2014) to adjust assessments on NAEP anchoring, but also to reproduce the trends found in these reports. Since, a significant number of items were used in both assessments, the trends found in these reports may reduce the bias which occurs when we only use NAEP anchoring.

bias, we also chose not to include data from an assessment for which there is no raw data available. This is the reason why we excluded International Assessment of Educational Progress (IAEP)[8] and Monitoring Learning Achievement (MLA) assessments. Indeed, specific corrections are often needed for mean scores from student achievement tests in order to obtain comparable scores between countries and over time, even for the very recent student achievement tests such as TIMSS 2003. For instance, some countries took part in several student achievement tests, but with different sampling procedures. With original raw data, it is possible to exclude specific populations in order to obtain more comparable data on learning outcomes.

Thirdly, since our data include countries from several continents, in addition to estimating an average effect of education on economic growth, we can also test for heterogeneity in these effects. This paper is the first one in literature that assesses within a causal framework the differences in the amplitude of cognitive skills and growth relationship.[9] In particular, we exploit the availability of more than 80 countries in our data and provide estimates separately by (i) income level of countries, (ii) regions, and (iii) total factor productivity. This analysis provides novel evidence on the cognitive skills and economic growth relationship. For example, among other regions, our analysis provides results for Arab countries and Sub-Saharan African countries, a region the growth experience of which received little attention by previous studies due to data constraints.

Nelson and Phelps (1966) very early suggested that the most powerful technologies are adopted and implemented more quickly by those economies richest in human capital. From this viewpoint, it is the *level* of education which raises the *growth rate* of the economy, by accelerating the assimilation of technical progress. Consequently, if a country devotes, in any one year, more resources to education and thus increases its stock of human capital, the result will be an increase in the growth rate of the economy. According to the ideas of Nelson and Phelps, it may be that education also increases the capacity to carry out strictly economic choices, in

[8] Unfortunately, raw data for IAEP are not available. We would like to thank ETS and NCES for their support.

[9] Castelló-Climent and Hidalgo-Cabrillana (2012) develop a theoretical model of human capital investments distinguishing between low- and high-quality education. Using the Hanushek and Kimko (2000) dataset, they show the education quality has a positive effect on growth only when quality is relatively high and conclude that quality may be not growth enhancing unless students achieve a minimum level of knowledge. Their main empirical exercise, however, does not control for potential endogeneity of cognitive skills and includes few developing countries compared to our data. Where they use an IV strategy to address endogeneity lagged values of education quality is used as an instrument but the authors caution against the validity of this instrument.

particular suitable allocation of resources. This approach thus supposes that the output of education is all the higher as there are major opportunities for technological imitation and adaptation in an unstable universe. For testing this hypothesis, we also conduct an analysis that tests whether the effect of minimum and advanced level of cognitive skills differ between countries. This analysis aims at answering which types of skills matter most for the economic growth of less developed and more developed regions. We find that high income countries need to focus on the share of pupils reaching the advanced level while low income countries and in particular Sub-Saharan African countries may invest more on education policies which focus on the share of pupils who reach the minimum level in mathematics and science.

We also control for potential endogeneity and measurement error by using an IV estimation strategy. Since our dataset includes more countries, the instruments used by HW in their panel data analysis are incomplete in our case. New instruments for panel regressions have been proposed recently in the literature (Islam *et al.*, 2014; Adams-Kane & Lim, 2014). We use this alternative set of instruments (government effectiveness and health measures) in an IV-GMM specification to test the robustness of results to endogeneity and measurement bias issues.

Lack of comparable data on cognitive skills is a general issue for growth empirics research. After improving the data by including more developing countries, our analysis yields five main results. i) While we cannot find a robust effect of the quantity of schooling (measured as initial years of education), the coefficient associated with our updated cognitive skills variable is quite strong over most estimations. These results confirm those reported by HW. ii) Our results show that including more developing countries increases the overall impact of cognitive skills on economic growth by about 25%. iii) Moreover, we find that the magnitude of the effect is highest for low-income countries, followed by high-income and middle-income countries. iv) Lastly, a focus on the share of basic and top performers within each country highlights different effects between subsamples. While in high-income countries the share of top performers in student achievement tests has a strong and positive effect on economic growth, it is the share of students reaching the minimum level which has the most impact on economic growth for countries from Arab States and Sub-Saharan Africa. Hence, after taking into account potential reverse causality issues, measurement error and omitted variable bias, the results provide strong evidence that investing in the quality of education enhances the economic growth of countries.

In section 2, we outline a simple growth model that forms the basis of our estimation. Section 3 presents the data sources and general methodology used to construct our database on the test scores measure. Section 4 estimates the contribution of the quality of education to economic growth in a cross-section dataset. Section 5 deals with potential endogeneity and measurement error bias. Section 6 includes a panel regression where we test for the effect of improved test scores on changes in economic growth. We then explore potential heterogeneity of the impact of cognitive skills in economic growth (Section 7). Lastly, section 8 concludes our analysis.

2. A Simple Growth Model

Following HW, we use a simple growth model: a country's growth rate (g) is a function of the skills of workers (H) and other factors (X). These factors include initial levels of income and technology, specific institutional dimensions, and other factors that are used in the growth empirics. Skills are often referred to simply as the workers' human capital stock. Our specification assumes that H is a one-dimensional index and that growth rates are linear in these inputs:

$$g = \gamma H + \beta X + \varepsilon \quad (1)$$

The most important specification issue in this framework is the nature of the skills (H) and where they might come from. In the educational production function literature (Hanushek, 2002), skills are explained by many factors such as family inputs (F), the quantity and quality of inputs provided by schools (qS), individual ability (A), and other relevant factors (Z) which include labor market experience, health, and other specific characteristics:

$$H = \alpha F + \beta(qS) + \gamma A + \delta Z + v \quad (2)$$

In our specification, the schooling term (qS) combines school attainment (S) and its quality (q). However, human capital is a latent variable that cannot be directly observed. Hence, we need a correct measure of human capital in order to test its impact on economic growth. The main existing theoretical and empirical work on growth begins by taking the quantity of schooling of workers (S) as a direct measure of H. Following the pioneering analysis of Hanushek and Kimko (2000), we focus directly on the cognitive skills component of human capital and evaluate H with test-score measures of mathematics, science, and reading achievement. There are many advantages of using measures of educational achievement (Hanushek and Woessmann, 2012). Firstly, they capture outputs of schooling by focusing on differences in the knowledge and ability generated by schools. Secondly, since they include all the general skills, they do not only rely on school skills but also skills from other sources (families and general ability). Another important advantage of using cognitive skills is the ability to assess the importance of different policies designed to affect the quality aspects of schools since cognitive skills allow for differences in performance among students with the same quantity of schooling.

3. Data and methodology

The dataset related to cognitive skills used in this paper builds upon the work of Altinok *et al.* (2014) and updates the 1960-2007 data to 1960-2012. Based on new data sources and the alternative method of anchoring, there are several innovations in this dataset compared to previous research. The construction of this data benefits from international student achievement tests (ISATs) as well as regional student achievement tests (RSATs). ISATs include the well-known TIMSS, PIRLS and PISA tests.[10] Along with these international assessments, three major RSATs conducted in Africa and Latin America, such as LLECE, SACMEQ or PASEC[11], which were never used in previous research on the effect of cognitive skills on economic growth. This help us to extend the available data to a larger set of countries, in particular improving the representation of developing world. For instance, our updated dataset includes 27 countries of sub-Saharan Africa which were not captured by previous research.[12] The resulting updated database in this paper includes comparable cognitive skills for 125 countries, as compared to HW who take into account 77 countries between 1960 and 2000.[13] This allows us to test the education quality and economic growth relationship with a much extended data set and also address

[10] Respectively the Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS) and Programme for International Student Assessment (PISA).

[11] Respectively the Latin American Laboratory for Assessment of the Quality of Education (LLECE), the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and the Program on the Analysis of Education Systems (PASEC).

[12] A description of various existing learning assessments is provided in Appendix A and detailed information on each assessment is provided in Table A.1.

[13] It should be noted that the number of countries included inside estimations is always lower than the number of countries for which we have comparable data on cognitive skills. The main reason is the lack of data on other explanatory variables. For instance, whereas HW compiled comparable data on cognitive skills for 77 countries, only 50 of them were included in different estimations. In our case, whereas we have data for 120 countries, our sample is reduced to 80 countries.

heterogeneity in the effects of education quality on economic growth. While the dataset of HW includes data for about 4,7 billion of people, our updated dataset increases this number by about 10%. Most importantly, while the HW study covers around 220 million people from Arab states and Sub-Saharan Africa, our updated dataset comprise more than double this figure (approximately 500 million people). This larger representation of the developing countries permits us to deeply explore potential heterogeneity hypothesis of the relationship between cognitive skills and economic growth.

The methodology to generate comparable achievement scores across countries used in Altinok et al. (2014) aims at improving the seminal work by Barro and Lee (1996) and Barro (2001), and consists of a major update of a previous work by Altinok and Murseli (2007). Hanushek and Kimko (2000) and Hanushek and Woessmann (2012a) also use a method of anchoring for their database of cognitive skills across 77 countries. The alternative methodology for creating the data used in this paper differs from Hanushek and Woessmann (2012a) in that it takes into account several improvements made by ISATs since 1995 and enables the inclusion of the main regional assessments that were absent in previous datasets. Details of this methodology are provided in Appendix A.

Ideally, the evaluation of the impact of cognitive skills on economic growth would need measures of the skills of workers in the labor force. However, some of our measures of cognitive skills based on recent testing (e.g. the tests conducted after late 2000s) include students who are still in school. As has been highlighted by HW, this creates a tradeoff: incorporating more recent testing has the potential advantages of improved assessments and observations on a greater number of countries (especially developing countries) but it also weights any country measure more toward students and less toward workers.[14]

4. Baseline results

In this section, we report cross-sectional estimates of the cognitive skills and economic growth relationship based on equation (1). Since we use an extended dataset based on a different methodology to HW, we first replicate results from HW using their own data as well as our

[14] Two international tests (the International Assessment of Adult Literacy and the Programme for International Assessment of Adult Competencies) have suggested the possibility of panel estimation across countries, because they have tested adults rather than students (see Coulombe & Tremblay, 2006; Hanushek & Woessmann, 2015).

dataset confined to the HW sample. We then report results from our new extended data, which show to what extent earlier results reported in the literature hold for a larger set of countries and with more recent data. In addition, we estimate the relationship between cognitive skills and economic growth for various subsamples of countries and using different set of tests to construct education quality measures. Table 1 presents the baseline results. This table is divided into three parts. The first part (part A) replicates Table 1 from HW using the same dataset and sample of countries. In the second part (B) of Table 1, we use our dataset that expands the years used for calculation of test scores to 2012 but restrict the sample to the countries in HW. This allows us to check to what extent the expanded time span for the tests in our dataset provides additional information compared to previous research. Part (C) of Table 1 uses our updated dataset with the extended set of countries and aims to test the robustness of previous estimates to the inclusion of additional countries.[15]

Results from Table 1-A replicate the estimation of HW for the 50 countries with cognitive skills and economic data over the period 1960-2000. Following their methodology, we use version 6.1 of the Penn World Tables (Heston *et al.*, 2002), while the data on years of schooling come from Cohen and Soto (2007).[16] The first column of Table 1-A presents estimates of a simple growth model with school attainment. In the second column, adding cognitive skills increases the explained variance from 31% to 75%. Whether we include (col. 3) or exclude (col. 2) initial school attainment in 1960, results remain unchanged. Results associated with the coefficient of the "cognitive skills" variable are significant and their amplitude are quite similar to those reported by HW. When we include the mean of school attainment between 1960 and 2000 (col. 4), results remain the same, confirming the estimates of HW. The remaining columns of Table 1, Panel A provide alternative specifications. In column 5, we employ regression techniques that are robust to outliers (excluding Botswana and Nigeria). Including regional dummies reduces the estimated test effect from 1.2 to 1.0 (col. 6). In columns 7 and 8, we consider economic institutions. We control for institutional differences in openness of the economy and security of property rights in col. 7 and introduce fertility rates and location in the

[15] Because we need data for both economic growth and cognitive skills between 1960 and 2010, all former communist countries are eliminated even if they have test measures. This explains why our estimation does not include 125 countries.

[16] HW explain that they use an extended version of the Cohen and Soto (2007) data. However, they do not explicitly explain the methodology used. We predict results from the Barro and Lee (2013) dataset for missing values from Cohen and Soto (2007) data. This may explain slight differences in results.

tropics as additional controls in col. 8. Although the amplitude of the effect is greatly reduced, it remains significant in all specifications, confirming the results of HW. In all estimates where the cognitive skills variable is included, the initial years of schooling have no significant impact on economic growth.

In Table 1, panel B (1-B), we use the scores for cognitive skills based on the new data source (i.e. the updated version of Altinok *et al.*, 2014) but still restrict the sample of countries to that of HW. We also use the same methodology as in Panel A, thus allowing a comparison of the results between the two datasets. Across columns (1) to (8) of Panel B, coefficient estimates for our "cognitive skills" remain significant. The coefficients are estimated more precisely, as reflected by higher t-statistics, implying that our data are at least as predictive as the data used by HW for the restricted set of countries. The overall effect of cognitive skills on economic growth is however slightly higher in our dataset.[17]

In Table 1, Part C (1-C), we still use our alternative measure of cognitive skills, but now extend our sample from 50 to 84 countries. Most of the newly included countries are from Sub-Saharan Africa and Latin America (see Table A.3 for a full list of countries included in our regressions). In our dataset, similarly to HW, we exclude five countries which can be considered as outliers (Botswana, Gabon, Kenya, Luxembourg and Mauritania).[18] The amplitude of the cognitive skills effect diminishes with the additional countries, but remains significant, even after we control for openness and property rights (col. 7), and include additional controls such as fertility rates and tropical location (col. 8). Another important result relates to the coefficient associated with our measure of years of schooling. While it was insignificant in both panels B and C with the sample of 50 countries, in panel C the magnitude of the coefficient is always positive and it is significant in some specifications, notably in column 6 where we include region fixed effects.

[17] This can be explained by the fact that we do not include in our dataset results from IAEP and results that refer to the end of secondary schools. The bias included in the IAEP survey has been well documented in the literature (see for instance Rotberg, 1990; McLean, 1990; Goldstein, 1993). Moreover, since the survival rates to the last grade of secondary education greatly differ between countries, we prefer not to include results from TIMSS-Advanced in our dataset.

[18] Luxembourg is known as a country which has economic growth mainly based on tax-free policies, so the relationship between cognitive skills and economic growth can be flawed. The remaining African countries are excluded since either we only have one observation (Mauritania) or results are contradictory between assessments (Botswana, Kenya, Gabon).

Thus, the results of Table 1 show a strong positive relationship between cognitive skills and economic growth that remains significant across different specifications in a cross-country estimation. This confirms the findings of HW in a larger sample, including many developing countries. Secondly, the amplitude of the effect is quite similar to that in HW's paper, even after we include more countries in the estimation. Comparing results in column 8 in each panel (1-A; 1-B and 1.C) shows to what extent the amplitude of the effect changes between different indicators of cognitive skills and different samples of countries. In the HW estimates, an increase of one standard deviation of cognitive skills improves growth by about 1.2 percent (column 3). The effect is approximately similar with our data restricted to the HW sample of countries (1.3 percent). Including all countries with available data shows that the estimated effect (1.5 percent) is about 25% higher. While one-third of this increase can be attributed to the difference between our updated cognitive skills measures, the remaining part is due to the expansion of sample, and especially the inclusion of developing countries. The main finding is that for our full sample and with our updated cognitive skills measures, a move of one standard deviation of individual student performance translates into 1.5 percentage points difference in annual growth rates, other things being equal. It is also interesting to measure the level of one standard deviation in terms of score points. Since one standard deviation is equal to 100 points in our scale, this represents approximately the difference of performance between Greece (533 points) and South Korea (628 points). In addition, the difference between Turkey and the remaining OECD countries is approximately equal to 0.5 standard deviation.

Next, in Appendix Table A.4, we present the estimated cognitive-skill coefficients across different samples of observations. While Panel A of Table A.4 reproduces results from HW, in Panel B we use our dataset in order to check for the stability of estimates with a more comprehensive set of countries. Results from Table A.4, Panel A are similar to estimates from HW with only slight differences in some cases[19]. In Table A.4, Part B, we use our new cognitive-skill scores. Also in Panel B the time span is extended to 1960-2012. With additional 30 countries and the time span extended to 2012 the coefficients are in general estimated more precisely.

In the appendix Table A.5, we perform a further robustness analysis that considers alternative aggregation of test scores. Under the assumption of stable test performance over time, row A uses

[19] This may be either due to differences in methodology used in some estimations or the fact that upper secondary schools are excluded from our analysis.

test scores since 1995 that are thought be a product of a higher standard of sampling and quality control; row B restricts the tests in this time span to tests using only lower secondary scores. A drawback of using only the most recent tests is that this assumes the test performance to be quite stable over time, since we relate test performance measured since 1995 to the economic level data for 1960-2010. In order to test that higher past economic growth is not impacting our measured test performance, we restrict the test-score measure used in row C to all tests until 1995. Our results remain quite stable and robust, although the number of countries with available data is reduced from 80 to 46 countries. Rows D to F use test scores individually, while row G uses test scores jointly. While our results are in general similar to the results of HW, an important difference arises when mathematics, reading and science scores are entered jointly (row G). While the results of HW shows that math and science coefficients are significant in our results only the mathematics score remains significant and has the highest point estimate suggesting that mathematics skills may be the most important skill for economic growth.

The above results show a strong positive relationship between cognitive skills and economic growth using cross-sectional variation. Therefore, our results confirm the results of HW, and partially the hypothesis highlighted in Pritchett (2001) where one explanation of the controversies in the lack of significant effect of education on economic growth was quality differences between countries. While the results are robust across various specifications and subsamples, reverse causality and endogeneity bias may potentially be driving the results. Reverse causality would arise if higher economic growth enables countries to develop better education systems that yield higher test performance. The presence of other factors that are correlated with cognitive skills such as cultural factors, institutions, and access to natural resources that affect growth will lead to an endogeneity bias in our estimations. The following section addresses these identification issues.

5. Endogeneity and measurement error

In this section, we address the potential endogeneity of cognitive skills using various identification strategies. First, we provide IV-2SLS estimates for a subset of IVs proposed by HW for which we could obtain data. Second, we report results from an IV-GMM estimation using alternative instruments proposed in the recent literature.

HW propose to use measures of the institutional structure of the school systems as instruments for cognitive skill. Hanushek and Woessmann (2011) show that these factors are associated with international educational production. HW (2012a) provide supporting evidence that these institutional features are plausibly exogenous in the growth regressions. While HW provide a description of the data source for their analysis, we could not obtain all observations concerning instruments. As a result, we use a subset of these instruments and the number of observations in some cases is smaller than that of HW. In Table 2, we report results from models that use several institutional features - external exit exam system, catholic share in 1900, and relative teacher salary - as instruments in regressions.[20] Columns 1, 3, and 5 of Table 2 report results that use data from HW while columns 2, 4 and 6 use our updated data.

Accountability has been shown in the literature to be related to better student achievement (Bishop, 2006), for which the external exit exam serves as a proxy. The first specification in Table 2 uses the share of students in a country who are subject to external exit exams as an instrument for our measure of cognitive skills in the growth regression. Data for this instrument are available for 29 countries in Woessmann *et al.* (2009).[21] While the measure deals with the mid-1990s, exam regimes are relatively stable over time for countries. We include years of schooling as a second instrument for test scores, since this variable is not significant in previous estimations, once tests scores are controlled for. The first stage results confirm a positive association between external exit exams and cognitive skills, but the coefficient is not significant. The relevance of the instruments is tested in the first-stage regression. As a rule of thumb, the F-Statistic of a joint test whether all excluded instruments are significant should be bigger than 10 in case of a single endogenous regressor (Stock, Wright and Yogo, 2002). Years of schooling are significantly associated with test scores in the first stage, and the first-stage F-statistic is higher than the usual minimum level expected (i.e. 10). In the second stage, we find a positive and statistically significant effect of cognitive skills on economic growth and the estimated impact is close to the OLS estimate. In column 2, we report results from the same specification that uses our measure of cognitive skills. While results show a positive and significant effect, the amplitude of the effect is reduced with our alternative data for cognitive skills. The first-stage F value is also low in this case, which may lead to a weak instrument problem. However, results

[20] We also estimated models with other instruments reported by HW. However, data was lacking for a large number of countries, so we do not report these results in the paper. These results are available on request.
[21] We could only obtain data for 29 of the 43 countries as explained in HW.

based on the modification of the limited information maximum likelihood (LIML) estimator by Fuller (1977) provide estimates that are very similar to the 2SLS estimates. Thus, both the 2SLS and the Fuller estimates confirm that schooling-induced differences in cognitive skills are significantly related to economic growth.[22]

The third and fourth columns of Table 2 use teacher salaries relative to per-capita income as an instrument. Following Dolton and Marcenaro-Gutierrez (2011), HW note that this variable serves as a proxy for the overall quality of the teaching force in a cross-country perspective. For both column 3, which uses HW data, and column 4 that uses our data with 10 additional countries, this instrument cannot predict cognitive skills and the first stage F values and Sargan statistic reveal a weak instrument problem. Despite this problem, both the 2SLS and LIML yields positive and significant coefficient estimates for cognitive skills.

The next specification uses the share of Catholics in a country's population in 1900, which is shown to be associated with the share of privately operated schools in current school systems (West and Woessmann, 2010), as an instrument for student achievement. The validity of this instrument stems from the positive association of school choice with student achievement in OECD countries (see the review in Woessmann *et al.*, 2009). The results of the first stage specification that includes the Catholic share in 1970 as a control indicate a positive correlation between cognitive skills and historical Catholic shares. The second stage results in column 5 that includes 50 countries used in the estimation of HW show a positive and significant impact of cognitive skills on growth, with very similar point estimates to HW. Columns 6 reports results using the extended dataset that involves 80 countries. The extended sample leads to a significantly higher first-stage F-statistic, indicating that the instrument has a higher predictive power in this sample. Compared to column 5, while the magnitude of the estimated coefficient for cognitive skills is lower, these results also confirm the positive effect of cognitive skills on growth.

Instruments used in Table 2 are available only for a limited number of developing countries. Several papers use an alternative set of instruments (Islam *et al.*, 2014; Adams and Lim, 2014) that allow IV estimation involving a larger set of countries. In addition to using an alternative set of instruments, we also use GMM estimation instead of standard 2SLS. A key advantage of the

[22] The Sargan test does not reject the overidentifying restrictions of the model.

IV-GMM estimator over the IV-2SLS approach is that the former is more efficient in the presence of heteroscedasticity. IV-GMM is also our preferred method because under the strict assumption of no heteroscedasticity, the IV-GMM is asymptotically no worse than the IV-2SLS estimator (Baum, Schaffer, and Stillman, 2003).

The first set of alternative instruments are (1) disability-adjusted life years lost per 100,000 population (DALY) due to communicable, maternal, perinatal, and nutritional diseases (excluding DALY due to noncommunicable diseases such as cancer, cardiovascular diseases, and injuries which are unlikely to influence school performance) and (2) estimated death rates due to communicable, maternal, perinatal, and nutritional diseases per 100,000 population (EDR). Islam *et al.* (2014) argue that because infectious and parasitic diseases impair the ability to learn, reduce students' attention and concentration in the classroom, and increase student and teaching absenteeism, DALY serve as a good instrument for the quality of learning. DALY are also not likely to be influenced by growth because they are mainly driven by pathogen stress, which is determined by ecology (Guernier *et al.*, 2004). For the same reasons underlying DALY, EDR serves as the second instrument. While these two instruments have a large overlap, Islam *et al.* (2014) explain the advantages of each one over the other and uses them separately in their analysis.

Estimation results using these new instruments are presented in Table 3. While in previous IV estimations, only 50 countries were included, our extensive sample now includes 78 countries, an increase of 60% in the number of countries. We first include DALY as the only instrument (column 1). Results from the first stage indicate an expected (negative) and significant relation with cognitive skills. The F-statistic at 47 is higher than the threshold of 10 and much higher than the F-statistics reported in Table 3. Columns 2 to 5 use as instruments either only EDR, only DALY, or both, and introduce initial years of schooling as an additional control. All of the resulting estimates suggest a positive impact of cognitive skills on growth where the magnitude of estimated coefficients is remarkably robust across specifications and also quite close to the estimate reported by column 6 of Table 2, which uses the extensive set of countries. The Fuller modification has been made for all estimates and does result in quite similar coefficient estimates, showing that this instrument is quite useful in the cognitive skills-economic growth relationship[23]. The Sargan statistic also does not reject the overidentification test.

Columns 6 and 7 distinguish between OECD and non-OECD countries to assess whether the effect of cognitive skills on economic growth differs between developed and developing economies. While cognitive skills have a significant and positive effect on economic growth, the effect is found to be larger for non-OECD countries. We explore the distinction of economic level of countries in more detail in Table 5.

Adams and Lim (2014) argue that the potential effect of government effectiveness on the per capita income of countries is likely to be driven mainly through its mediating effect on the delivery of education. Given the fact that policies that can be more directly associated with government effectiveness tend to be insignificant in standard cross-country growth regressions, and the absence of a robust relationship between public education expenditures and growth (Levine and Renelt, 1992; Sala-i-Martin *et al.*, 2004), the quality of public financial management is unlikely to have a direct effect on economic growth. As a result, the measure of government effectiveness can be considered as a valid instrument for our cognitive skills measure. We use the "Worldwide Governance Indicators" as our government effectiveness measure, which captures perceptions regarding the quality of public services and the quality of the civil service (Kaufmann, Kraay & Mastruzzi, 2011) and serves as a proxy for the quality of educational service delivery. Using the years in which this measure is available (1998, 2000, and annually from 2002 to 2006) we compute a mean score of government effectiveness for the 1988-2006 period. Column 8 uses government effectiveness and DALY as instruments and find that both variables are correlated with cognitive skills in the first stage. The coefficient estimate associated with our cognitive skills variable in the second stage remains quite stable, compared to the estimation where DALY was included as an instrument (see col. 1). However, the Sargan statistic rejects the overidentification test, suggesting that our instruments are no longer valid. Therefore, we only include years of schooling and government effectiveness as instruments (column 9). These two instruments satisfy Sargan test and we obtain a coefficient estimate for cognitive skills that is positive and significant which is quite similar in magnitude to other estimates in Table 4.

---

[23] Fuller's modification of the LIML estimator is more robust than 2SLS in the presence of weak instruments. Moreover, this modification provides better performance in the simulations by Hahn et al. (2004). We set the user-specified constant (Fuller 1977's alpha) to a value of one, but our results are hardly affected if we set alpha to four.

A global comparison between different estimates from Table 1 to Table 3 shows that IV estimate is higher than OLS estimate. In particular, while a move of one standard deviation of individual student performance translates into 1.5 percentage points difference in annual growth rates in OLS estimates (Table 1, column 3), this effect turns out to be higher by about 25% with IV estimates (Table 3). The downward bias observed in OLS estimates may be stemming from measurement issues, especially for low income countries which took part to student assessments tests like PASEC or SACMEQ. In these assessments, the methodology of scaling is less precise than in international student achievement tests like PISA or TIMSS. Another possible explanation relates with bias occurring when we anchor regional student achievement tests with international student achievement tests. Since, the items in each assessments are not exactly similar, it may be possible that the anchoring methodology used in Altinok *et al.* (2014) underestimates the performance of pupils who participated in these regional assessments (PASEC, SACMEQ, LLECE). It is clear however that there are limits to use IV specifications, especially for cross-country regressions and for a limited number of countries. In the meantime, since coefficients are positive and significant in nearly all estimations regardless to the estimation technique used, we can reasonably think that the effect of cognitive skills on economic growth is quite robust.

6. Heterogeneity in the Impact of Cognitive Skills on Economic Growth

Countries place a high priority to investments in education and skills as a key driver of economic growth. The gains from these investments, however, depend on the interactions between skills, technology, and physical capital. For example, investments in skills may result in larger productivity gains in countries where skill supply is scarce compared to countries where skill supply is relatively abundant. Although there are many studies that assess the mean effect of cognitive skills on growth across countries, there has been little research in the literature that addresses the heterogeneity of this relationship. The robustness tests in our analysis in Table A.4 showed that the division of the sample into OECD and non-OECD countries revealed a somewhat higher impact of cognitive skills on economic growth for non-OECD countries. A similar result is reported by HW when these two sets of countries are compared. Castelló-Ciment and Hidalgo-Cabrillana (2012) also provide evidence of differences in the effect of education quality on economic growth across the quality distribution

A second important issue regarding the heterogeneous effects of skills is which types of skills matter most for economic growth. Acemoglu and Zilibotti (2001) shows that a mismatch between supply of skills and the adopted technology leads to low productivity while Hanushek (2013) provides evidence that the impact of high performers on growth differs between OECD and non-OECD countries. Potential differences in the impact of different types of skills on growth has important policy implications since the countries that aim to improve cognitive skills face the choice of targeting improvements across the whole distribution or placing more emphasis on a specific part of the distribution, such as the bottom or the top.

In this section, we aim to extend the existing literature in a number of ways. We first provide further evidence of the heterogeneity of the relationship between cognitive skills and growth, presenting results for various subsamples. Secondly, we conduct an analysis that tests whether the effect of minimum and advanced level of cognitive skills differ between countries. Our third contribution is related to the estimation methodology. The previous literature that considers the heterogeneity of the relationship between cognitive skills and growth do not address endogeneity of cognitive skills. Using a larger sample of countries, we also address the endogeneity issue through a number of alternative instruments. Our analysis yields a rich set of results that indicate significant differences in the impact of cognitive skills on growth and to sheds further light on growth experiences of countries. Also, using a single data set that involves a consistently defined human capital measure and applying the same methodology for estimation provides comparable results across subsamples. This overcomes the challenge of synthesizing results over different studies that use different methodologies and measures of human capital in different country contexts and the resulting uncertainty in the policy arena as to the most effective type of education or skills for growth

6.1. Distinction of different subsamples

In this section, we provide estimates of the effects of cognitive skills on economic growth across different subsamples. We divide the sample into several parts and provide estimates separately by (i) income level of countries, (ii) total factor productivity, and (iii) regions. Higher income countries employ a higher level of capital stock and enjoy higher total factor productivity. Hence, the role of skills in growth for these countries may differ from those of low income countries. There are also marked differences by geography in the growth experiences of countries. The role skills play in these growth experiences has received little attention. For example, our analysis provides results for Arab countries and Sub-Saharan African countries, the continent that could not often be studied separately by previous studies due to data constraints.

The results are presented in Table 4 which is divided into 2 panels. The first panel reports results from OLS regressions (panel A) whereas the second panel (panel B) report results from the IV-GMM estimation. In order to test whether our results are driven by the use of specific instruments, we use different combinations of instruments presented by rows B1 through B6. In all of the IV-GMM estimates initial years of schooling is used as an instrument in combination with one or two other instrumental variables.[24] In Panel B1, we use the government effectiveness as an instrument that proved to be a valid instrument for the whole sample in Table 4. In Panel B2, we use initial school drop-out rate for primary education as an instrument. Since pupils may leave schools because they do not receive a high standard of education, school drop-out rate for primary school may serve as a good instrument for education quality or cognitive skills. Hanushek *et al.* (2008), for example, show in a developing country context that a student is much less likely to remain in school if attending a low-quality school rather than a high-quality school. However, since growth rate of the economy could also impact on drop-out rate, in order to satisfy the exclusion restriction, we use the *initial* level of school dropout as an instrument[25]. A combination of government effectiveness, initial school drop-out, and initial years of schooling is used as our instruments in panel B3. Primarily used by Islam *et al.* (2014), DALY was highly correlated with our cognitive skills variable in the first stage results of the IV estimation (see column 3 of Table 3) and it is included as an instrument in panels B4 and B5. Final specification uses the overall level of income inequality (measured with Gini index) as an instrument. A recent study shows a positive correlation between income inequality and inequality of education (Inter-American Bank, 1999) while Krueger (2012) and Corak (2013) show that countries with more

[24] Previous sections provided evidence for the validity of initial years of schooling as an instrument in the cognitive skills growth relationship. Nevertheless, we have also carried out estimations that does not use initial years of education as an instrument and obtained results that are very similar to those presented in Table 6. Due to space considerations these results are not presented.

[25] Since data availability differs greatly between countries, the year of the initial value of drop-out rate in primary education varies between countries. However, for most countries, the initial year is 1970.

inequality as measured by Gini coefficients have less intergenerational mobility. Overall level of inequality may thus capture disparities along the income distribution in access to education and its quality, hence lead to reductions in cognitive skills. Cingano (2014) provides support for this channel. The study finds that the main mechanism through which inequality affects growth is by undermining education opportunities for children from poor socio-economic backgrounds, lowering social mobility and hampering skills development. While we hypothesize an effect of inequality on growth only through its effect on cognitive skills, inequalities in education and income and growth may be jointly determined. In order to avoid reverse causality, we use the *initial* level of the Gini coefficient for each country as an instrument[26].

In this section, in order to obtain comparable effects in terms of standard deviations, we also standardize the cognitive skills variable in each sub-sample (with a mean equal to 0 and a standard deviation equal to 1). This allows us to directly compare the effect of cognitive skills expressed in terms of standard deviations between each sub-sample. Given the large set of results, we only report the coefficient estimate of the cognitive scores variable, the first-stage F-statistic, and the number of countries included in each subsample in brackets, but do not to present the first stage results.

In column 1 of Table 4, in the first two rows, we reproduce results from Tables 1 and 3 where our cognitive skills variable has a positive and significant impact on economic growth, whether we consider the OLS or the IV estimations. In rows B2 to B6 of the first column IV estimation using different sets of instruments provide coefficient estimates that range between 1.7 and 2.1 and are all larger than the OLS estimate. According to the OLS estimation an increase of one standard deviation in cognitive skills produces an increase in annual economic growth of about 1.5 percentage points. By using the IV estimation technique, the overall effect of cognitive skills is increased by about 25% (1.9%). As explained in Madsen (2014) regarding educational achievement, one reason for the increased effect may be the downward bias due to measurement error.

The results in columns 2 to 8 that distinguish between the economic levels of countries provide important insights. Comparing columns 2 and 3 shows that while the effect of cognitive skills is positive and significant for both low and high income countries, both the OLS and IV

[26] Similarly to drop-out rate, the initial level of Gini coefficient differs between countries. Due to data constraints, the initial year is often around 1980.

results indicate that its amplitude is about 60% higher for the low-income countries. This emphasizes that the promotion of education policies that focus on the quality of education has especially large payoffs in least developed regions. These results are in contrast with Castelló-Ciment et al. (2012) that reports no effect of education quality on economic growth in countries at the bottom end of the quality distribution but a positive effect in countries with higher education quality. However, our results are in line with the results of HW that report a higher impact of cognitive skills for OECD countries. In countries with low levels of education quality, improvements in quality may lead to substantial improvements in productivity of workers. Higher estimated effects of quality on growth in low income countries may be due these productivity gains.

Estimation results by geographical region are presented in columns 4 to 6. IV estimates for each region shows a positive and significant impact of cognitive skills on economic growth. Compared to other regions, we find much larger effect of cognitive skills on economic growth for the Asian countries. It is possible that the early-period growth explosion of East Asia is mainly due to high level of cognitive skills in this region, compared to other regions like Latin America (Hanushek & Woessmann, 2016). The lowest coefficient estimates, on the other hand, are obtained for Latin America. As we focus on regions, with much smaller sample sizes, some of the F statistics are now lower than 10. Only in the case of Latin America, however, they are systematically below this threshold. Hence, we do not have a strong support for a positive effect of cognitive skills on growth for this region.

Besides the standard distinction of economic levels of countries and geographical locations, we also divided the sample into two parts, in the spirit of Nelson and Phelps (1966). It is possible that countries which are far from the technology frontier, i.e. with a low total factor productivity in 1960, will benefit more from higher cognitive skills levels than others countries. To test this possibility, we separate the sample by distinguishing low initial total factor productivity (TFP) countries and high initial TFP countries, using the median level of TFP in 1960 (columns 7 and 8). Results confirm that countries which are far from their technology frontier benefit more from cognitive skills than other countries. Comparing columns 7 and 8, the effect of cognitive skills is doubled for these countries in the standard OLS estimation. When we correct for possible endogeneity, measurement error and omitted variable bias by using the IV GMM estimation technique, the difference between the two groups becomes even larger. Another important finding is that the extent of bias between OLS and IV estimates is the largest for Arab States and Sub-Saharan Africa. This may be due to lower quality of assessment for this region.[27] In conclusion, our cognitive skills variable is quite stable and in most subsamples has a positive and significant impact on economic growth. We find that the magnitude of the effect is higher for the low-income countries and across regions investing in the quality of education appears to be most rewarding for Arab States and Sub-Saharan African countries.

6.2. The ingredients of growth: innovators and/or imitators?

In Tables 1 to 4, our updated cognitive skills indicators were included as mean scores, without any focus on the within country distribution of cognitive skills. However, it is important to question whether the top-performers or the share of pupils reaching a minimum level has the highest impact on economic growth. Our updated dataset provides the possibility of asking the question of how to allocate education resources between the lowest and the highest achievers. There are two main views regarding the channel through which education enhances growth. The first view argues for investing in the top performers who would boost innovation (Nelson and Phelps, 1966; Aghion and Howitt, 1998; Vandenbussche, Aghion, and Meghir, 2006; Galor, 2011) while the alternative view argues for a more egalitarian school system to ensure well-educated masses (Mankiw, Romer & Weil, 1992). Aghion and Cohen (2004) distinguish economies of imitation from economies of innovation. The first group of economies must invest primarily in the school levels supporting the imitation and implementation of new techniques, that is to say, primary and secondary education. This group includes countries with low and middle incomes. In order to encourage economic growth, the second group of countries must contribute to technological innovation and have at their disposal a large mass of skilled labor. This justifies a major investment in higher education supporting economic growth. The developed countries belong to this second group of economies. These alternative views are reflected in different policy goals such as the Bologna Process that aims at developing high

[27] Contrary to other assessments where modern psychometric procedures were included, the PASEC assessment had no Rasch scaling of scores which may reduce survey quality and explain why the estimated IV coefficient is higher than the one found with OLS technique. See Wagner (2011).

quality standards in the education sectors for European countries and "Education 2030" objective that aims to provide the majority of pupils with a minimum level in both mathematics and reading (UNESCO, 2015).

Altinok *et al.* (2014) distinguishes between "advanced level students" and "minimum level students" that allows us to test the effects of attaining minimum skill levels and reaching advanced level skills on economic growth. In this dataset, the minimum level threshold is 400 test-score points in the adjusted international scale, while the advanced level threshold is defined as 600 points. The minimum level can be benchmarked to level 1 of PISA assessment where students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined (OECD, 2013). These students may be able to perform mathematical tasks quickly, such as reading a single value from a well-labeled table. The international median of this share of students is 73%, ranging from Malawi with 20% to Korea and Chinese Taipei with 95%. The "advanced level", on the other hand, is approximately anchored to level 5 of the PISA scale, where students can develop and work with models for complex situations, identifying constraints and specifying assumptions (OECD, 2013). They can select, compare, and evaluate appropriate problem-solving strategies for working with complex problems related to these models. The international median of this share of students is 11% in our sample, ranging from less than 0.7% (El Salvador) to 63% (Korea).

The correlation rate between the share of pupils reaching advanced and minimum levels is not perfect, although it is still quite high (r = 0.82), indicating that these differences are not fully comparable to a standard deviation. However, the correlation between the mean score of cognitive skills and the share of pupils reaching the minimum level is higher (r = 0.96) than its correlation with the advanced level (r = 0.87). Figure 2 presents the possible relationship between the shares of pupils reaching each level, suggesting the existence of an inverted U-shaped relationship. It is indeed possible to achieve relatively high median performance, both with a relatively equitable spread (Korea, Finland) and a relatively unequal spread (Belgium, Switzerland). The same is true for the developing countries with low average performance, as shown by the contrast between Mauritius' higher inequality and Thailand's much greater equality between low and high achievers (Figure 2).

In Table 5, we present the estimation results for the relationship between the share of pupil reaching the advanced level ("Advanced Level") or the minimum level ("Minimum Level"). Both distributional measures of cognitive skills are significantly related to economic growth, when entered either individually or jointly (columns 1-3). From the estimates in column 3 with the OLS technique, a 10 percentage point increase in the share of students reaching the minimum level is associated with 0.4 percentage points higher annual growth, while a ten percentage point increase in the share of advanced level students is associated with 0.2 percentage points higher annual growth. Compared to results of HW, while the magnitude of the coefficient estimate for the minimum level is quite similar, the effect of top-performing students is lower in our larger sample. Note, however, that it may be less difficult to increase the minimum share than to boost the advanced share, as suggested by the fact that the standard deviation of the minimum share is almost twice that of the advanced share across countries (0.26 against 0.17), hence it may be more appropriate to provide a comparison in terms of standard deviations. Expressed in standard deviations, increasing each share by roughly half a standard deviation (8 percentage points for "advanced level" performing share and 13 percentage points for "minimum level" performing share) yields a quite similar growth effect of roughly 0.3 percentage points

The inclusion of measures of economic institutions, fertility, and tropical geography do not change the overall effect of basic performers (column 5), but results for advanced performers in columns 4 and 6 show that only the share of basic performers remains significant, suggesting that a significant part of the effect of advanced performers comes from improved institutions.[28] Some other potential differences between regions may also explain why the advanced performers do not have a positive effect on economic growth. When we include regions fixed effects (column 7), we find a positive effect of advanced performers. Columns 8 and 9 split the sample into two groups based on the economic level of countries. These results show that the effect of the basic literacy share or the top-performers share appears to be stronger for developing countries.

[28] Since the number of countries decrease from 80 to 68 when we include institutional controls, it may be possible that the coefficient is no longer significant due to the lower number of countries. In order to test for this possibility, we ran the estimation without these controls with restricted sample of countries (68 countries). The results (not shown) confirm that the share of pupils reaching the advanced level has a positive and significant effect on economic growth. We also tried the same exercise for regressions with both the share of pupils reaching the minimum and advanced level. Only the coefficient associated with the former remain positive and significant, while the latter loses its significance.

Above results may suffer from endogeneity bias. For example, a country with high economic growth may be more able than others to invest in high quality universities and boost the share of pupils reaching the advanced level. Higher economic growth, on the other hand, may permit developing countries to invest in primary and secondary schools, allowing more pupils to improve their cognitive skills. In previous research, while HW present OLS estimates for the effect of basic and advanced performers on economic growth, their analysis do not address endogeneity issues. In the next section, we address this endogeneity issue and also explore the effects of basic and advanced performers in greater depth by using different subsamples. In Table 6, we conduct an analysis similar to Table 4 for both advanced and basic performers. While in panel A standard OLS estimations are presented, Panel B provides IV-GMM estimates that correct for measurement error and endogeneity. In all estimations, the both the top performers share and the basic literacy share are included. Given these two endogenous variables, in IV-GMM estimations, we need at least two instruments for identification. To test robustness of our results, we provide four separate IV-GMM estimates in panels B1-B4 where each panel uses a different set of instruments. Governance effectiveness (GE) and DALY, which are powerful predictors for cognitive skills, are common instruments across these panels. In addition to these instruments, panel B1 uses initial years of education, panel B2 uses initial years of education and drop out of primary education, panel B3 uses survival rate to the last grade of primary education, panel B4 uses initial years of schooling and initial level of Gini index as instruments. Controlling for endogeneity, IV GMM estimates for the whole sample (column 1) provide quite stable coefficients for both advanced and minimum levels indicating a positive and significant effect of basic performers but an insignificant effect for advanced performers.

Above results for the overall sample may be hiding heterogeneity in the impact of skills on growth. The basic performers may be essential component of growth in developing countries as imitators while advanced performers may be crucial for innovation that spurs growth in developed countries. In order to test this hypothesis, countries are separated according to their economic level in columns 2 and 3. In all estimations for high income countries in column 2, regardless of the estimation technique or the instruments used, results indicate that an advanced level of cognitive skills is an important factor of economic growth for high-income countries. The coefficient estimate for minimum performers, however, is marginally significant only in two of four IV panels and the magnitude of the coefficient is much lower than that for advanced performers. For low income countries in column 3, we get the opposite result that minimum performers enhance growth more than advanced performers. This suggests that developing countries which focused on the provision of mass education rather than providing subsidies for elites grew faster than other developing countries.

Since our dataset includes a significant number of developing countries, we provide more detailed analysis by distinguishing between three regions (Arab states and Sub-Saharan Africa (SSA); Asia; and Latin America). The results are presented in columns 4 to 6. While the share of students with a minimum level of cognitive skills have the greatest impact on economic growth in Arab states and SSA, we find the exact opposite effect for Latin American countries, confirming the previous results of Hanushek & Woessmann (2012b). In Arab states and SSA countries, in all estimations, the share of top performers has a negative effect on economic growth. This result should be viewed with caution, because the share of top performers in most countries of this region is very low. The results for Latin America should be treated with caution too since in most estimations, we face a weak instrument problem. In panel B4, where the instruments appear to be valid, the positive effect of the share of top performers is positive and significant. Another important result concerns Asian countries. While in the OLS estimation, both measures of cognitive skills have a positive and significant effect on economic growth, only the share of pupils reaching the basic level seems to positively enhance economic growth. The main reason for the lack of significance of the share of top performers in IV estimates is the potential reverse causality: countries with higher economic growth may be investing more on the education of pupils with high skills. Although results are not robust in all sub-samples, they provide some interesting information about the importance of investing not only for the whole population, but also for specific sub-populations, which differ between the economic level of countries. These results should however be also tested in a panel dataset in order to test to what extent *in increase* of the share of advanced (minimum) level students enhances economic growth.

8. Conclusion

Among all the explanations for economic growth, one that is generally accepted concerns the level of human capital. The seeming obviousness of the idea, however, has met with inconsistencies in the existing literature. The most robust macro-economic analyses reveal contradictions in the relationship between education and growth. Pritchett (2001) showed that very often the impact of education on growth is negative and significant. However, the majority of these studies have ignored the qualitative dimension of human capital, recognizing only the purely quantitative indicators.

The use of national or international achievement tests in mathematics and sciences fills this gap in qualitative measurement. Hanushek and Kimko (2000), Barro (2001) and Hanushek and Woessmann (2012a) used qualitative variables, but did not exploit all of the international assessments or all the countries surveyed (respectively 36, 43 and 50 countries included in their samples). In this research study, we used an updated dataset on cognitive skills for a larger number of countries. Consequently, our sample includes more developing countries than the previous studies and the timespan is larger since we include the most recent assessments.

The estimate of the relationship between education and growth reveals the positive role played by cognitive skills: when the *qualitative* dimension of education is taken into account, we find positive and significant effects of cognitive skills on economic growth. Whatever the specifications selected (with or without variables of control), the effect of cognitive skills is always positive and significant on the growth rate of the economy.

An important problem had to be addressed: as economic growth also plays a part in explaining the quality of education systems, it is necessary to make use of a simultaneous equations model. By using an instrument variable estimation technique, the effect of education on growth is maintained. Education thus has a direct and causal effect on the growth rate of an economy. The amplitude of the effect varies slightly between estimations. However, it can be agreed that an increase of one standard deviation of cognitive skills may increase economic growth by about 1 percentage point, which is approximately the same as the result found in Hanushek and Woessmann (2012a).

The main advantage of using a larger dataset is probably the possibility of distinguishing between different subsamples. We look for potential differences in the amplitude of the effect of cognitive skills on economic growth between high-income and low-income countries, but also between different regions. While the overall effect of cognitive skills is higher for developing countries, it remains positive and significant in all subsamples. The highest amplitude is found for Sub-Saharan African countries, where an increase of one standard deviation of cognitive skills improves economic growth by about 2.5 percentage points.

Besides the standard mean value of cognitive skills, the distribution of education quality within each country may also be a potential factor of economic growth. Instead of simply relying on standard deviations of cognitive skills and economic growth, our study includes the share of pupils reaching the minimum level or the advanced level in mathematics/sciences as potential factors explaining the economic growth of countries. Both measures have a positive and significant effect on economic growth, although the share of minimum performers tends to be more robust in subsamples. Another important finding relates to the fact that it is the share of top performers which leads to more economic growth in developed countries, whereas the share of pupils reaching the minimum level appears to be crucial in countries from Arab states and Sub-Saharan African countries. The possibility of distinguishing countries according to their economic level or the region in which they are included is therefore an important improvement which needs to be continued in future research.

References

Acemoglu, D. and Zilibotti, F. 2001. Productivity Differences. Quarterly Journal of Economics. 116(2): 563-606.

Adams-Kane, J., & Lim, J. J. 2014. Institutional quality mediates the effect of human capital on economic performance. World Bank Policy Research Working Paper 6792, Washington.

Aghion P. and Cohen E., 2004. Éducation et Croissance. La Documentation française, Paris, 2004.

Aghion P. and Howitt P., 1998. Endogeneous Growth Theory. MIT Press, Cambridge.

Altinok, N., Diebolt, C. & de Meulemeester, J.-L. 2014. A New International Database on Education Quality: 1960-2010. Applied Economics 46 (11), 1212-1247.

Altinok, N., & Murseli, H. 2007. International database on Human Capital Quality. Economics Letters 96 (2), 237-244.

Barro, R. 1991. Economic growth in a cross-section of countries. Quarterly Journal of Economics 106 (2), 407–443.

Barro, R.J. 2001. Education and Economic Growth. in Helliwell, J.F. (Ed.), The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being (pp. 14-41). Paris: OECD Press.

Barro, R.J., & Lee, J.W. 1993. International Comparisons of Educational Attainment. Journal of Monetary Economics 32, 363-394.

Barro, R.J., & Lee, J.W. 2010. New Data Set of Educational Attainment in the World: 1950-2010, NBER Working Paper 15902.

Barro, R.J., & Lee, J.W. 2013. A new data set of educational attainment in the world, 1950–2010. Journal of development economics 104, 184-198.

Baum, C.F., M.E. Schaffer, & Stillman, S. 2003. Instrumental Variables and GMM: Estimation and Testing. The Stata Journal 3, , 1–31.

Bils, M., & Klenow, P. J. 2000. Does schooling cause growth?. American economic review, 1160-1183.

Bishop, J.H. 2006. Drinking from the fountain of knowledge: Student incentive to study and learn—Externalities, information problems, and peer pressure. In E. A. Hanushek & F. Welch (Eds.), Handbook of the Economics of Education (pp. 909–944). Amsterdam: North Holland.

Breton, T.R. 2011. The quality vs. the quantity of schooling: What drives economic growth?. Economics of Education Review 30 (4), 765-773.

Castelló-Climent, Amparo and Ana Hidalgo-Cabrillana 2012. The role of educational quality and quantity in the process of economic development, Economics of Education Review, vol. 31(4), pages 391-409.

Chinapah, V. 2003. Monitoring Learning Achievement (MLA) Project in Africa. Association for the Development of Education in Africa (ADEA). Paris: International Institute for Educational Planning.

Coulombe, S., & Tremblay, J.-F. 2006. Literacy and Growth. Topics in Macroeconomics 6(2). Article 4.

Cohen, D., & Soto, M. 2007. Growth and human capital: good data, good results. Journal of economic growth 12 (1), 51-76.

Corak, Miles, 2013. Income Inequality, Equality of Opportunity, and Intergenerational Mobility. Journal of Economic Perspectives, 27, no. 3, 79–102.

Dolton, P., & Marcenaro-Gutierrez, O.D. 2011. If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance.Economic Policy 26 (65), 5-55.

Sala-i-Martin, X., Doppelhofer, G., & Miller, R.I. 2004. Determinants of long-term growth: A Bayesian Averaging of Classical Estimates (BACE) approach. American Economic Review 94 (4), 813–835.

Durlauf, S.N., Johnson, P.A., & Temple, J.R. 2005. Growth econometrics. Handbook of economic growth 1, 555-677.

Easterly, W., & Levine, R. 1997. Africa's growth tragedy: policies and ethnic divisions. The Quarterly Journal of Economics, 1203-1250.

Fuller, W.A. 1977. Some properties of a modification of the limited information estimator. Econometrica 45(4), 939–954.

Glaeser, E.L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. 2004. Do institutions cause growth?. Journal of economic Growth 9 (3), 271-303.

Glewwe, P., Maïga, E., & Zheng, H. 2014. The Contribution of Education to Economic Growth: A Review of the Evidence, with Special Attention and an Application to Sub-Saharan Africa. World Development 59, 379-393.

Goldstein, H. 1993. Interpreting international comparisons of student achievement. Report Prepared for UNESCO, Paris.

Guernier, V., Hochberg, M. E., & Guégan, J. F. 2004. Ecology drives the worldwide distribution of human diseases. PLoS biology 2 (6), e141.

Hahn, J., Hausman, J. A., & Kuersteiner, G. 2004. Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. Econometrics Journal, 7(1), 272–306.

Hanushek, E.A., & Kimko, D.D. 2000. Schooling, labor-force quality, and the growth of nations. American Economic Review, 1184-1208.

Hanushek, E.A. 2002. Publicly provided education. In A. J. Auerbach & M. Feldstein (Eds.), Handbook of Public Economics. (pp. 2045–2141). Amsterdam: Elsevier.

Hanushek, E., Lavy, V. & Hitomi, K. 2006. Do students care about school quality? Determinants of dropout behavior in developing countries, Journal of Human Capital, 2008, vol. 2, no. 1, p. 69-105.

Hanushek, E.A., & Woessmann, L. 2011. The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), Handbook of the Economics of Education, Vol. 3 (pp. 89–200). Amsterdam: North Holland.

Hanushek, E.A., & Woessmann, L. 2012a. Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, And Causation. Journal of Economic Growth 17(4), 267-321.

Hanushek, E. A., & Woessmann, L. 2012b. Schooling, educational achievement, and the Latin American growth puzzle. Journal of Development Economics 99 (2), 497-512.

Hanushek, E. A., & Woessmann, L. 2015. *The Knowledge Capital of Nations. Education and the Economics of Growth*. The MIT Press, Cambridge, MA.

Hanushek, E. A., & Woessmann, L. 2016. Knowledge capital, growth, and the East Asian miracle. *Science*, 351(6271), 344-345.

Hanushek, E.A. 2013. Economic growth in developing countries: The role of human capital. Economics of Education Review 37, 204–212.

Hanushek, E.A., Schwerdt, G., Wiederhold, S., & Woessmann, L.. 2013. *Returns to Skills around the World: Evidence from PIAAC. European Economic Review 73. 103-130.*

Hanushek, E.A., Schwerdt, G., Wiederhold, S., Woessmann, L. 2015. Returns to Skills around the World: Evidence from PIAAC. European Economic Review 73, 103-130.

Heston, A., Summers, R., & Aten, B. 2002. Penn world table. Center for International Comparisons at the University of Pennsylvania.

Inter-American Development Bank. 1999. Facing up to Inequality in Latin America: Economic and Social Progress in Latin America, 1998-99 Report. Johns Hopkins University Press.

Islam, R., Ang, J. B., & Madsen, J. B. 2014. Quality-Adjusted Human Capital and Productivity Growth. Economic Inquiry 52 (2), 757-777.

Kaufmann, D., Kraay, A., & Mastruzzi, M. 2011. The worldwide governance indicators: methodology and analytical issues. Hague Journal on the Rule of Law 3 (02), 220-246.

Keeves, J.P.1992. *The IEA Science Study III: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon Press.

Krueger, Alan, 2012. ‘‘The Rise and Consequences of Inequality in the United States,’’ Speech at the Center for American Progress, Washington, DC, January 12, 2012.

Lee, J. W., & Barro, R. J. 2001. Schooling quality in a cross–section of countries. Economica 68 (272), 465-488.

Levine, R., & Renelt, D. 1992. A sensitivity analysis of cross-country growth regressions. The American economic review, 942-963.

Mankiw, N. G., Romer, D., & Weil, D. N. 1992. A contribution to the empirics of economic growth (No. w3541). National Bureau of Economic Research.

McLean, L. 1990. Let’s call a halt to pointless testing, *Education Canada*, 30(3), 10-13.

Nelson R. and Phelps E., 1966. Investment in Humans, Technological Diffusion and Economic Growth. American Economic Review, 61, 69-75.

OECD. 2013. *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume 1),* PISA, OECD Publishing.

Cingano, F. (2014), “Trends in Income Inequality and its Impact on Economic Growth”, OECD Social, Employment and Migration Working Papers, No. 163, OECD Publishing.

Pritchett, L. 2001. Where has all the education gone?. The World Bank Economic Review 15 (3), 367-391.

Ramirez, F. O., Luo, X., Schofer, E., & Meyer, J. W. 2006. Student achievement and national economic growth. American Journal of Education 113 (1), 1-29.

Robitaille, D.F. & Garden, R.A. (Eds.).1989. *The IEA Study of Mathematics II: Context and Outcomes of School Mathematics*. Oxford: Pergamon Press.

Rotberg, I.C. (1990). I never promised you first place. *Phi Delta Kappan*, 72, 296-303.

Sachs, J. D. 2001. Tropical underdevelopment (No. w8119). National Bureau of Economic Research.

Sala-i-Martin, X. X. 1997. I just ran four million regressions (No. w6252). National Bureau of Economic Research.

Stock, J.H., Wright, J.H., and Yogo, M. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics*, 20(4), 518-29.

UNESCO. 2000. With Africa for Africa. Toward quality education for all. UNESCO. Paris: Human Sciences Research Council.

UNESCO. 2015. *Education For All 2000-2015: Achievements and Challenges*. Paris, UNESCO Publishing.

Wagner, D. 2011. Smaller, Quicker, Cheaper. Improving Learning Assessments for Developing Countries. UNESCO: International Institute of Educational Planning. Paris.

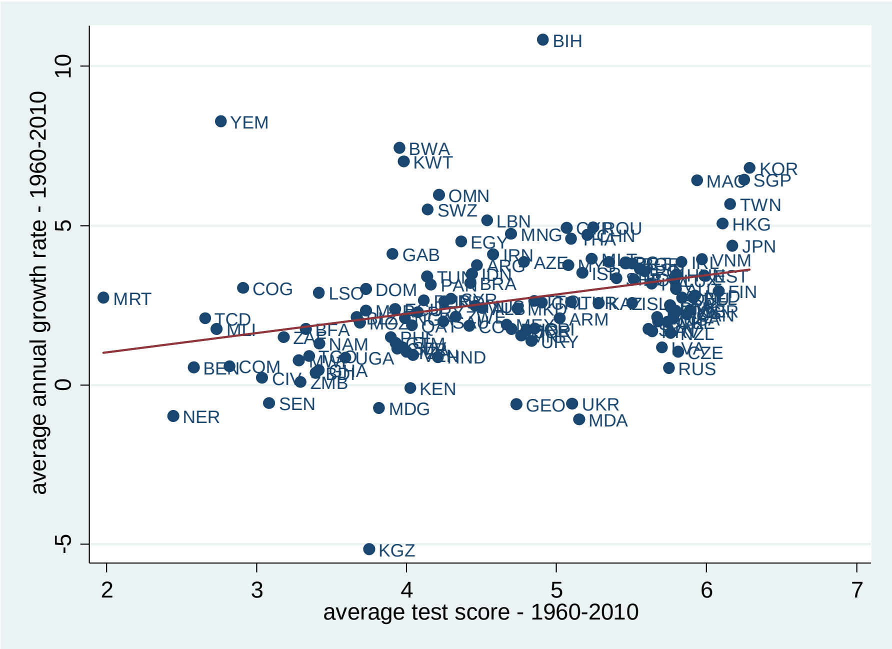
West, M. R., & Woessmann, L. 2010. ‘Every Catholic Child in a Catholic School’: Historical Resistance to State Schooling, Contemporary Private Competition and Student Achievement across Countries. The Economic Journal 120(546), F229-F255.

Woessmann, L. 2003. Central exit exams and student achievement: International evidence. In P.E. Peterson & M. R. West (Eds.), No child left behind? The politics and practice of school accountability (pp. 292–323). Washington D.C.: Brookings Institution Press.

Woessmann, L., Luedemann, E., Schuetz, G., & West, M. R. 2009. School accountability, autonomy,and choice around the world. Cheltenham, UK: Edward Elgar.

Figures and Tables

Figure 1. Relationship between economic growth and test scores around the World.



Note: 125 countries are included. The correlation rate is equal to 0.30.

Figure 2. Relationship between the proportion of pupils reaching the minimum level and the advanced level

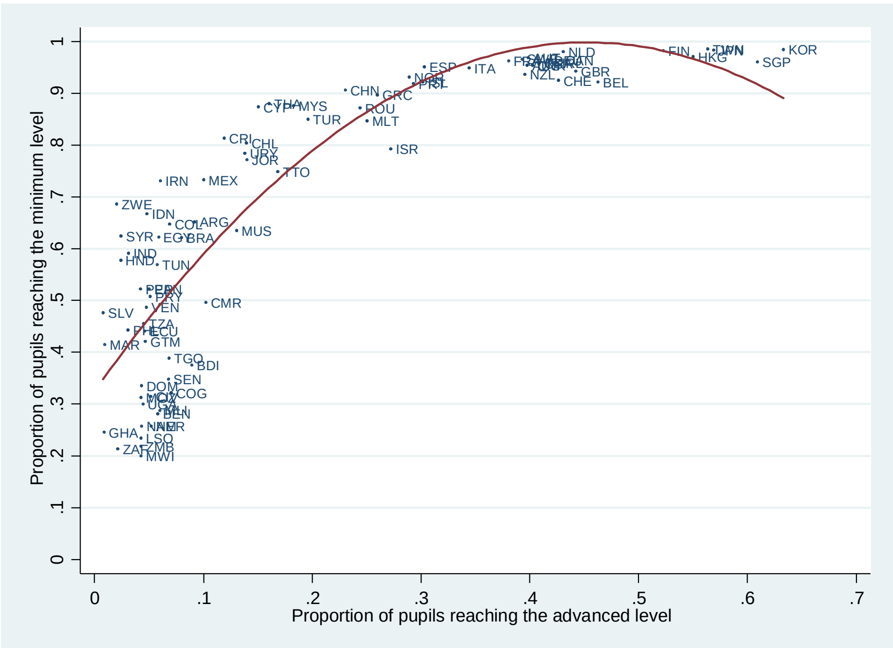


Table 1. Standard estimates of the effect of cognitive skills on economic growth

	(1)	(2)	(3)	(4)[(a)]	(5) [(b)]	(6) [(c)]	(7) [(d)]	(8) [(e)]	(9) [(f)]
(A) Data from Hanushek and Woessmann (2012a), sample from Hanushek and Woessmann (2012a)									
Cognitive skills		1.238	1.199	1.224	1.102	1.006	0.853	0.594	1.191
		(8.62)	(7.38)	(6.88)	(8.13)	(3.33)	(5.02)	(5.18)	(6.04)
Years of schooling 1960	0.408		0.050	0.014	0.064	0.070	-0.003	-0.014	-0.067
	(4.41)		(0.79)	(0.18)	(0.76)	(0.79)	(0.04)	(0.23)	(0.89)
GDP pc 1960	-0.399	-0.294	-0.321	-0.300	-0.317	-0.270	-0.334	-0.315	-0.975
	(4.85)	(9.21)	(8.49)	(7.48)	(5.74)	(5.10)	(7.18)	(6.80)	(4.38)
(B) Data from updated Altinok et al. (2014), Sample from Hanushek and Woessmann (2012a)									
Cognitive skills		1.312	1.316	1.383	1.382	1.185	1.023	0.669	1.360
		(8.86)	(7.24)	(6.84)	(11.13)	(4.44)	(4.57)	(3.56)	(6.44)
Years of schooling 1960	0.408		-0.004	-0.062	-0.045	0.006	-0.038	-0.018	-0.079
	(4.41)		(0.06)	(0.76)	(0.69)	(0.07)	(0.52)	(0.27)	(1.11)
GDP pc 1960	-0.399	-0.319	-0.317	-0.294	-0.301	-0.243	-0.327	-0.314	-1.171
	(4.85)	(9.41)	(8.18)	(7.30)	(6.29)	(4.19)	(6.73)	(6.49)	(6.50)
(C) Data from updated Altinok et al. (2014), Sample from updated Altinok et al. (2014)									
Cognitive skills		1.629	1.510	1.382	1.587	1.101	1.531	0.720	1.610
		(13.14)	(10.50)	(8.36)	(10.40)	(3.96)	(5.66)	(2.72)	(9.92)
Years of schooling 1960	0.465		0.115	0.168	0.099	0.149	0.055	0.003	0.036
	(4.90)		(1.83)	(2.51)	(1.29)	(2.28)	(0.83)	(0.06)	(0.54)
GDP pc 1960	-0.220	-0.270	-0.310	-0.315	-0.294	-0.314	-0.281	-0.308	-1.073
	(3.82)	(9.43)	(8.21)	(8.63)	(7.29)	(7.18)	(6.37)	(6.84)	(7.41)
(A) Observations	50	50	50	50	52	50	47	45	50
(B) Observations	50	50	50	50	52	50	47	45	50
(C) Observations	84	84	80	80	85	80	68	68	80
(A) R-squared (adj.)	0.313	0.753	0.756	0.754		0.778	0.800	0.803	0.667
(B) R-squared (adj.)	0.313	0.792	0.792	0.794		0.823	0.820	0.791	0.760
(C) R-squared (adj.)	0.232	0.719	0.729	0.739		0.756	0.714	0.750	0.670

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960-2000 for sample from Hanushek and Woessmann (HW) (2012a), 1960-2010 for sample from updated Altinok et al. (2014). Regressions include a constant. Test scores are average of math and science, primary through end of secondary school (for HW data) or through lower-secondary school (for Altinok et al. data), all years. Absolute t-statistics in parentheses

[(a)] Mean years of schooling refers to the average between 1960 and 2000 (HW data), 2010 (ADM data).

[(b)]Robust regression including the two outliers of Botswana and Nigeria (with rreg robust estimation implemented in Stata).

[(c)] Specification includes dummies for the eight world regions taken in HW.

[(d)] Specification includes additional controls for openness and property rights

[(e)] Specification includes additional controls for openness, property rights, fertility, and tropical location.

[(f)]GDP per capita 1960 measured in logs

Table 2. From schooling institutions to education quality to economic growth: instrumental variables estimates (HW's instruments)

	(1) [a]	(2) [a]	(3) [a]	(4) [b]	(5) [b]	(6) [b]
Second stage						
2SLS						
Cognitive skills	1.155	1.467	1.137	1.860	1.473	1.917
	(7.47)	(3.77)	(5.68)	(4.95)	(7.71)	(8.74)
Catholic share in 1970					0.061	-0.289
					(0.21)	(1.28)
Fuller(1) modification of LIML	1.155	1.452	1.221	1.841	1.485	1.907
Cognitive skills	(7.66)	(4.04)	(3.66)	(5.20)	(7.45)	(8.91)
First stage (dependent variable: Cognitive skills)						
External exit exam system	0.225	0.103				
	(1.28)	(0.60)				
Initial years of schooling	0.359	0.143	0.282	0.170	0.317	0.244
	(5.82)	(2.99)	(3.53)	(2.84)	(4.38)	(5.23)
Catholic share in 1900					2.666	2.160
					(1.33)	(3.37)
Relative teacher salary			-0.002	-0.002		
			(1.14)	(0.90)		
Catholic share in 1970					-3.408	-2.350
					(1.63)	(3.26)
No. of countries	29	30	37	47	50	80
Centered R^2	0.720	0.770	0.439	0.677	0.731	0.701
First-stage F-statistic	17.00	5.70	6.60	4.58	9.61	21.86
Sargan statistic	0.030	0.179	1.111	0.941	1.540	0.017
p-value	(0.864)	(0.673)	(0.292)	(0.332)	(0.215)	(0.900)
Durbin-Wu-Haussman X^2 test	0.022	0.272	0.253	1.448	2.499	5.329
p-value	(0.883)	(0.602)	(0.615)	(0.229)	(0.114)	(0.021)
Cognitive skills measure						
Hanushek and Woessmann data	Yes	No	Yes	No	Yes	No
Altinok et al. data	No	Yes	No	Yes	No	Yes

Notes: Dependent variable (of the second stage): average annual growth rate in GDP per capita, 1960-2000. Control variables: Initial per capita and a constant. Test score are average of math and science, primary through secondary school, all years. t-statistics in parentheses unless otherwise noted.

(a) Data regarding cognitive skills is from Hanushek and Woessmann (2012a) dataset.

(b) Data regarding cognitive skills is from updated Altinok *et al.* (2014) dataset.

Table 3. From schooling institutions to education quality to economic growth: instrumental variables estimates (part 2)

	(1)	(2)	(3)	(4)	(5)	(6) [(a)]	(7) [(b)]	(8)	(9)
Second stage									
GMM									
Cognitive skills	1.913	1.898	2.012	1.986	1.938	1.866	2.203	1.897	1.875
	(9.74)	(9.09)	(5.47)	(5.06)	(11.03)	(5.82)	(11.25)	(11.02)	(10.45)
Fuller(1) modification of LIML	1.905	1.889	2.090	2.061	1.927	1.886	2.165	1.900	1.864
Cognitive skills	(9.87)	(9.24)	(5.60)	(5.19)	(11.08)	(5.99)	(10.97)	(11.06)	(10.51)
First stage (dependent variable: Cognitive skills)									
DALY	-0.323		-1.556	-1.408	-0.272	-3.089	-0.238	-0.237	
	(6.27)		(5.14)	(4.48)	(6.19)	(3.24)	(5.11)	(5.09)	
Initial years of schooling				0.046	0.146	0.003	0.175		0.146
				(1.32)	(3.60)	(0.09)	(3.13)		(3.37)
Early Death Rates (EDR)		-10.173	42.883	38.290					
		(5.56)	(4.24)	(3.62)					
Governance effectiveness								0.514	0.637
								(5.69)	(7.30)
No. of countries	78	78	78	78	78	27	51	77	79
Centered R^2	0.668	0.670	0.523	0.524	0.664	0.759	0.616	0.676	0.708
First-stage F-statistic	47.00	30.91	39.43	37.47	35.55	5.31	26.41	50.71	42.62
Sargan statistic			1.764	1.932	0.078	0.428	0.433	0.035	0.219
p-value			(0.184)	(0.165)	(0.780)	(0.513)	(0.510)	(0.852)	(0.640)
Durbin-Wu-Haussman X^2 test	3.900	2.960	0.666	0.497	7.089	0.580	7.139	5.878	3.567
p-value	(0.048)	(0.085)	(0.414)	(0.481)	(0.008)	(0.447)	(0.008)	(0.015)	(0.058)

Notes: Dependent variable (of the second stage): average annual growth rate in GDP per capita, 1960-2010. Control variables: Initial per capita and a constant. Test score are average of math and science, primary through lower secondary school, all years. t-statistics in parentheses unless otherwise noted. Data relative to cognitive skills is from updated Altinok *et al.* (2014) dataset.

[(a)]Sample of OECD countries.

[(b)] Sample of non-OECD countries

Table 4. Sensitivity of estimated effects of cognitive skills to the economic level of countries and regions

	(1) All countries	(2) High Income Countries[a]	(3) Low Income Countries[a]	(4) Arab States & Sub-Saharan Africa	(5) Asian Countries	(6) Latin American Countries	(7) High TFP countries	(8) Low TFP countries
A- OLS								
Cognitive skills	1.510	0.899	1.600	0.892	1.710	-0.062	0.917	1.836
	(10.50)	(7.68)	(8.18)	(2.96)	(16.22)	(0.23)	(7.83)	(12.68)
Adj. R^2 (Observations)	0.729 (80)	0.717 (40)	0.765 (40)	0.424 (25)	0.958 (14)	0.524 (17)	0.627 (36)	0.858 (36)
B1- IV-GMM								
Cognitive skills	1.875	1.306	1.833	2.500	1.236	0.542	0.956	2.483
	(10.45)	(8.12)	(10.16)	(2.26)	(5.30)	(1.71)	(5.55)	(10.15)
F statistic (observations)	42.62 (79)	29.46 (39)	19.33 (40)	1.56 (25)	13.27 (14)	7.06 (17)	16.61 (35)	16.49 (36)
B2- IV-GMM								
Cognitive skills	2.051	1.225	2.074	1.886	1.650	1.497	0.740	2.214
	(9.80)	(7.11)	(9.87)	(4.20)	(6.19)	(2.43)	(2.33)	(9.56)
F statistic (observations)	27.51 (74)	13.76 (35)	9.45 (35)	6.87 (25)	4.97 (11)	1.97 (17)	5.16 (30)	9.75 (36)
B3- IV-GMM								
Cognitive skills	1.994	1.254	2.064	1.965	1.743	0.423	0.932	2.410
	(10.12)	(7.70)	(10.69)	(4.57)	(7.42)	(2.08)	(4.08)	(10.89)
F statistic (observations)	27.80 (73)	21.23 (34)	11.17 (39)	6.28 (25)	4.79 (11)	4.36 (17)	12.70 (29)	13.52 (36)
B4- IV-GMM								
Cognitive skills	1.938	1.258	1.963	1.843	1.630	1.128	0.839	2.352
	(11.03)	(6.91)	(12.02)	(5.43)	(8.37)	(2.23)	(5.57)	(11.88)
F statistic (observations)	35.55 (78)	9.14 (39)	16.66 (39)	11.82 (25)	6.29 (12)	4.02 (17)	40.62 (34)	14.37 (36)
B5- IV-GMM								
Cognitive skills	1.912	1.209	1.966	1.883	1.635	0.534	0.821	2.439
	(11.38)	(8.29)	(12.05)	(5.79)	(9.74)	(1.78)	(6.78)	(12.22)
F statistic (observations)	37.26 (77)	8.79 (38)	16.95 (39)	8.50 (25)	24.44 (12)	4.24 (17)	64.67 (33)	13.30 (36)
B6- IV-GMM								
Cognitive skills	1.712	0.954	1.779	0.773	0.912	1.476	1.085	2.073
	(10.95)	(8.17)	(10.84)	(1.93)	(3.28)	(2.27)	(6.49)	(10.07)
F statistic (observations)	33.24 (78)	29.37 (38)	13.45 (40)	9.40 (25)	9.84 (14)	2.04 (17)	21.69 (36)	11.43 (34)

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960-2010 for sample from updated Altinok *et al.* (2014). Regressions include a constant. Test scores are average of math and science, primary through lower secondary school, all years. Absolute t-statistics in parentheses. Each panel with IV estimations includes different instruments but always initial years of schooling (hereafter Yrs). B1: Yrs + governance effectiveness. B2: Yrs + drop out of primary education. B3: Yrs + governance effectiveness + drop out of primary education B4: Yrs + DALY. B5: Yrs + governance effectiveness + DALY. B6: Yrs + initial level of Gini index

[a] Countries above/below sample median of GDP per capita 1960

Table 5. Estimated effects of advanced and minimum level of cognitive skills

	(1)	(2)	(3)	(4)[a]	(5)[a]	(6)[a]	(7)[b]	(8)[c]	(9)[d]
Advanced level	6.070		4.186	1.272		0.381	3.064	1.829	2.346
	(9.25)		(5.64)	(0.82)		(0.23)	(2.63)	(1.32)	(2.09)
Minimum level		5.091	2.108		2.263	2.170	1.903	3.000	5.094
		(8.80)	(2.00)		(3.21)	(2.84)	(2.00)	(3.10)	(4.51)
Years of schooling 1960	0.259	0.197	0.164	0.020	0.028	0.023	0.177	0.151	0.198
	(3.04)	(3.06)	(2.42)	(0.33)	(0.49)	(0.40)	(2.50)	(2.46)	(1.43)
GDP pc 1960	-0.309	-0.317	-0.322	-0.321	-0.308	-0.310	-0.325	-0.246	-0.342
	(7.38)	(7.74)	(8.13)	(6.26)	(6.56)	(6.04)	(7.21)	(6.43)	(1.25)
Observations	80	80	80	68	68	68	80	40	40
R-squared (adj.)	0.542	0.685	0.703	0.712	0.740	0.741	0.751	0.718	0.746

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960-2010. Regressions include a constant. Test scores are either the share of pupil reaching the advanced level ("Advanced Level") or the minimum level ("Minimum Level"). Absolute t-statistics in parentheses

[a] Specification includes additional controls for openness, fertility, and tropical location.

[b] Specification includes dummies for the eight world regions reported in HW.

[c] Sample of High Income Countries (countries above sample median of GDP per capita 1960)

[d] Sample of Low Income Countries (countries below sample median of GDP per capita 1960)

Table 6. Sensitivity of estimated effects of advanced and minimum levels of cognitive skills to subsamples

	(1) All countries	(2) High Income Countries[(a)]	(3) Low Income Countries[(a)]	(4) Arab States & SSA countries[(b)]	(5) Asian Countries	(6) Latin Am. Countries
A- OLS						
Advanced level	2.108	1.889	2.346	-17.694	4.143	16.165
	(2.00)	(1.32)	(2.09)	(1.62)	(2.98)	(1.84)
Minimum level	4.186	2.997	5.095	6.230	4.460	-3.537
	(5.64)	(3.10)	(4.51)	(2.77)	(3.15)	(2.31)
Adj. R^2 (Observations)	0.703 (80)	0.718 (40)	0.746 (40)	0.346 (28)	0.955 (14)	0.650 (17)
B1- IV-GMM						
Advanced level	1.983	5.558	2.655	-51.739	-1.393	28.499
	(0.77)	(2.54)	(0.73)	(1.70)	(0.40)	(2.58)
Minimum level	6.412	2.322	7.939	13.396	10.514	-2.748
	(4.59)	(1.82)	(4.05)	(3.46)	(2.07)	(1.08)
F statistic (observations)	7.83 (77)	11.17 (38)	2.77 (39)	2.47 (28)	35.48 (12)	6.11 (17)
F statistic (observations)	38.27 (77)	7.17 (38)	11.76 (39)	5.01 (28)	9.93 (12)	3.68 (17)
B2- IV-GMM						
Advanced level	2.265	6.248	2.892	-53.987	0.725	32.146
	(0.86)	(2.79)	(0.80)	(1.82)	(0.21)	(3.07)
Minimum level	6.458	1.329	7.975	12.901	7.927	-4.695
	(4.63)	(1.07)	(3.99)	(3.58)	(1.82)	(2.08)
F statistic (observations)	6.52 (72)	13.94 (33)	2.19 (39)	1.87 (28)	8.00 (10)	4.26 (17)
F statistic (observations)	31.74 (72)	13.75 (33)	11.33 (39)	7.25 (28)	4.27 (10)	3.35 (17)
B3- IV-GMM						
Advanced level	2.154	4.474	2.448	-64.185	4.975	6.724
	(0.80)	(1.81)	(0.69)	(2.01)	(3.03)	(0.38)
Minimum level	6.376	2.503	8.244	10.896	3.698	2.826
	(4.52)	(1.83)	(3.89)	(3.61)	(1.80)	(0.54)
F statistic (observations)	6.40 (72)	12.52 (33)	2.49 (39)	2.69 (33)	7.35 (10)	3.84 (17)
F statistic (observations)	42.32 (72)	12.56 (33)	14.31 (39)	12.43 (33)	4.95 (10)	3.53 (17)
B4- IV-GMM						
Advanced level	-0.259	4.644	0.768	-59.66	-0.575	24.304
	(0.12)	(3.21)	(0.23)	(2.09)	(0.19)	(2.54)
Minimum level	6.811	1.156	8.186	13.00	9.321	-2.128
	(4.94)	(1.04)	(4.22)	(3.38)	(2.21)	(0.95)
F statistic (observations)	13.41 (75)	35.73 (36)	2.74 (39)	2.01 (28)	16.53 (12)	10.31 (17)
F statistic (observations)	43.81 (75)	29.10 (36)	11.37 (39)	5.41 (28)	13.41 (12)	3.63 (17)

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960-2010. Regressions include a constant. Independent variables include the share of pupil reaching the advanced level ("Advanced Level") or the minimum level ("Minimum Level"). Absolute t-statistics in parentheses. Each panel with IV estimations includes different instruments but always governance effectiveness (GE) and DALY which are powerful predictors for cognitive skills. B1: GE + DALY + initial years of education. B2: GE + DALY + initial years of education + drop out of primary education. B3: GE + DALY + survival rate to the last grade of primary education; B4: GE + DALY + initial years of schooling + initial level of Gini index.

[(a)] Countries above/below sample median of GDP per capita 1960; (b) SSA countries refers to Sub-Saharan Africa

Appendix (not for publication)

Appendix A: Presentation and anchoring methodology of student assessment tests

In this appendix, we present international and regional student achievement tests (respectively ISAT and RSAT) used in this study, and the methodology used in order to obtain the updated dataset based on Altinok, Diebolt, de Meulemeester (2014). For a more detailed presentation, please consult this paper.

We provide below first a short presentation of the various existing learning assessments.

The *International Association for the Evaluation of Educational Achievement* (IEA) was the first body to measure individual learning achievement for international comparative purposes in the early 1960s. The surveys include the highly regarded "Trends in International Mathematics and Science Study" (TIMSS) and "Progress in International Reading Literacy Study" (PIRLS). TIMSS test aims at evaluating skills of students in grades 4 and 8 in mathematics and science, while PIRLS is based on a test based on reading in Grade 4. [29] Another well-known international assessment is PISA (Programme for International Student Assessment). The Organisation for Economic Co-operation and Development (OECD) launched PISA in 1997. More generally, PISA has assessed the skills of 15-year-old every three years since 2000 in countries that together account for almost 90% of the global economy – i.e. a major part of the World GDP. Until now, five rounds of PISA are available (every three years over 2000-2012).

Two other international assessments are available. Drawing on the experience of the *National Assessment of Educational Progress* (NAEP), the *International Assessment of Educational Progress* (IAEP) comprises two surveys first conducted in 1988 and 1991. Under a joint UNESCO and UNICEF project, learning achievements have been assessed as part of the *Monitoring Learning Achievement* (MLA) programme on a vast geographical scale in more than 72 countries (Chinapah, 2003). This programme of assessment is flexible and ranges from early childhood, basic and secondary education to non-formal adult literacy. However, all of the data

[29] A grade consists of a specific stage of instruction in initial education usually covered during an academic year. Students in the same grade are usually of similar age. This is also referred to as a 'class', 'cohort' or 'year' (Glossary of UIS website available at: http://glossary.uis.unesco.org/glossary/en/home).

have not been published. Supplementing national reports, a separate report on MLA I was drafted for 11 African countries (Botswana, Madagascar, Malawi, Mali, Morocco, Mauritius, Niger, Senegal, Tunisia, Uganda and Zambia; see UNESCO, 2000). As the microdata of IAEP and MLA is not available, we preferred not to include these assessments in our database.

Three major regional assessments (RSATs) have been conducted in Africa and Latin America. The *Southern and Eastern Africa Consortium for Monitoring Educational Quality* (SACMEQ) grew out of a very extensive national investigation of the quality of primary education in 15 African countries in 1995-1999, 2000-2002 and 2007. Following a different approach, surveys under the *Programme d'Analyse des Systèmes Educatifs* (PASEC, or "Programme of Analysis of Education Systems") of the Conference of Ministers of Education of French-Speaking Countries (CONFEMEN) have been conducted in the French-speaking countries of sub-Saharan Africa since 1993. Finally, the network of national education systems in Latin American and Caribbean countries, known as the *Latin American Laboratory for Assessment of the Quality of Education* (LLECE), was established in 1994 and is coordinated by the UNESCO Regional Bureau for Education in Latin America and the Caribbean. Assessments conducted by the LLECE focused on learning achievements in reading, mathematics and science[30] in grades 3 and 4[31] in 13 countries of the subcontinent in 1998 and for grade 3 and 6 pupils in 2006.

All achievements tests undertaken and the main information concerning them are summarized in Table A.1 below. The methodology used to adjust them is presented in section A2 below and in more details in Altinok *et al.* (2014).

A.1. Presentation of student assessment tests

A.1.1. International learning assessments

The International Association for the Evaluation of Educational Achievement (IEA) was the first body to measure individual learning achievements and conduct recurrent surveys for

[30] Science skill was included in the second round only.
[31] A grade is a stage of instruction usually equivalent to one complete year. Hence, grade 3 represents the third year of compulsory schooling – i.e. of primary education in most countries.

international comparative purposes as soon as in the early 1960s. The surveys include the highly regarded "Trends in International Mathematics and Science Study" (TIMSS) and "Progress in International Reading Literacy Study" (PIRLS).

A. TIMSS. The major survey series from IEA is the "Trends in International Mathematics and Science Study" (TIMSS). Five TIMSS rounds have been held to date. The first, conducted in 1995, covered 45 national educational systems and three groups of learners[32]. 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. The last round were done in 2011 and covered 77 countries/areas[33].

B. PIRLS. The other major IEA survey is the "Progress in International Reading Literacy Study", also known as PIRLS. Three major rounds of PIRLS have held up to 2011: in 2001, in 2006 and in 2011. PIRLS survey tests pupils from primary schools in reading literacy[34]. The last PIRLS round was done together with TIMSS (2011) and included 60 countries/areas.

C. PISA. The Organisation for Economic Co-operation and Development (OECD) is another international organisation that has carried out standardised international comparisons of pupils achievements. The OECD launched its "Programme for International Student Assessment" (PISA) in 1997 to meet the need for readily comparable data of student performance. In 2009, 72 countries/areas participated, while 64 countries/areas expect to take part to PISA 2012. Unlike the IEA surveys, PISA assesses only 15-year-old pupils, whatever their school level, whereas the grade is the main criterion in selecting pupils for IEA assessments (and all over student achievement tests).

A.1.2. Regional learning assessments

Three major regional assessments have been conducted in Africa and Latin America.

[32] IEA assessments defined populations relative to specific grades, while assessments as PISA focus on age of pupils. In IEA studies, three different group of pupils are generally assessed: pupils from grade 4, grade 8 and from the last grade of secondary education. Some Canadian provinces or states in the United States of America have occasionally taken part in the IEA surveys. For the sake of simplicity, these regions are not included in the number of countries participating in the surveys.

[33] Since the micro data is not yet released, we did not included TIMSS 2011 dataset in our database. However, the database will be updated regularly with new released data.

[34] Similarly to TIMSS, pupils from Grade 4 were chosen.

D. SACMEQ. The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) is an assessment which concerns mainly anglophone Sub-Saharan countries. The first SACMEQ round took place between 1995 and 1999. SACMEQ I thus covered seven different countries and assessed performance in reading at grade 6. In the second round, which was held between 2000 and 2002 and covered 14 countries and one territory (Zanzibar), performance in mathematics and reading was assessed. The target cohort consisted of grade 6 pupils, as under SACMEQ I. The third SACMEQ round (SACMEQ III) is covering the same countries as in 2002 (plus Zimbabwe) and focuses on achievements levels of grade 6 pupils.

E. PASEC. Surveys under the "Programme d'Analyse des Systèmes Éducatifs" (PASEC, or "Programme of Analysis of Education Systems") of the Conference of Ministers of Education of French-Speaking Countries (CONFEMEN) have been conducted in the French-speaking countries of sub-Saharan Africa. This assessment contains results for primary school performance in mathematics and in French. In both CP2 (the second grade in primary school) and CM1 (grade 5), more than 15 countries have taken part to PASEC assessments. In order to simplify the analysis, we will consider two different rounds of PASEC: the first round includes assessments occurred between 1996 and 2003, whereas the PASEC II takes into account evaluations which have been done between 2004 and 2010.

F. LLECE. The network of national education systems in Latin American and Caribbean countries, known as the Latin American Laboratory for Assessment of the Quality of Education (LLECE), was formed in 1994. Assessments conducted by the LLECE thus focused on learning achievement in reading and mathematics in grades 3 and 4 in 13 countries of the subcontinent (Casassus et al., 1998, 2002), namely Argentina, Bolivia, Brazil, Chile, Columbia, Costa Rica, Cuba, Dominican Republic, Honduras, Mexico, Paraguay, Peru and the Bolivarian Republic of Venezuela (Casassus et al., 1998). In 2006, the second round of the LLECE survey was initiated in the same countries as LLECE I. Our analysis will include only LLECE II results, since the grade tested is the last grade in all countries.

A.2. Methodology of anchoring student assessment tests

We hereafter present the methodology used to compare the main database underlying our indicators of quality of student achievement (IQSA). For a more detailed presentation, please refer to the original paper (Altinok *et al*., 2014).

In their pioneering paper, Lee and Barro (2001) used direct results from ISATs, without any specific methodology for adjusting potential differences between all the series. They used instead a regression technique – *i.e.* the seemingly unrelated regression – which allows to obtain different constants between each test, and hence to allow for potential differences between tests over years and over skills.

Another method of anchoring has been used by Hanushek and Kimko (2000). These authors adjusted ISATs between 1964 and 1995 by using results from NAEP (*National Assessment of Educational Progress*[35]). Their methodology is only based on United States scores, and the data is limited to the period of 1964-1995. A recent paper by Hanushek and Woessmann (2012) aimed at correcting some of these imperfections by using an approach that assumes stability over time of the variance of quality of student achievement in a restricted number of OECD countries. The authors suggest two criteria for a group of countries to serve as a standardization benchmark for performance variation over time. Firstly, the countries have to be member states of the relatively homogenous and economically advanced group of OECD countries in the whole period of ISATs observations. Second, the countries should have had a substantial enrollment in secondary education already in 1964. Then, the authors suggest 13 countries that meet both of these measures of stability which are named “OECD Standardization Group” (OSG) of countries[36]. Hanushek and Woessmann (2012) assume that the cross-country variation among the OSG countries do not vary substantially since 1964. By using this assumption, they build new indicators of student achievements and educational quality. Their main measure of cognitive skills is a simple average of all standardized math and science test scores of the ISATs in which a country participated. Their database includes combined measure for the 77 countries that have ever participated in any of the math and science tests.

[35] A description of the NAEP can be found in Altinok et al. (2014).

[36] The OSG countries are: Austria, Belgium, Canada, Denmark, France, Germany, Iceland, Japan, Norway, Sweden, Switzerland, the United Kingdom, and the United States.

As the authors explain in their paper, a major issue with this methodology concerns countries that are far from the measured OECD performance. In particular, countries far off the scale of the original test scores may not be well represented because the tests may be too hard and thus not very informative for them. This bias may be more important when analyses are focused on developing countries, which is the case of our study.

Moreover, the methodology used – i.e. the "OSG of countries" – does not take into account several improvements made by ISATs since 1995. The International Association for the Evaluation of Educational Achievement (IEA) and the Organisation for Economic Co-operation and Development (OECD) teams prepared modern ISATs in order to allow intertemporal comparisons. By using another methodology, Hanushek and Woessmann (2012) chose a specific approach and neglected the recent improvements made by psychometricians, as the Item Response Theory (IRT). Moreover, they do not clearly show to what extent their main assumption – i.e. the variation between the OSG of countries is stable – is corroborated by results from modern ISATs, as these ones permit to compare countries performance over time.

Another limit deals with the absence of Regional Student Achievement Tests (RSATs) in their database. Hanushek and Woessmann (2012) only focused in ISATs since the methodology used is based on developed economies. In our paper, we provide an alternative methodology which enables the possibility to include main regional assessments. As these ones are focused on developing countries, our study permits to analyze more deeply the question of universal primary education and specific analysis for developing economies.

Below we provide details of the steps used to generate the data.

Step 1: Distinction between assessments (A.1, A.2 and B)

As the methodology of anchoring differs between assessments, we allocate each assessment to three different groups of surveys (groups A.1, A.2 and B). Surveys grouped in survey series A are from IEA and OECD, while assessments from survey series B are from RSATs (PASEC, SACMEQ and LLECE). In Table A.1., we present the different assessments used in our study with the classification used.

Concerning surveys A.1. and A.2., two complementary methods of anchoring can be used in order to obtain comparable scores. The first one is related with the permanent anchoring of each score with the results of USA. Since the USA took part to all international assessments, it remains possible to obtain comparable scores between assessments by anchoring the performance of this country with a national assessment. The surveys included in this group are mainly the ones which occurred until 1995. More precisely, we include all ISATs in reading until 2001, and all ISATs in mathematics and science until 1995 for IEA surveys, while surveys for PISA are included until 2000[37].

Other assessments (PIRLS [2006], TIMSS [1999, 2003, 2007], PISA [2003, 2006, 2009]) are grouped in the survey series A.2. This second group of assessments is adjusted with another methodology (presented in step 3).

Step 2: Adjustment of Survey series A.1.

The methodology used in survey series A.1 involved recurrent adjustment to the *US National Assessment of Educational Progress* (NAEP) survey in the way described by Hanushek and Kimko (2000). The NAEP has been the main instrument used to measure the learning achievement of pupils in the United States of America since 1969. Based on the Hanushek and Kimko (2000) methodology, we can scale the level of each ISAT relative to the comparable test performance of the United States, by using the time series evidence on the performance of U.S. students. Following assessments are included in this group: FIMS, FISS, RLS, TIMSS 1995, PIRLS 2001 and PISA 2000 for reading.

Step 3: Adjustment of survey series A.2.

If recent assessments such as PISA 2009 or PIRLS 2006 were to be adjusted according to the above procedure, all survey scores would be based on scores obtained in the USA. However, recent surveys have been designed to allow analysis of country across time. The same adjustment coefficients as those calculated for the series A surveys are used in order to achieve a single linear conversion of country scores, and this procedure does not compromise the comparability of the scores obtained by countries participating in the same survey series. As highlighted above, this main difference with the methodology used by Hanushek and Woessmann (2012a) allows us to avoid a potential bias in estimating trends in pupils performance for countries for which scores

[37] In the case of mathematics, we also include PISA 2003 survey, while in science, PISA 2003 and 2006 surveys are included too, since the PISA datasets does not allow direct comparability of scores between 2000 and 2003 for mathematics and between 2000 and 2006 for science.

are already comparable over time. As our anchoring methodology is a simple linear transformation of surveys, the trends of schooling performance observed in PISA reports are still present in our database (see OECD, 2010)[38].

Step 4: Adjustment of survey series B

For the last group of surveys (called « Series B »), we cannot use a simple anchoring method on the NAEP, since the United States did not participate to any regional assessment. We selected countries which participated in at least two different surveys so as to establish a comparison between the surveys. The IEA surveys have been chosen as reference survey as they cover most of the countries and as the economic levels of participating countries is the most heterogeneous.

As some countries took part simultaneously to several assessments, we can suppose that the differences found between assessments are exogenous to the performance of these countries.

Step 5. Cross country dataset

It may be of interest to compare countries' average performance in international and regional surveys. First, countries' average score for all surveys at the same educational level is calculated. Next, each country's average score in each skill and in all primary education surveys is calculated. The same is then done for secondary education and each skill. We then obtain 6 different series of cross country data for educational achievement, since we distinguish for the level of education (primary versus secondary) and the skill evaluated (mathematics, science and reading). The mean score in each level is then computed by averaging scores from each skill. Moreover, we aggregated results from each education level in order to obtain a single general score of schooling performance for each country. Since, the dataset for each level is unbalanced, we firstly predicted scores for all countries by using all the existing information (i.e. with general scores), and then we obtained the total scores for education (primary + secondary).

Step 6. Update of the database

Since original database covers student assessment test only until 2007, we udpated it with exactly the same methodology by using recent assessments published by IEA (TIMSS 2011, PIRLS

[38] However, a problem occur when for some countries, we detect a big difference in trends between IEA and PISA assessments. For instance, in can be possible that for a low number of countries, the performance of pupils increased in TIMSS while we observed a decrease in PISA. Instead of merging both variations - which would lead to a stagnation of score - we prefer to focus primarily to IEA results.

2011) and OECD (2012). Hence, our updated dataset covers a period from 1965 to 2012, which represents approximately 50 years.

Table A.1 Review of student achievement tests included in the study

No	*Year*	*Organization*	*Abbr.*	*Subject*	*Countries/ Areas*	*Grade/Age*	*Survey Series*
2	1964	IEA	FIMS	M	12	7, FS	A.1.
3	1970-71	IEA	SRC	R	15	4,8, FS.	-
4	1970-72	IEA	FISS	S	19	4,8, FS.	A.1
5	1980-82	IEA	SIMS	M	19	8, FS	A.2
6	1983-1984	IEA	SISS	S	23	4,8, FS	A.2
8	1990-1991	IEA	RLS	R	32	3-4, 7-8	A.1
9	1995,1999,2003,2007,2011	IEA	TIMSS	M,S		3-4, 7-8, FS	A.1 (1995), A.2. (Other years - except 2011)
11	1997, 2006	UNESCO	LLECE	M,S,R	13	3,6	B.
12	1999, 2002, 2007	UNESCO	SACMEQ	M,R	7, 15,16	6	B
13	1993-2001, 2002-2012	CONFEMEN	PASEC	M ;R		2, 5	B
14	2001, 2006, 2011	IEA	PIRLS	R	35, 41, 55	4	A.1 (2001) ; A.2. (Other years - except 2011)
16	2000-2012 (every 3 years)	OECD	PISA	M,S,R	43, 41, 57, 75	Age 15	A.1 (2000 for reading ; 2003 for maths ; 2006 for science); A.2. (Other years for reading ; until 2003 for maths ; until 2006 for science)

Note: For the meaning of abbreviations, please consult "Abbreviations" page. « FS » means "Final Secondary". Only assessments for which there is an information in "Survey Series" column are included in our dataset.

Table A.2. Descriptive statistics

	Obs.	Mean	Std. Dev.	Min.	Max.
Original HW dataset					
AAGR in GDP pc 1960-2000	50	3.051	1.403	1.112	7.396
Cognitive skills*	50	4.546	0.611	3.089	5.452
GDP pc 1960	50	4.995	3.686	0.682	14.978
Years of schooling, 1960	50	5.289	2.739	0.620	10.540
Mean years of schooling, 1960-2000	50	7.154	2.619	1.804	11.650
Updated Altinok *et al.* dataset					
AAGR in GDP pc 1960-2010	80	2.560	1.486	-0.983	6.814
Cognitive skills*	80	4.654	1.028	2.443	6.289
GDP pc 1960	80	4.157	3.879	0.323	17.056
Years of schooling, 1960	80	3.783	2.504	0.195	10.170
Mean years of schooling, 1960-2010	80	6.078	2.600	0.896	11.743
Instrumental variables estimates					
External exit exam system	30	0.595	0.473	0	1
Initial years of schooling	30	5.657	2.488	0.908	10.170
Catholic share in 1900	80	0.323	0.413	0	1
Relative teacher salary	47	7.140	23.097	0.25	157.38
DALY	78	1.162	1.696	0.046	6.176
EDR	78	0.036	0.050	0.002	0.185
Governance effectiveness	77	0.347	1.021	-1.73	2.1

* Cognitive skills variables are standardized in all estimations with a mean equal to 0 and a standard deviation equal to 1.

Table A.3. International data on cognitive skills

Country	Cognitive[a]	Basic[b]	Advanced[c]	Data available in HW[d]	Data available in our updated database[e]	Country included in HW regressions[f]	Country included in our regressions[g]
Albania	4.506	0.674	0.098	1	1	0	0
Algeria	4.127	0.586	0.008	0	1	0	0
Argentina	4.469	0.652	0.092	1	1	1	1
Armenia	5.021	0.841	0.159	1	1	0	0
Austria	5.797	0.969	0.398	1	1	1	1
Azerbaijan	4.783	0.808	0.083	0	1	0	0
Bahrain	4.698	0.765	0.086	1	1	0	0
Belgium	5.913	0.922	0.462	1	1	1	1
Belize	3.667	0.415	0.052	0	1	0	0
Benin	2.581	0.281	0.058	0	1	0	1
Bosnia	4.910	0.868	0.100	0	1	0	0
Botswana	3.951	0.469	0.022	1	1	0	0
Brazil	4.425	0.622	0.080	1	1	1	1
Bulgaria	5.505	0.905	0.343	1	1	0	0
Burkina Faso	3.328	0.383	0.075	0	1	0	0
Burundi	3.395	0.375	0.089	0	1	0	1
Cameroun	3.938	0.496	0.102	0	1	0	1
Canada	5.869	0.964	0.428	1	1	1	1
Chad	2.657	0.283	0.057	0	1	0	0
Chile	4.900	0.805	0.139	1	1	1	1
China	5.204	0.906	0.230	1	1	1	1
Colombia	4.418	0.648	0.069	1	1	1	1
Comoros	2.819	0.307	0.068	0	1	0	0
Congo	2.909	0.322	0.071	0	1	0	1
Costa Rica	4.854	0.813	0.119	0	1	0	1
Croatia	5.637	0.969	0.323	0	1	0	0
Cuba	5.402	0.867	0.304	0	1	0	0
Cyprus	5.066	0.874	0.150	1	1	1	1
Czech Rep.	5.810	0.973	0.401	1	1	0	0
Côte d’Ivoire	3.037	0.314	0.051	0	1	0	1
Denmark	5.762	0.955	0.402	1	1	1	1
Dominican Rep.	3.728	0.336	0.043	0	1	0	1
Ecuador	3.924	0.441	0.046	0	1	0	1
Egypt	4.364	0.622	0.059	1	1	1	1
El Salvador	3.976	0.476	0.008	0	1	0	1
Estonia	5.985	0.971	0.475	1	1	0	0
Finland	6.080	0.982	0.523	1	1	1	1
Portugal	5.757	0.963	0.381	1	1	1	1
Gabon	3.906	0.497	0.059	0	1	0	0
Georgia	4.731	0.767	0.092	0	1	0	0
Germany	5.838	0.965	0.417	1	1	0	1
Ghana	3.417	0.246	0.009	1	1	1	1
Greece	5.349	0.898	0.260	1	1	1	1

Country	Cognitive(a)	Basic(b)	Advanced(c)	Data available in HW(d)	Data available in our updated database(e)	Country included in HW regressions(f)	Country included in our regressions(g)
Guatemala	3.929	0.421	0.047	0	1	0	1
Honduras	4.208	0.578	0.024	0	1	0	1
Hong Kong	6.108	0.972	0.550	1	1	1	1
Hungary	5.802	0.962	0.422	1	1	0	0
Iceland	5.504	0.922	0.301	1	1	1	1
India	4.253	0.591	0.031	1	1	1	1
Indonesia	4.436	0.667	0.048	1	1	1	1
Iran I.R.	4.578	0.731	0.060	1	1	1	1
Ireland	5.833	0.959	0.425	1	1	1	1
Israel	5.172	0.793	0.272	1	1	1	1
Italy	5.632	0.949	0.344	1	1	1	1
Japan	6.173	0.984	0.569	1	1	1	1
Jordan	4.797	0.771	0.140	1	1	1	1
Kazakhstan	5.279	0.908	0.235	0	1	0	0
Kenya	4.024	0.468	0.049	0	1	0	0
Korea Rep.	6.289	0.984	0.633	1	1	1	1
Kuwait	3.982	0.512	0.021	1	1	0	0
Kyrgyzstan	3.750	0.340	0.024	0	1	0	0
Latvia	5.701	0.969	0.362	1	1	0	0
Lebanon	4.537	0.731	0.058	1	1	0	0
Lesotho	3.416	0.235	0.043	0	1	0	1
Liechenstein	5.973	0.954	0.472	1	1	0	0
Lithuania	5.669	0.972	0.339	1	1	0	0
Portugal	5.785	0.959	0.420	1	1	0	0
Macao	5.936	0.966	0.476	1	1	0	0
Macedonia F.Y.R.	4.740	0.744	0.115	1	1	0	0
Madagascar	3.817	0.490	0.142	0	1	0	0
Malawi	3.282	0.200	0.043	0	1	0	1
Malaysia	5.077	0.877	0.183	1	1	1	1
Mali	2.735	0.288	0.060	0	1	0	1
Malta	5.235	0.847	0.250	0	1	0	1
Mauritania	1.980	0.228	0.046	0	1	0	0
Mauritius	4.459	0.635	0.130	0	1	0	1
Mexico	4.666	0.734	0.100	1	1	1	1
Moldova	5.149	0.886	0.177	1	1	0	0
Mongolia	4.698	0.791	0.053	0	1	0	0
Montenegro	4.764	0.772	0.126	0	1	0	0
Morocco	3.729	0.415	0.009	1	1	1	1
Mozambique	3.688	0.313	0.043	0	1	0	1
Namibia	3.417	0.257	0.043	0	1	0	1
Netherlands	5.924	0.980	0.431	1	1	1	1
Nicaragua	4.025	0.470	0.044	0	1	0	0
Niger	2.443	0.257	0.052	0	1	0	1
Nigeria	3.988	0.483	0.015	1	1	0	0
Norway	5.514	0.931	0.289	1	1	1	1

Country	Cognitive[(a)]	Basic[(b)]	Advanced[(c)]	Data available in HW[(d)]	Data available in our updated database[(e)]	Country included in HW regressions[(f)]	Country included in our regressions[(g)]
Oman	4.216	0.589	0.039	0	1	0	0
Palestine	4.287	0.597	0.049	1	1	0	0
Panama	4.163	0.523	0.050	0	1	0	1
Papua New Gui.	4.867	0.912	0.072	0	1	0	0
Paraguay	4.074	0.507	0.051	0	1	0	1
Peru	4.116	0.522	0.042	1	1	1	1
Philippines	3.894	0.443	0.031	1	1	1	1
Poland	5.588	0.933	0.327	1	1	0	0
Portugal	5.460	0.919	0.293	1	1	1	1
Qatar	4.035	0.481	0.071	0	1	0	0
Romania	5.243	0.872	0.244	1	1	1	1
Russian Fed.	5.749	0.968	0.396	1	1	0	0
Saudi Arabia	4.245	0.592	0.029	1	1	0	0
Senegal	3.085	0.348	0.068	0	1	0	1
Serbia	5.398	0.918	0.267	1	1	0	0
Seychelles	4.103	0.535	0.050	0	1	0	0
Singapore	6.250	0.961	0.609	1	1	1	1
Slovakia	5.678	0.954	0.362	1	1	0	0
Slovenia	5.613	0.955	0.320	1	1	0	0
South Africa	3.180	0.213	0.021	1	1	1	1
Spain	5.554	0.951	0.303	1	1	1	1
Swaziland	4.142	0.521	0.024	1	1	0	0
Sweden	5.793	0.966	0.393	1	1	1	1
Switzerland	5.739	0.926	0.426	1	1	1	1
Syrian A.R.	4.296	0.624	0.024	0	1	0	1
Tanzania U.R.	3.998	0.456	0.045	0	1	0	1
Thailand	5.096	0.880	0.160	1	1	1	1
Togo	3.350	0.388	0.068	0	1	0	1
Trinidad & T.	4.853	0.749	0.168	0	1	0	1
Tunisia	4.138	0.570	0.057	1	1	1	1
Turkey	5.098	0.850	0.196	1	1	1	1
Uganda	3.594	0.300	0.044	0	1	0	1
Ukraine	5.105	0.885	0.152	0	1	0	0
United Arab Em.	4.944	0.782	0.181	0	1	0	0
Uruguay	4.833	0.784	0.138	1	1	1	1
Venezuela	4.042	0.487	0.048	0	1	0	1
Vietnam	5.971	0.968	0.482	0	1	0	0
Yemen	2.761	0.282	0.046	0	1	0	0
Zambia	3.293	0.219	0.043	0	1	0	1
Zimbabwe	4.329	0.686	0.020	1	1	1	1

Notes:
[(a)] Average test score in math and science, primary through lower secondary, all years.
[(b)] Share of students reaching basic literacy
[(c)] Share of students reaching advanced level in achievement tests
[(d)] Data for cognitive skills is available in the HW database
[(e)] Data for cognitive skills is available in our updated database

[f] Data for cognitive skills is both available in the HW database and in growth regressions
[g] Data for cognitive skills is both available in our updated database and in growth regressions

Table A.4. Sensitivity of estimated effects of cognitive skills to the sample of countries and time periods

	(1) Full	(2) OECD	(3) Non-OECD	(4) W/o East Asia	(5) 1960-1980	(6) 1980-2000 (A) 1980-2010 (B)	(7) Score-schooling outliers[b]	(8) Score-schooling core[b]
Test-score specification								
(A) Data from Hanushek and Woessmann (2012a)								
All math and science	1.120	1.144	1.273	0.795	0.961	1.714	1.140	1.634
	(7.38)	(7.08)	(5.66)	(7.21)	(3.32)	(9.12)	(6.71)	(6.52)
Only lower secondary	1.177	1.128	1.231	0.777	0.985	1.629	1.111	1.484
	(6.84)	(6.39)	(5.68)	(6.15)	(3.46)	(7.87)	(5.19)	(6.48)
(B) Data from updated Altinok, Diebolt, de Meulemeester (2014)								
(B1) All math and science	1.510	1.471	1.524	1.268	1.681	1.681	1.394	0.745
	(10.50)	(3.41)	(7.70)	(6.52)	(6.84)	(8.13)	(8.72)	(1.88)
(B2) Only lower secondary	1.107	1.057	1.109	1.023	1.484	1.143	0.944	1.295
	(9.06)	(3.50)	(6.30)	(5.74)	(6.78)	(7.77)	(8.36)	(4.85)
No. of countries (A)	50	26	24	38	50	50	25	25
No. of countries (B1)	80	27	53	68	80	80	41	39
No. of countries (B2)	56	27	29	45	56	56	28	28

Notes: Reported numbers are the coefficient on test scores in each model specification. Dependent variable: Unless noted otherwise, average annual growth rate in GDP per capita, 1960-2000 (Hanushek and Woessmann, 2012a), 1960-2010 (updated data from Altinok *et al.*,2014)). Control variables: Initial GDP per capita, initial years of schooling and a constant. Test scores: Unless noted otherwise, average of math, reading and science, primary through secondary school, all years. Absolute t-statistics in parentheses

[b] Countries with largest (outliers)/smallest (core) residuals when regressing years of schooling on test scores

Table A.5. Sensitivity of estimated effects of cognitive skills to the measurement of skills

	(1) Full	(2) OECD	(3) Non-OECD	No. of countries (Non-OECD countries)
Test-score specification				
(A) Only since 1995	1.392	1.419	1.335	76 (49)
	(9.37)	(3.55)	(6.52)	
(B) Only lower secondary since 1995	1.107	1.057	1.109	56 (29)
	(9.06)	(3.50)	(6.30)	
(C) Only until 1995	1.545	0.771	1.491	46 (22)
	(7.62)	(2.04)	(3.86)	
(D) Only math	1.401	1.283	1.354	76 (49)
	(10.81)	(4.04)	(7.42)	
(E) Only science	1.012	0.957	1.021	61 (34)
	(5.83)	(2.61)	(4.24)	
(F) Only reading	1.443	1.848	1.331	73 (46)
	(7.51)	(2.99)	(4.74)	
(G) All subjects entered jointly				54 (27)
Math	1.726	1.432	1.811	
	(4.71)	(2.52)	(3.61)	
Science	-0.208	-0.367	-0.354	
	(0.77)	(0.88)	(0.96)	
Reading	-0.487	0.359	-0.451	
	(0.73)	(0.41)	(0.48)	

Notes: Reported numbers are the coefficient on test scores in each model specification. Dependent variable: average annual growth rate in GDP per capita, 1960-2010. Control variables: Initial GDP per capita, initial years of schooling and a constant. Test scores: Unless note otherwise, average of math, reading and science, primary through lower secondary school, all years. Data relative to cognitive skills is from updated Altinok *et al.* (2014) dataset. Absolute t-statistics in parentheses.