J Econ Growth (2012) 17:267–321
DOI 10.1007/s10887-012-9081-x

Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation

Eric A. Hanushek · Ludger Woessmann

Published online: 14 July 2012

Abstract We develop a new metric for the distribution of educational achievement across countries that can further track the cognitive skill distribution within countries and over time. Cross-country growth regressions generate a close relationship between educational achievement and GDP growth that is remarkably stable across extensive sensitivity analyses of specification, time period, and country samples. In a series of now-common microeconometric approaches for addressing causality, we narrow the range of plausible interpretations of this strong cognitive skills-growth relationship. These alternative estimation approaches, including instrumental variables, difference-in-differences among immigrants on the U.S. labor market, and longitudinal analysis of changes in cognitive skills and in growth rates, leave the stylized fact of a strong impact of cognitive skills unchanged. Moreover, the results indicate that school policy can be an important instrument to spur growth. The shares of basic literates and high performers have independent relationships with growth, the latter being larger in poorer countries.

Keywords Cognitive skills · Long run growth · Causation and identification · School quality · Educational achievement

E. A. Hanushek (✉)
Hoover Institution, Stanford University, Stanford, CA 94305-6010, USA
e-mail: hanushek@stanford.edu
URL: www.hanushek.net

E. A. Hanushek · L. Woessmann
CESifo, University of Munich, Munich, Germany

E. A. Hanushek
NBER, National Bureau of Economic Research (NBER), Cambridge, MA, USA

L. Woessmann
Ifo Institute for Economic Research, University of Munich, Poschingerstr. 5, 81679 Munich, Germany
e-mail: woessmann@ifo.de
URL: www.cesifo.de/woessmann

1 Introduction

Schooling and human capital investments have been a central focus of development policy, but doubts have arisen as disappointments with results grow. Nowhere is this more apparent than in the case of growth policy, where schooling investments have not appeared to return the economic outcomes promised by theoretical growth models.[1] Prior analyses into the specification of empirical cross-country growth models lead to a warranted skepticism about the identification of causal growth effects. Our analysis of newly developed measures of skill differences based on international tests of math and science suggests, however, that one of the most significant problems underlying these prior concerns is the valid measurement of human capital across countries. We find that accurately measuring differences in educational achievement, which we refer to simply as cognitive skills, dramatically improves our ability to explain variations in long-run growth across countries. Moreover, while having limitations in macroeconomic applications, a set of microeconometric approaches can be employed in the cross-country setting to rule out many of the common concerns that undermine causal interpretations.

As a simple summary observation, world policy attention today focuses on the lagging fortunes of Sub-Saharan Africa and of Latin America. Considerably less attention goes to East Asia, and, if anything, East Asia is proposed as a role model for the lagging regions. Yet to somebody contemplating development policy in the 1960s, none of this would be so obvious. Latin America had average income exceeding that in Sub-Saharan Africa and the Middle East and North Africa regions, and both of these exceeded East Asia (see Appendix Table 8).[2] Further, Latin America had schooling levels that exceeded those in the others, which were roughly equal. Thus, on the basis of observed human capital investments, one might have expected Latin America to pull even farther ahead while having no strong priors on the other regions. The unmistakable failure of such expectations, coupled with a similar set of observations for separate countries in the regions, suggests skepticism about using human capital policies to foster development. But, this skepticism appears to be more an outgrowth of imperfect measurement of human capital investments than an empirical reality.

The measurement issues become apparent when we introduce direct measures of cognitive skills from international tests of math and science into the growth picture. The entire picture changes. Figure 1 plots regional growth in real per capita GDP between 1960 and 2000 against average test scores after conditioning on initial GDP per capita in 1960.[3] Regional annual growth rates, which vary from 1.4 % in Sub-Saharan Africa to 4.5 % in East Asia, fall on a straight line with an $R^2 = 0.985$. But, school attainment, when added to this regression, is unrelated to growth-rate differences. Figure 1 suggests that, conditional on initial income levels, regional growth over the last four decades is completely described by differences in cognitive skills.

In the upsurge of empirical analyses of why some nations grow faster than others since the seminal contributions by Barro (1991, 1997) and Mankiw et al. (1992), a vast literature of cross-country growth regressions has tended to find a significant positive association between

[1] See, for example, Pritchett (2006).

[2] Japan was significantly ahead of the rest of the East Asia region, but its exclusion does not change the regional ordering (see Appendix Table 8).

[3] Regional data come from averaging all countries with available data in a region. The 50 countries are not chosen to be representative but instead represent the universe of countries that participated in international tests and had available the requisite economic data. Still, Appendix A shows that the average 1960 incomes for all countries in each region are quite similar to those for our subset of countries. The division of Europe into three regions illustrates the heterogeneity within OECD countries, but a combined Europe also falls on the line in Fig. 1.

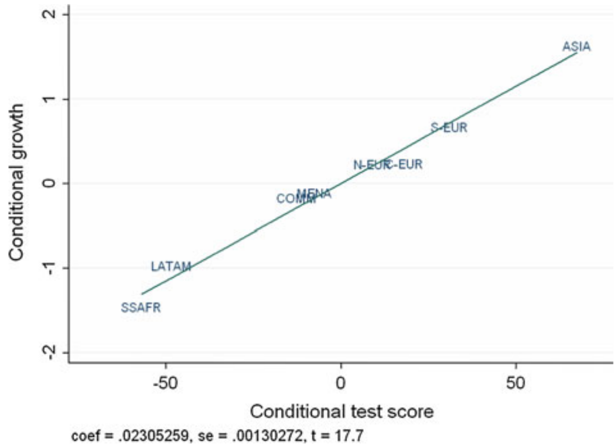

Fig. 1 Cognitive Skills and Growth across World Regions. *Notes*: Added-variable plot of a regression of the average annual rate of growth (in percent) of real GDP per capita in 1960–2000 on the initial level of real GDP per capita in 1960 and average test scores on international student achievement tests. Authors' calculations. See Table 8 for a list of the countries contained in each world region. Region codes: East Asia and India (ASIA), Central Europe (C-EUR), Commonwealth OECD members (COMM), Latin America (LATAM), Middle East and North Africa (MENA), Northern Europe (N-EUR), Southern Europe (S-EUR), Sub-Saharan Africa (SSAFR)

quantitative measures of schooling and economic growth.[4] But, all analyses using average years of schooling as the human capital measure implicitly assume that a year of schooling delivers the same increase in knowledge and skills regardless of the education system. For example, a year of schooling in Peru is assumed to create the same increase in productive human capital as a year of schooling in Japan. Equally as important, this measure assumes that formal schooling is the primary source of education and that variations in the quality of nonschool factors have a negligible effect on education outcomes.

In this paper, we concentrate directly on the role of cognitive skills. This approach was initiated by Hanushek and Kimko (2000), who related a measure of educational achievement derived from the international student achievement tests through 1991 to economic growth in 1960–1990 in a sample of 31 countries with available data. They found that the association of economic growth with cognitive skills dwarfs its association with years of schooling and raises the explanatory power of growth models substantially. Their general pattern of results has been duplicated by a series of other studies over the past 10 years that pursue different tests and specifications along with different variations of skills measurement.[5]

But should we interpret the tight relationship between cognitive skills and growth as reflecting a causal relationship that can support direct policy actions? Questions about the identification of underlying causal effects in cross-country growth models have existed for a long time and go beyond just the impact of human capital. Beginning with Levine and Renelt (1992), plentiful evidence of the general sensitivity to alternative samples and specifications has convinced many that cross-country empirical models are not fruitful policy investigations. In terms of schooling, Bils and Klenow (2000) provide convincing evidence of the

[4] For extensive reviews of the literature, see Topel (1999), Krueger and Lindahl (2001), Pritchett (2006), and Hanushek and Woessmann (2008). The robustness of the association is highlighted by the extensive analysis by Sala-i-Martin et al. (2004): Of 67 explanatory variables in growth regressions on a sample of 88 countries, primary schooling turns out to be the most robust influence factor (after an East Asian dummy) on growth in GDP per capita in 1960–1996.

[5] Detailed discussion of these studies is available in Hanushek and Woessmann (2011a).

endogeneity of school attainment in growth models. Further, it is unclear to what extent prior attempts to deal with endogeneity, such as the panel data approaches of Barro (1997) and Vandenbussche et al. (2006), have been successful in a setting where the dominant information is found in the cross-country variation.[6] Perhaps the strongest evidence on causality has been related to the importance of fundamental economic institutions using identification through historical factors (Acemoglu et al. 2001, 2005), but this has not yielded clear advice about the kinds of feasible policies that will lead to national payoffs, and it itself has been subject to question (Glaeser et al. 2004).

When estimating the effect of cognitive skills on growth, the main causality concerns relate to reverse causality and to omitted country variables such as inherent difference in nations' culture and economic institutions that are correlated both with economic growth and with cognitive skills or their determinants. We assess these issues from a number of angles with the objective of narrowing the range of threats to a causal interpretation. Of course, it is virtually impossible to identify causality in a thoroughly convincing manner given the limited observations underlying cross-country growth models. Each approach we employ deals with one or more common concerns such as the influence of cultural differences, faulty measurement of cognitive skills, or simple reverse causality. But each relies upon strong maintained hypotheses that may or may not be completely persuasive.

Our analysis, while building on Hanushek and Kimko (2000), provides new evidence about the potential causal interpretation of the cognitive skills-growth relationship. The development of a new data series on cognitive skills (Sect. 3), expanded to 64 countries for some analyses, permits approaches to estimation not previously possible. We are able to improve on the underlying measurement, to increase the country observations to a broader range of development experiences, to extend the period of observed long-term growth to 1960–2000 (and to 2007 in some specifications), and to add both a longitudinal and a within-country distributional dimension to the database. We begin by showing that the relationship between cognitive skills and economic growth is extraordinarily robust to alternative samples defined by different time periods and sets of countries and to different specifications of the skills measure and of the growth relationship (Sect. 4).

The core of the paper applies a series of approaches to identification of causal parameters now common in microeconometric studies to the macroeconomic analysis of growth, although the application to cross-country estimation remains difficult. The important new analyses include estimation with instrumental variables (Sect. 5) and consideration of intertemporal changes in growth rates within countries (Sect. 7). More recent U.S. data also permit important refinements to the analysis of cognitive skills on the labor market earnings of immigrants (Sect. 6) previously introduced in Hanushek and Kimko (2000), including the specification of full difference-in-differences models.

Each of our three approaches deals with a particular class of reverse causation or omitted variables. By identifying skill variation stemming from institutional school policies in the countries, the instrumental-variable models highlight the role of schools while addressing issues of simple reverse causality and of inherent cultural difference across nations that might be related to attitudes and performance in learning. By focusing on U.S. labor-market outcomes for immigrants, the difference-in-differences approach deals not only with reverse causality but also with the possibility that cultural differences or economic institutions of national economies may be correlated with favorable educational outcomes. By using the intertemporal dimension of our new database, our longitudinal analysis of changes in growth rates eliminates stable country-specific factors in a general way in the spirit of country fixed

6 Aghion et al. (2009) approach causality by relying on within-country variation.

effects. In each of the three investigations, we explicitly describe the assumptions that are key to interpreting the results. Importantly, the different approaches rely on different assumptions, guard against different threats to identification, and would fail for different reasons.

A related aspect of these separate causal investigations is the pinpointing of a specific policy role for improved school quality. While variations in cognitive skills can arise from various influences—families, culture, health, and ability—the instrumental-variable results indicate that schools, and in particular institutional structures of school systems, are one way for improvement available to policy makers. This conclusion is reinforced by how country of schooling—U.S. versus home country—is important for identifying individual skills in the immigrant analysis.

A final issue addressed is that average test scores do not adequately reflect the range of policy options facing a nation. Specifically, one could institute policies chiefly directed to the lower end of the cognitive distribution, such as the Education for All initiative, or one could aim more at the top end, such as the focused technological colleges of India. In an analysis enabled by the detailed country-specific distributional dimension of our new micro database, we are able to go beyond simple mean difference in scores and provide the first estimates of how growth is affected by the distribution of skills within countries and how it might interact with the nation's technology (Sect. 8). We find improving both ends of the distribution to be beneficial and complementary. The importance of the highly skilled is even more important in developing countries that have scope for imitation than in developed countries that are innovating.

2 A simple growth model with cognitive skills

We begin with a very simple growth model: a country's growth rate (g) is a function of the skills of workers (H) and other factors (X) that include initial levels of income and technology, economic institutions, and other systematic factors. Skills are frequently referred to simply as the workers' human capital stock. For simplicity in Eq. (1), we assume that $H$is a one-dimensional index and that growth rates are linear in these inputs, although these are not important for our purposes.[7]

$$g = \gamma H + \beta X + \varepsilon \tag{1}$$

It is useful at this stage to understand where the skills (H) might come from. As discussed in the extensive educational production function literature (Hanushek 2002), these skills are affected by a range of factors including 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 so forth as in:

$$H = \lambda F + \phi(qS) + \eta A + \alpha Z + \nu \tag{2}$$

The schooling term combines school attainment (S) and its quality (q).

Human capital is nonetheless a latent variable that is not directly observed. To be useful and verifiable, it is necessary to specify the measurement of H. The vast majority of existing

[7] The form of this relationship has been the subject of considerable debate and controversy. As we write it, it can be consistent with both basic endogenous growth models such as Lucas (1988), Romer (1990), and Aghion and Howitt (1998) and neoclassical growth models such as Mankiw et al. (1992). We allow for conditional convergence in the empirical specifications, and the parameters estimated suggest very long transitional periods from any perturbation off of a balanced growth path. We generally cannot adequately distinguish among alternative forms of the underlying growth process. While considering the growth implications of various policy changes, we can, however, investigate directly the sensitivity of GDP projections to the alternative models (see Hanushek and Woessmann 2011b).

theoretical and empirical work on growth begins—frequently without discussion—by taking the quantity of schooling of workers (S) as a direct measure of H.

A more compelling alternative is to focus directly on the cognitive skills component of human capital and to measure H with test-score measures of mathematics, science, and reading achievement.[8] The use of measures of educational achievement has a number of potential advantages. First, they capture variations in the knowledge and ability that schools strive to produce and thus relate the putative outputs of schooling to subsequent economic success. Second, by emphasizing total outcomes of education, they incorporate skills from any source—families, schools, and ability. Third, by allowing for differences in performance among students with differing quality of schooling (but possibly the same quantity of schooling), they open the investigation of the importance of different policies designed to affect the quality aspects of schools.[9]

3 Consistent international measures of cognitive skills

This analysis starts with the development of new measures of international differences of cognitive skills derived from educational achievement tests. We would ideally have measures of the skills for workers in the labor force, but our measures of cognitive skills come from data on testing for students who are still in school. This creates a trade-off: incorporating more recent testing has the potential advantages of improved assessments and observations on a greater number of countries but it also weights any country measures more toward students and less toward workers.[10] We begin with an expansive inclusion of more recent tests but then investigate the impact of this choice through extended robustness checks that take more restrictive choices.

The measures developed here extend those developed in Hanushek and Kimko (2000) to add new international tests, more countries, and intertemporal and within-country dimensions. They also deal with a set of problems that remained with the early calculations.[11]

[8] Some researchers have suggested that test scores should be thought of as a measure of school quality (q), leading to use of test scores times years of schooling as a measure of H, but this ignores the influence of family factors and other elements of Eq. (2) that have been shown to be very important in determining cognitive skills.

[9] Some recent work has introduced the possibility that noncognitive skills also enter into individual economic outcomes (see importantly Bowles et al. 2001; Heckman et al. 2006; Cunha et al. 2006). Hanushek and Woessmann (2008) integrate noncognitive skills into the interpretation of general models such as above and show how this affects the interpretation of the parameter on school attainment and other estimates. While there are no agreed-upon measures of noncognitive skills, at the aggregate level they might well be incorporated in "cultural differences," something that we address in the analysis below.

[10] The reliance on schooling-based measures of skills also makes it clear why it is not possible to employ panel data estimation even though tests are spread across almost four decades for some nations. Any panel study would require measuring the cognitive skills of the labor force at different points in time, something that is not possible with the sporadic measurement of student skills. Only one international test—the International Assessment of Adult Literacy—has suggested the possibility of panel estimation across countries because it has tested adults rather than students (see Coulombe and Tremblay 2006). Nonetheless, such analysis requires very strong assumptions about the mapping of observed age patterns of skills onto changes in labor force skills over time. Further, most of the variance in growth and in test scores is found across countries, not across time for individual countries—suggesting that panel data do not deal effectively with the most acute estimation issues. As shown in Appendix B, the testing has involved voluntary participation by a time-varying group of countries in tests that assess varying subject matters and grade/age ranges of students.

[11] The correlation across the common 30 countries of the new test measures developed here and those in Hanushek and Kimko (2000) is 0.83. Appendix B assesses the importance for growth modeling of the differences in their measures and those developed here.

Between 1964 and 2003, twelve different international tests of math, science, or reading were administered to a voluntarily participating group of countries (see Appendix Table 10). These include 36 different possible scores for year-age-test combinations (e.g., science for students of grade 8 in 1972 as part of the First International Science Study or math of 15-year-olds in 2000 as a part of the Programme on International Student Assessment). Only the United States participated in all possible tests.

The assessments are designed to identify a common set of expected skills, which were then tested in the local language. It is easier to do this in math and science than in reading, and a majority of the international testing has focused on math and science. Each test is newly constructed, usually with no effort to link to any of the other tests.

We wish to construct consistent measures at the national level that will allow comparing, say, math performance of 13-year-olds in 1972 to that in 2003. This would permit us to compare performance across countries, even when they did not each participate in a common assessment, as well as track performance over time. It would also provide the ability to aggregate scores across different years, ages, and even subjects as appropriate. The details of this construction along with the final data are found in Appendix B, and here we simply sketch the methodology. Because the test distribution is normal within the OECD sample, our construction of aggregate country scores focuses on transformations of the means and variances of the original country scores in order to put them each into a common distribution of outcomes.

Comparisons of the difficulty of tests across time are readily possible because the United States has participated in all assessments and because there is external information on the absolute level of performance of U.S. students of different ages and across subjects. The United States began consistent testing of a random sample of students around 1970 under the National Assessment of Educational Progress (NAEP). By using the pattern of NAEP scores for the U.S. over time, it is possible to equate the U.S. performance across each of the international tests.

The comparison of performance of other countries to the U.S. requires a distance metric for each test. Each assessment has varying country participation and has different test construction so that the variance of scores for each assessment cannot be assumed to be constant. Our approach is built on the observed variations of country means for a group of countries that have well developed and relatively stable educational systems over the time period.[12] We create the "OECD Standardization Group" (OSG) by using the thirteen OECD countries that had half or more of the relevant population attaining a secondary education in the 1960s (the time of the first tests). For each assessment, we then calibrate the variance in country mean scores for the subset of the OSG participating to the variance observed on the PISA tests in 2000 (when all countries of the OSG participate). The identifying assumption of this approach is that the *variance* in the mean performance among a group of relatively stable education systems does not change substantially over time.

By combining the adjustments in levels (based on the U.S. NAEP scores) and the adjustment in variances (based on the OECD Standardization Group), we can directly calculate standardized scores for all countries on all assessments. Each age group and subject is normalized to the PISA standard of mean 500 and individual standard deviation of 100 across OECD countries. We can then aggregate scores across time, ages, and subjects as we desire.

12 The development of aggregate scores by Hanushek and Kimko (2000) and by Barro (2001) assumed that the test variances across assessments were constant, but there is no reason for this to be the case. Our approach is in the spirit of Gundlach et al. (2001).

The international testing protocols have evolved over time so that recent assessments employ careful sampling rules, restrictions on the extent of any student exclusions, and modern psychometric testing procedures, while earlier testing less consistently met current standards. This variation in testing quality potentially affects parts of our analysis, because the earlier (but poorer) tests relate to relevant members of the labor force during our period of observation for economic growth. The more recent testing involves students not observed to be in the labor force. As a result, most of the estimation relies upon an assumption that the average scores for a country tend to be relatively stable over time and that the differences among countries are a good index of the relative skill differences of the workforces. This assumption is partially tested below, and, while there are some observed score changes, the overall rankings of countries show considerable stability. For the 693 separate test observations in the 50 countries employed in our growth analysis, 73 % of the variance falls between countries. The remaining 27 % includes both changes over time in countries' scores and random noise from the testing. Our averaging procedure will minimize the noise component at the cost of obscuring any differences over time for each country. In Sect. 7 below, we use the intertemporal variation in scores for the subset of countries with enough observations to estimate the systematic changes as opposed to test noise. For the 15 countries employed in the analysis of score trends in that section, 85 % of the variance lies between countries, and the remaining 15 % within countries will be more heavily the result of systematic trends in scores.

The assessments give cognitive skill measures for tested students. Thus, exclusion rates (say for handicapped children) or differential student enrollment and attendance could affect the estimation. Direct investigation of these issues, at least for tests since 1995 when reporting is sufficient, indicates that the growth analysis is not affected by variations in testing.[13]

4 Stability of the cognitive skills-growth relationship

The basic growth model in Eq. (1) is estimated for the 50 countries with cognitive-skill and economic data over the period 1960–2000. Cognitive skills are measured by the simple average of all observed math and science scores between 1964 and 2003 for each country, although we test the sensitivity of the results to inclusion of varying time periods and subsets of tests. The income data come from version 6.1 of the Penn World Tables (Heston et al. 2002), while the data on years of schooling are an extended version of the Cohen and Soto (2007) data.[14]

We have concentrated on issues surrounding the measurement of cognitive skills, but other questions have recently been raised about the accuracy and reliability of both economic and schooling data. In Appendix D, we investigate the impact of using the latest Barro and Lee (2010) data on school attainment and of substituting the most recent Penn World Table data (version 7.0), which provides additional evidence confirming our basic results through the period up to 2009. Because neither of these alternatives materially affects our results, we simply combine these sensitivity studies in the Appendix.

[13] Hanushek and Woessmann (2011c) show that average tests can be affected by these: greater exclusions and higher enrollment rates are correlated with higher scores. Nonetheless, the variations caused by these factors are orthogonal to growth rates, so they do not bias our estimated skill parameters. The problems are potentially more severe with the earliest tests, but the reported information for these is insufficient for any analysis.

[14] Because we need comparable data on economic growth over the 1960–2000 period, all former communist countries are eliminated even if they have test measures. Appendix B provides details on the country sample, and Appendix C provides descriptive statistics for each of the analyses below.

Table 1 Years of schooling versus cognitive skills in growth regressions

	(1)	(2)	(3)	(4)[a]	(5)[b]	(6)[c]	(7)[d]	(8)[e]	(9)[f]
Cognitive skills		2.015	1.980	1.975	1.933	1.666	1.265	1.239	1.985
		(10.68)	*(9.12)*	*(8.28)*	*(8.29)*	*(5.09)*	*(4.06)*	*(4.12)*	*(7.83)*
Years of schooling 1960	0.369		0.026	0.024	0.025	0.047	0.004	−0.049	−0.090
	(3.23)		*(0.34)*	*(0.78)*	*(0.29)*	*(0.54)*	*(0.05)*	*(0.66)*	*(1.02)*
GDP per capita 1960	−0.379	−0.287	−0.302	−0.298	−0.298	−0.255	−0.351	−0.310	−0.879
	(4.24)	*(9.15)*	*(5.54)*	*(6.02)*	*(5.04)*	*(3.12)*	*(6.01)*	*(5.73)*	*(3.39)*
No. of countries	50	50	50	50	52	50	47	45	50
R^2 (adj.)	0.252	0.733	0.728	0.728		0.706	0.784	0.797	0.637

Notes Dependent variable: average annual growth rate in GDP per capita, 1960–2000. Regressions include a constant. Test scores are average of math and science, primary through end of secondary school, all years. Absolute t-statistics in parentheses

[a] Measure of years of schooling refers to the average between 1960 and 2000

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

[c] Specification includes dummies for the eight world regions depicted in Fig. 1

[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

The central finding of the statistical analysis is the importance of cognitive skills in explaining international differences in long-run growth rates. As a comparison to prior cross-country analyses, the first column of Table 1 presents estimates of a simple growth model with school attainment.[15] While this model explains one-quarter of the variance in growth rates, adding cognitive skills increases this to three-quarters of the variance. The test score is strongly significant with a magnitude that is unchanged by whether initial school attainment in 1960 is excluded (col. 2) or included (col. 3).

School attainment is not statistically significant in the presence of the direct cognitive-skill measure of human capital. This does not change when attainment is measured as the average between 1960 and 2000 (col. 4), rather than at the beginning of the period. This finding, of course, does not mean that schooling is irrelevant. Measured skills are closely related to schooling—a point we emphasize below—but life-cycle skill accumulation depends upon the learning earlier in life. We measure achievement at various points during primary and secondary education. Even if tertiary schooling is simply additive, knowledge at earlier points in education will strongly influence the ultimate skill accumulation when students enter the labor force. But, as James Heckman and his colleagues have emphasized, there is a dynamic complementary of investments such that further schooling has a larger impact on skills if it builds on a larger base developed earlier (Cunha and Heckman 2007). The simple point is that "skill begets skill through a multiplier process" (Cunha et al. 2006, p. 698), such that additional attainment has a lessened impact if built upon lower basic skills.[16] It does suggest

[15] While not the focal point of this analysis, all specifications include GDP per capita in 1960, which provides consistent evidence for conditional convergence, i.e., countries with higher initial income tend to grow more slowly.

[16] Relatedly, a variety of people place extra weight on tertiary education (e.g., Ehrlich 2007). However, without building on strong basic skills, such investment appears to have little extra value. In analysis of growth across both developed and developing countries, tertiary education has little added value in explaining economic growth after consideration of measured cognitive skills on international tests with the exception that U.S. investments in higher education have signaled increased growth (Hanushek and Woessmann 2011b).

that simply investing in further schooling without ensuring commensurate improvements in cognitive skills does not lead to economic returns.

One standard deviation in test scores (measured at the OECD student level) is associated with a two percentage point higher average annual growth rate in GDP per capita across 40 years. (This effect is equivalent to one percentage point per country-level standard deviation, thus making it virtually identical to the more limited estimates in Hanushek and Kimko 2000). Such impacts are clearly large in substantive economic terms, and below we provide alternative perspectives on the magnitude of these effects.

The remaining columns of Table 1 provide alternative perspectives on these basic results. Estimating the model with regression techniques robust to outliers yields virtually identical coefficient estimates to those including Nigeria and Botswana, the two significant outlier countries in the growth equation (col. 5).[17] Because the robust model assigns essentially zero weight to these two observations, they are dropped from the remaining models. Including fixed effects for the eight world regions depicted in Fig. 1 (so that no between-region variation in test scores is used in the estimation) reduces the estimated test effect to 1.7 (col. 6).

Columns 7 and 8 consider economic institutions. Acemoglu et al. (2001, 2005) argue that historic factors surrounding the colonization of nations affected economic institutions and that we can thus isolate the causal impact of institutions on growth. We add institutional differences for openness of the economy and security of property rights (col. 7)[18] and for these plus fertility rates and location in the tropics (col. 8) into our growth models.[19] These reduce the estimated test-score effect to around 1.25, but the effect of cognitive skills remains strongly statistically significant. On the other hand, Glaeser et al. (2004) argue that the colonists brought human capital in addition to knowledge of good societal institutions and that it is more likely that better human capital led both to the development of good institutions and higher economic growth. This latter perspective highlights the difficulty of using measures that reflect economic institutions from near the end of the observed growth period and that might better be thought of as outcomes of growth itself. In their spirit, we interpret the reduced estimates of test scores in columns 7 and 8 as a lower bound on the true effect, since the institutional measures include any direct effects of cognitive skills on the development of good institutions.[20] Additionally (not shown), the stock of physical capital per adult in

Footnote 16 continued
The difficulty is that it is impossible to identify the impact of higher education as opposed to other unmeasured determinants of economic growth in the United States.

17 The specific robust regression technique reported is Stata's *rreg* command, which eliminates gross outliers with Cook's distance measure greater than one and iteratively down weights observations with large absolute residuals. The OLS estimate of the test effect in the 52-country sample is 1.752 (t-statistic 5.75). Nigeria and Botswana each participated only in a single international test.

18 The measure of openness is the Sachs and Warner (1995) index reflecting the fraction of years between 1960 and 1998 that a country was classified as having an economy open to international trade, based on five factors including tariffs, quotas, exchange rate controls, export controls, and whether or not a socialist economy. Following Acemoglu et al. (2001), the measure of security of property rights is an index of the protection against expropriation risk, averaged over 1985–1995, from Political Risk Services, a private company which assesses the risk that investments will be expropriated in different countries.

19 Openness and security of property rights enter the model (jointly) significantly; fertility and tropical location do not.

20 A separate analysis of institutions has evolved in the developed countries of the OECD. These countries all have strong property rights, open economies, and institutions that generally favor economic growth that explain differences between developed and developing countries. In order to explain differences within developed countries, considerable analysis has gone into how various regulations of labor and product markets within the OECD affect economic growth (e.g., Nicoletti and Scarpetta 2003 on product market regulations and Cingano et al. (2010) on labor market regulation). Nonetheless, none of the explicit measures of economic

1960 does not enter the basic growth model significantly and does not affect the test-score coefficient.

Finally, the precise specification of the growth model is the subject of considerable debate within macroeconomics. While there are many nuances of the argument, it can be framed as a simple contrast. The endogenous growth model indicates that increases in human capital can lead to permanent differences in growth rates, because a better-educated workforce leads to a larger stream of new ideas that produces technological progress at a higher rate (e.g., Lucas 1988 or Romer 1990). By contrast, in the augmented neoclassical growth model, changes in human capital lead to higher steady-state levels of income but do not affect the long-run growth path (e.g., Mankiw et al. 1992). Our estimates, which include the level of initial GDP per capita, allow for conditional convergence, but it is difficult to distinguish between temporary "catch-up" growth and long-run differences in growth in the empirical model. Those who favor the neoclassical model, however, favor estimation that includes the log of initial income. Column 9 shows these results, where the impact of cognitive skills on growth is little changed from the linear alternative. Thus, the results do not appear to be the result of the specific empirical model that is estimated.

While the estimated effect of test scores varies some across these different specifications, the cognitive-skill coefficients are always very significant and the variation is quite limited: A move of one standard deviation of individual student performance translates into 1.2–2.0 percentage points difference in annual growth rates, other things equal. How much is one standard deviation in performance? The difference between the U.S. average and the top performers on the PISA tests is approximately 0.4 standard deviations, while the difference between the average Mexican student and the rest of the OECD was approximately one standard deviation.

Two other important questions that relate to interpretation arise. The first set of issues is whether the sample of countries or years of observation heavily influences the results, thus implying that the results are potentially driven by other, unmeasured factors. The second is whether the specific measure of cognitive skills drives the estimates.

Table 2 provides the matrix of estimated cognitive-skill coefficients across different samples of observations. The columns consider sample sensitivity and concentrate on whether the overall results are driven by specific subsets of countries, which might indicate that the cognitive-skill measures simply proxy for other facets of the economies. The top row focuses on the average of all observed math and science scores—as presented previously—while, as explained below, the second row relies on just lower-secondary-school scores which may be a more reliable measure of skill differences. Each entry comes from a separate regression that includes GDP per capita in 1960 and school attainment.

The first two comparisons (col. 2–3 and col. 4–5) present evidence on whether cognitive skills are more or less important in developed countries. The first comparison divides the estimation into the 23 OECD countries and 27 non-OECD countries, while the second comparison divides countries into above and below the median level of per-capita GDP in 1960. The statistically significant difference of high-income and lower (below median)-income countries indicates that developing countries are somewhat more affected by cognitive skills than developed countries.[21] This larger impact of skills in low-income countries is consistent

Footnote 20 continued
regulations identified in this work is useful at explaining the large differences in long-run growth rates among OECD countries (Hanushek and Woessmann 2011b). Thus, at least for the developed countries, omission of this expanded set of economic institutions does not appear to bias our growth estimates.

[21] While not shown, the school attainment measures are insignificantly related to growth even among the developing countries where the levels are low and where there is considerable cross-country variance.

Table 2 Sensitivity of estimated effects of cognitive skills to the sample of countries and time periods

Country/year sample	(1) Full	(2) OECD	(3) Non-OECD	(4) High-income[a]	(5) Low-income[a]	(6) W/o East Asia	(7) 1960–1980	(8) 1980–2000	(9) 1980–2000[b]	(10) Score-schooling outliers[c]	(11) Score-schooling core[c]
Test-score specification											
All math and science	1.980	1.736	2.056	1.287	2.286	1.301	1.522	2.996	3.782	1.888	2.175
	(9.12)	*(4.17)*	*(6.10)*	*(5.37)*	*(6.98)*	*(4.90)*	*(4.29)*	*(9.42)*	*(3.11)*	*(7.81)*	*(3.47)*
Only lower secondary	1.759	1.646	1.792	1.040	2.083	1.137	1.407	2.580	4.386	1.673	1.887
	(9.22)	*(4.02)*	*(6.19)*	*(4.70)*	*(7.44)*	*(4.82)*	*(4.56)*	*(8.88)*	*(4.49)*	*(7.83)*	*(3.45)*
No. of countries	50	23	27	25	25	40	50	50	25	25	25

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. Control variables: Initial GDP per capita, initial years of schooling, and a constant. Test scores: Unless noted otherwise, average of math and science, primary through end of secondary school, all years. Absolute t-statistics in parentheses

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

[b] Test scores refer only to tests performed until 1984

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

with the arguments by Glaeser et al. (2004) that nearly all poor countries in 1960 were dictatorships, some of which developed better societal institutions as an outcome of growth rather than a cause. The countries that did better in terms of growth were those with higher human capital, supporting the larger coefficient on human capital in the poor countries. Nonetheless, variations in math and science skills remain very important in distinguishing among growth rates of the developed countries.

A portion of the influence of cognitive skills comes from the high growth of East Asian countries. As shown in column 6, excluding the ten East Asian countries lowers the estimated impact of math and science skills to 1.3, but it remains highly significant in the remaining countries. In other words, the overall estimates, while influenced by the East Asian growth experience, are not simply identifying the high growth—high test-score position of East Asia, which would raise the possibility that the growth relationships might be driven by other factors that were simply correlated with East Asian test performance.

The growth estimates are meant to identify long-run factors, but the sample period of 1960–2000 includes sub-periods of world stagnation, fast growth, and financial crises. Some have suggested, for example, that the observed growth rates are dominated by the early-period growth explosion of East Asia and that this changed considerably with the financial crises of the late 1990s (Ramirez et al. 2006). Our results (col. 7 and 8) indicate, however, a consistent impact of cognitive skills across the period that, if anything, has grown stronger in the second half of our observations. Indeed, the estimated impact doubles in the most recent period, consistent with various arguments that, at least for the U.S. and OECD countries, the importance of skills has increased (Murnane et al. 1995; Katz and Autor 1999; Goldin and Katz 2008).

The analysis has relied on assessments given throughout the period of economic observation. This choice is made to maximize the number of countries and to include the more precise testing of recent periods, but it raises questions of reverse causality. If greater growth provides added resources that can be used to improve schools and test scores, our estimates could suffer from simultaneity bias. One direct set of estimates addresses this issue.[22] The same impact on 1980–2000 growth is found when we restrict the test scores to measures obtained before 1985 (available for only 25 countries), i.e., when we use test scores nearly fully pre-dating the growth period (col. 9). In fact, the point estimate for cognitive skills becomes substantially larger in this specification. By using pre-determined test scores, this specification excludes the possibility of simple reverse causation. The conclusion that simple reverse causation is not driving the results is reinforced in analyses using data updates that extend the economic series to 2009 (Appendix D).[23] The possibility of reverse causation from economic growth to test scores is also unlikely because additional educational spending (which might become affordable with higher growth) does not systematically relate to better test scores (e.g., Hanushek 2002).

Levels of schooling and cognitive scores are correlated across our sample ($r = 0.62$), in part because of the differences between developed and developing countries. Still, in many cases countries with similar years of schooling have very different test scores (see Appendix Fig. 6). The separation of the impact of cognitive skills from that of school attainment in our estimation relies upon information where these two diverge, and it might be a peculiar set of countries in terms of growth where the pattern of school attainment and skills varies most.

---

22 A second set of analyses directly addresses the resource question and finds that international tests are not driven by differences in resources across countries. See the review in Hanushek and Woessmann (2011a).

23 Results using only test scores that pre-date the analyzed growth period (not shown) are also robust when combined with our other robustness checks pursued in Table 1.

The final two columns divide countries based on deviations of cognitive scores from school attainment. Specifically, the "score-schooling outliers" are the 25 countries with the largest residuals when test scores are regressed on attainment, and the "score-schooling core" are the 25 with the smallest residuals. Interestingly, the relationship between cognitive skills and growth is virtually the same across these two samples, revealing that the results are not driven by "peculiar" countries in the production of cognitive skills.

The preceding results hold looking across columns, but the pattern also obtains for the alternative measures of test scores. The estimated coefficients using only lower-secondary-school math and science scores are systematically a little smaller than those from all scores, which may reflect attenuation bias when using fewer test observations in the construction of the cognitive-skill measure, but there are no changes in patterns across any of the columnar comparisons. This test-score measure excludes any test in primary schooling or in the final year of secondary education. Test scores at the end of the secondary level, which combine the knowledge accumulated over primary and secondary schooling, may be most relevant for the labor force, but, at the same time, the duration of secondary education differs across countries, so that tests performed in the final year of secondary schooling may not be as readily comparable across countries. Further, given differing school completion rates, tests in the final year of secondary schooling may produce samples with differential selectivity of test takers. Yet neither the primary-school tests nor the tests in the final secondary year are crucial for the results.[24]

Table 3 provides more detail on sensitivity to the measure of cognitive skills, comparing several additional plausible alternatives for the aggregation of scores, including using math, science, and reading scores separately. We also provide breakdowns by OECD and non-OECD countries, although this breakdown makes little qualitative difference, and we concentrate on the variations in aggregate test information found in the table rows.

Results are qualitatively the same when using only scores on tests performed since 1995 (row A). These recent tests have not been used in previously available analyses and are generally viewed as having the highest standard of sampling and quality control. Likewise, results are robust to using tests scores since 1995 for just lower secondary grades (row B).

A drawback of using only the more recent tests is that such an approach requires a strong version of the assumption that test performance is reasonably constant over time, because it relates test performance measured since 1995 to the economic data for 1960–2000. To make sure that higher previous economic growth is not driving the measured test performance, the test-score measure used in row C disregards all tests since the late 1990s. Our results turn out to be robust, with a point estimate on the test-score variable that is significantly higher (although the sample is reduced to 37 countries). Our results are also robust to using the average early test scores as an instrument for the average of all test scores in a two-stage least-squares regression, in order to utilize only that part of the total test-score measure that can be traced back to the early test scores (row D). In sum, the results do not appear to be driven by either early or late test scores alone.

The remainder of the table investigates different combinations of the math, science, and reading tests. While we were concerned about the reliability of the reading tests and thus have focused on math and science, the use of reading tests provides similar results in the growth models (rows E-G). In a specification that enters the different subjects together (panel H), the three are always jointly significant at the 1 % level and higher, even though the science effect gets smaller and the reading effect loses significance in the joint model.

[24] Hanushek and Woessmann (2011c) also provide direct analysis of potential biases from test exclusions and enrollment rates and find that they are not an important concern.

Table 3 Sensitivity of estimated effects of cognitive skills to the measurement of skills

Country sample	(1) Full	(2) OECD	(3) Non-OECD	No. of countries
Test-score specification				
(A) Only since 1995	1.814	1.473	1.850	47
	(9.91)	*(3.80)*	*(6.74)*	
(B) Only lower secondary since 1995	1.644	1.379	1.657	47
	(9.57)	*(3.49)*	*(6.48)*	
(C) Only until 1995	3.156	1.377	3.668	37
	(6.57)	*(1.93)*	*(4.44)*	
(D) Early as instrument for average[a]	2.341	1.212	2.915	37
	(7.71)	*(1.98)*	*(5.80)*	
(E) Only math	2.009	1.529	2.063	47
	(8.98)	*(4.62)*	*(5.81)*	
(F) Only science	1.576	1.769	1.556	50
	(7.00)	*(3.28)*	*(4.48)*	
(G) Only reading	2.351	1.616	2.529	46
	(6.21)	*(3.33)*	*(3.68)*	
(H) All subjects entered jointly				41
Math	1.662	2.270	1.882	
	(3.69)	*(2.97)*	*(1.97)*	
Science	1.007	−2.414	1.270	
	(2.34)	*(1.62)*	*(1.92)*	
Reading	−0.793	1.333	−1.457	
	(1.15)	*(1.44)*	*(0.94)*	

Notes Reported numbers are the coefficient on test scores in each model specification. Dependent variable: Average annual growth rate in GDP per capita, 1960–2000. Control variables: GDP per capita 1960, years of schooling 1960, and a constant. Test scores: Unless noted otherwise, average of math and science, primary through end of secondary school, all years. Absolute t-statistics in parentheses

[a] 2SLS with average of test scores until 1995 as instrument for average of all test scores

The overall picture from this sensitivity analysis is that the estimated effect of cognitive skills on growth is quite robust to a range of samples, specifications, and measurements. This finding contrasts sharply with many previous analyses that use years of schooling as the human capital measure, beginning with Levine and Renelt (1992) and continuing through Pritchett (2006). But of course the similarity of findings, while ruling out some specification and measurement issues, cannot guard against all plausible threats to the identification of causal growth relationships.

The main theme of this paper is that cross-sectional growth regressions using existing variation across countries provide stylized facts about long-term development but that their interpretation may be hampered by endogeneity biases. Endogeneity of cognitive skills could, for example, arise because nations with conditions favorable to economic growth also produce high test performance. This correlation could arise because cultural factors, historically good economic institutions, variations in health status, or any other set of factors that lead to strong economic performance might also be systematically related to high cognitive skills. Indeed, it does not matter whether such relationships are causal or purely associational. If these factors are omitted from the growth estimation, they will tend to bias the coefficient on cognitive skills. Likewise, as suggested previously, there might be reverse causality if economic growth facilitates investments in the school system or increases family resources that improve cognitive skills.

The following three sections approach the interpretation of these stylized facts about growth from different viewpoints. Each approach to dealing with potential interpretative problems is clearly inconclusive, but each does work to eliminate specific sets of concerns. These further analyses also highlight the potential for policies based upon improved school quality.

5 Variations in cognitive skills driven by schools: instrumental variable models

One general concern in cross-country investigations is that cultural features influence both economic behavior and school outcomes. Even if the cognitive skills-growth relationship is causal, the results presented so far would only be relevant for school policy if the variation in cognitive skills emanating from school policies is in fact related to economic growth. As noted, cognitive skills are likely to depend not only on formal schooling but also on non-school factors such as families, peers, and ability. Therefore, it is important to establish any links with school policy levers.

One means of addressing the set of issues is to use measures of the institutional structure of the school systems as instruments for the cognitive-skill measure, thereby focusing only that part of the international variation in cognitive skills that can be traced back to international differences in school systems. We use several institutional features—notably the existence of external exit exam systems, the share of privately operated schools, the impact of varying Catholic church history, the centralization of decision-making, and relative teacher pay—that have been shown in the literature on international educational production to be associated with student achievement (see Hanushek and Woessmann 2011a for a review and evaluation of the micro evidence).

While other school policies such as those surrounding educational spending levels may well be endogenous to the growth process, these institutional features can plausibly be assumed uncorrelated with the regression disturbances of our growth models. First, many educational institutions such as the existence and extent of private schooling reflect long-standing polices embedded in education law and thus are not outcomes of the growth process per se (see, for example, the review of private schooling across countries in Glenn and De Groof 2002).[25] Second, while there have been some trends in these institutions—such as the slow movement toward decentralizing school decision-making—there is no suggestion that this reflects either growth or other systematic differences in cultural and economic systems.[26] Third, there is empirical support from the literature on educational production that these institutional effects on student learning are robust to including regional fixed effects in cross-country analyses, to within-country analyses, and to the use of historical instruments (see Woessmann 2003a; West and Woessmann 2010; Hanushek and Woessmann 2011a). These results suggest that institutional impacts are not driven by cultural differences and do not suffer directly from reverse causality.

25 If private school attendance rates are also related to the religious composition of countries, this instrument might still by problematic because of religious impacts on economic behavior; see Barro and McCleary (2003). However, evidence in West and Woessmann (2010) shows that, if anything, such impacts would be likely to lead to an underestimation of the effect of private school competition.

26 Glenn and De Groof (2002, p. 267), note that "there has been in most Western democracies a slow but very marked shift in the allocation of responsibility for the organization and control of education, in the public as well as the nonpublic education sector, through decentralization of various aspects of decision-making to the local school community." The cross-country details suggest no obvious political or cultural differences in these trends.

External exit exam systems are a device to increase accountability in the school system that has been repeatedly shown to be related to better student achievement (see Bishop 2006 for a review).[27] The first specification reported in Table 4 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. The first-stage results confirm a statistically significant association between external exit exams and cognitive skills. The effect of cognitive skills on economic growth in the second stage of the instrumental variable (IV) estimation is statistically significant and close to the OLS estimate.[28] However, the relatively low F-statistic of the instrument in the first stage indicates the possibility of a weak instrument problem. Instruments that are only weakly correlated with the endogenous explanatory variable may actually increase estimation bias and compromise the reliability of the conventional asymptotic approximations used for hypothesis testing. Thus, we also report estimates based on the modification of the limited information maximum likelihood (LIML) estimator by Fuller (1977), but the results are hardly affected.[29] While the confidence band of the conditional likelihood ratio test proposed by Moreira (2003) and Andrews et al. (2007) gets large at the upper end in this specification, difference from zero still reaches significance at the 10 % level.[30]

Because initial years of schooling are insignificant in the growth model once test scores are controlled for (both in the OLS and in the IV specification), another possibility is to include years of schooling as a second instrument for test scores.[31] This approach is also suggested by the prior model as long as cognitive skills are a measure of human capital in Eq. (1). Specification (2) of Table 4 reveals that years of schooling are significantly associated with test scores in the first stage, and the first-stage F-statistic increases substantially. The Sargan test does not reject the overidentifying restrictions of the model, suggesting that, if external exit exams are a valid instrument, years of schooling are also valid. Both the 2SLS and the Fuller estimates, as well as inference based on Moreira confidence bands, confirm that schooling-induced differences in cognitive skills are significantly related to economic growth.

School choice, as measured by the share of privately operated schools in a system, consistently shows a positive association with student achievement in OECD countries (see the review in Woessmann et al. (2009), along with West and Woessmann (2010)) and provides an additional instrument. In our sample, the share of private enrollment in a country is significantly positively associated with cognitive skills in the first stage of our IV model (specification (3) of Table 4).[32] The second-stage estimate of the growth model confirms our

[27] Data on external exit exams are available for 43 countries in Woessmann et al. (2009), who update Bishop (2006)'s collection from reviews of comparative-education studies, educational encyclopedia, government documents, background papers, and interviews with national representatives. The measure refers roughly to the mid-1990s, but exam regimes are relatively stable over time for countries.

[28] The Durbin–Wu–Hausman test does not reject the exogeneity of cognitive skills at conventional levels.

[29] Fuller's modification of the LIML estimator is more robust than 2SLS in the presence of weak instruments and performs relatively well 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.

[30] Likewise, the Anderson–Rubin χ^2 statistic (3.06) of this just-identified model indicates significance at the 8 % level. Note that the LIML estimators, around which the Moreira bands are centered, differ from the reported 2SLS estimates only in the third digit in all our models.

[31] School attainment will also be affected by enrollment in higher education, which is not explicitly modeled. The results here suggest that international differences in cognitive skills still remain the dominant factor in growth.

[32] The data on private enrollment as percentage of total enrollment in general secondary education are from UNESCO (1998) and refer to 1985, the earliest year with consistent data. For greater consistency of the time

Table 4 From schooling institutions to cognitive skills to economic growth: instrumental variable estimates

	(1)	(2)	(3)[a]	(4)	(5)[a]	(6)
Second stage						
2SLS						
Cognitive skills	2.151	2.023	2.978	2.207	3.914	1.749
	(2.73)	*(5.81)*	*(5.84)*	*(6.54)*	*(4.17)*	*(5.77)*
Catholic share in 1970				0.003		
				(0.01)		
Fuller (1) modification of LIML						
Cognitive skills	2.121	2.022	2.969	2.197	3.797	1.753
	(3.01)	*(5.94)*	*(5.93)*	*(6.64)*	*(4.17)*	*(5.92)*
Moreira 95 % confidence band						
Cognitive skills	[−3.888, 19.871]	[1.190, 2.868]	[1.734, 4.343]	[1.465, 3.093]	[2.063, 7.006]	[0.865, 2.525]
p-value	*(0.100)*	*(0.001)*	*(0.0004)*	*(0.0001)*	*(0.0000)*	*(0.007)*
First stage (dependent variable: cognitive skills)						
External exit exam system	0.286	0.286				
	(2.01)	*(2.01)*				
Initial years of schooling		0.176	0.137	0.186	0.065	0.161
		(4.11)	*(4.19)*	*(4.32)*	*(2.06)*	*(3.05)*
Private enrollment share			0.520			
			(2.36)			
Centralization (share) of decisions on organization of instruction					−0.941	
					(3.24)	
Catholic share in 1900				2.301		
				(2.15)		
Relative teacher salary						0.188
						(2.19)
Catholic share in 1970				−2.801		
				(2.46)		

Table 4 continued

	(1)	(2)	(3)[a]	(4)	(5)[a]	(6)
No. of countries	43	43	20	50	18	34
Centered R^2	0.752	0.753	0.791	0.743	0.590	0.819
First-stage F-statistic	4.04	10.28	12.15	10.60	13.35	6.94
Sargan statistic		0.033	0.158	0.193	0.011	0.377
p-value		*(0.856)*	*(0.691)*	*(0.661)*	*(0.917)*	*(0.540)*
Durbin–Wu–Hausman χ^2 test	0.034	0.003	0.113	0.479	4.744	0.081
p-value	*(0.855)*	*(0.957)*	*(0.737)*	*(0.489)*	*(0.029)*	*(0.776)*

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

[a] Dependent variable: average annual growth rate in GDP per capita, 1980–2000; sample of OECD countries

previous results—schooling-induced differences in cognitive skills are significantly related to economic growth. Again, the Sargan test does not reject the validity of the overidentifying restrictions, and the Durbin–Wu–Hausman test presents no evidence of endogeneity of the cognitive-skill measure. Results are also very similar without years of schooling as a second instrument.

An additional way to exploit the effect of private competition is to use the historical origins of international variation in the size of the private school sector.[33] In particular, West and Woessmann (2010) show that the opposition of 19th century Catholic church doctrine to state schooling provides a natural experiment in that the share of Catholics in a country's population in 1900 is associated with the share of privately operated schools in its current school system, even after controlling for current Catholic shares. In this spirit, specification (4) of Table 4 uses the Catholic share in 1900 as an instrument for the cognitive skill measure and controls for the Catholic share in 1970 in both stages of the IV model to ensure that results are not driven by effects of religious affiliation per se on cognitive skills and on economic growth (see Hanushek and Woessmann (2012) for details). The first stage shows that historical Catholic shares are indeed positively related to cognitive skills, whereas the opposite is true for modern Catholic shares (which are insignificant in the second stage). This IV specification, which we can estimate for our full sample of 50 countries, confirms a significant effect of cognitive skills on economic growth that is very close to the OLS estimate. The F-statistic of the instruments in the first stage is just above 10, and LIML estimates and Moreira bands confirm that the result is not driven by weak instruments problems.[34]

A further institutional feature regularly shown to be positively associated with student achievement is the extent to which schools (or at least local decision-makers) are autonomous to make their own decisions about the organization of instruction (see Woessmann 2003b). Specification (5) of Table 4 shows that the share of decisions on the organization of instruction that are made at the central government level is significantly negatively associated with our cognitive-skill measure. The second-stage estimators, robust to potentially weak instruments, confirm the significantly positive effect of cognitive skills on economic growth.[35]

Finally, given the crucial importance of teacher quality in educational production, our final IV specification uses the *relative* position of teacher salaries in the income distribution of a country as an instrument for cognitive skills (see Hanushek and Woessmann 2012 for details). Following Dolton and Marcenaro-Gutierrez (2011), we use teacher salaries relative to per-capita income as a proxy for the overall quality of the teaching force in a cross-country

Footnote 32 continued
spans, the dependent variable in this specification is economic growth in 1980–2000; results are robust to using growth in 1960–2000. Given that the results from the educational production literature mostly refer to the sample of OECD countries, we restrict the analysis to the OECD sample, for which 20 observations are available.

[33] The idea for this IV specification, as well as the IV specification using relative teacher salaries, was first presented by Hanushek and Woessmann (2012) in the context of analyzing the Latin American growth puzzle; that reference also provides additional detail on these two IV specifications.

[34] While the F-statistic of the instrument in the first stage is lower (at 4.61) in a model that does not use years of schooling as a second control, such a specification also confirms the significant main result and is also robust to LIML estimation and to the use of Moreira bands.

[35] Data on the percentage of decisions on the organization of instruction in public lower secondary education taken at the central level of government are available in Organisation for Economic Co-operation and Development (1998), available only for 1998. The IV results are very similar without using years of schooling as a second instrument, and the F-statistic of the excluded instrument is already above 10. In this specification, the estimated growth effect is even larger than the OLS estimate. Note, though, that the Fuller estimate is already closer to the OLS estimate, and the Moreira confidence bands include the OLS and other IV estimates.

perspective, which targets the point in the overall "ability" distribution from which a country draws its teacher population and circumvents issues of salary *levels* being themselves related to growth.[36] Again, the results confirm the significant growth effect of cognitive skills, in the same order of magnitude as the OLS estimates, as well as the robustness to LIML estimation and Moreira bands (specification (6) of Table 4).

One potential worry about the exogeneity of our instruments is that the institutional features of school systems may be correlated with economic institutions, which are themselves correlated with economic growth. To test whether this affects our identification, we add the two measures of differences in economic institutions that tend to enter most robustly in growth regressions—openness and security of property rights—to our IV models (remembering, however, our prior reservations about the distinct possibility that these economic institutions capture part of the human capital effect). Our basic result is unaffected. In fact, the measures of economic institutions do not enter significantly (individually or jointly) in any of the IV models except column 2, and the effect of cognitive skills remains significant in all specification except for the just-identified model (1). The point estimates for cognitive skills are hardly affected except for model (2), where—in line with the OLS results of Table 1—it is reduced to 1.1 (and to 1.3 in model (4)), similar to our OLS estimate of the lower bound for the effect.

The results suggest that improvements in cognitive skills generated in the school system—through institutional features affecting school quality—lead to higher long-run growth of economies. There are obvious limits of cross-country regressions with small data samples, and these are particularly salient in IV specifications. Caution is appropriate in interpreting IV results for our relatively small samples of countries and employing the aggregate nature of the institutional measures, but these cautions also make the statistical significance, reasonable precision, and quantitative robustness of the results based on quite different instruments even more striking.[37]

A significant concern remains, however. The institutional characteristics of the school system might still be related to important unmeasured aspects of economic institutions (either causally or correlationally). Nonetheless, any such problems must go beyond the traditional measures of differences in economic institutions that are commonly employed.

6 Comparing the impacts of U.S. and home-country education on the U.S. labor market

An alternative approach for assessing the causal importance of schools and of our measured skill differences on economic outcomes relies on microdata on earnings differences within a single labor market—the U.S. labor market. This strategy, first proposed by Hanushek and Kimko (2000), looks within a single labor market, thereby explicitly holding constant the quality of economic and cultural factors affecting the operations of the economy and focusing on whether measured cognitive skills directly relate to productivity.[38] Following a difference-

[36] The teacher salary data, based on OECD and UNESCO surveys, refer to teacher salaries at the top of the experience scale in 2003.

[37] The IV results hold when employing several of the institutional instruments jointly in the specification, but only one of them tends to capture statistical significance in the joint specifications.

[38] This analysis extends the original work in Hanushek and Kimko (2000) in several ways. By placing the analysis in the framework of a difference-in-differences model, it compares the earnings of late immigrants just to early immigrants from the same country; it dramatically expands both the sample of workers and the number of countries of origin for immigrants; it uses better test information for the comparisons; and it

in-differences strategy, we can compare the returns to skills of immigrants schooled in their country of origin to those of immigrants from the same country schooled within the United States. If it is the measured differences in cognitive skills and not other economically relevant attributes of the families and economies that are important, the impact of skills can be derived from the different earnings of immigrants who received their schooling at home and in the United States.

The structure of the estimation is derived from a standard Mincer (1974) wage equation augmented by measured cognitive skills such as:

$$\ln y_{ic} = \alpha_0 + \alpha_1 S_{ic} + \alpha_2 PE_{ic} + \alpha_3 PE_{ic}^2 + \gamma_y H_{ic} + \upsilon_{ic} \tag{3}$$

where y is annual earnings for immigrant i from country c, S is years of school attainment, *PE* (=age-*S*-6) is potential experience, H is cognitive skills, and υ is a random error.[39]

We look at immigrants to the U.S. who were either educated entirely in their country of origin or entirely in the United States.[40] (This excludes any individuals partially educated in both the U.S. and their home countries in order to obtain a clear separation of treatment and control groups.) We assign the average cognitive-skill score of the home country ($\overline{T}_c$) for each immigrant and estimate the Mincer earnings Eq. (3) as:

$$\begin{aligned}\ln y_{ic} = {} & \alpha_0 + \alpha_1 S_{ic} + \alpha_2 PE_{ic} + \alpha_3 PE_{ic}^2 \\ & + [\alpha_4 ORIGIN_i + \delta \overline{T}_c + \delta_O (\overline{T}_c \times ORIGIN_i)] + \upsilon_{ic}\end{aligned} \tag{4}$$

where *ORIGIN* is an indicator that is one if immigrant i was educated entirely in schools in the country of origin and zero otherwise and the combined terms in brackets indicate the skills of individuals from country c. The parameter δ_O is the relevant contrast in skills between home-country schooling and U.S. schooling. We interpret δ_O as a difference-in-differences estimate of the effect of home-country test scores on earnings, where the first difference is between home-country educated immigrants (the "treatment group") and U.S.-educated immigrants (the "control group") from the same country, and the second difference is in the average cognitive-skill score of the home country.[41] The parameter δ captures the bias that would emerge in standard cross-sectional estimates from omitted variables like cultural traits that are correlated with home-country test scores in the same way for all immigrants from

Footnote 38 continued
considers a range of sensitivity analyses such as excluding Mexican immigrants and including only immigrants from English-speaking countries.

39 Given that our growth specifications are most closely related to an endogenous-growth formulation, one cannot directly go from the estimated effects on individual productivity to the impact on growth rates. For that reason, in order to validate our micro estimates below, we concentrate on comparisons with other micro estimates of the impact of cognitive skills on productivity. While our analysis uses skill differences by country of origin to infer earnings differences among immigration in the U.S., Hendricks (2002) and Schoellman (2012) go the opposite way of using earnings differences of U.S. immigrations to infer cross-country differences in human capital.

40 Immigrants are individuals born in a foreign country. The sample includes all individuals age 25 or older currently in the labor force with wage and salary earnings of at least \$1,000 and not enrolled in school. Included immigrants had to have been born in a country with international test data (see Appendix B). The number of included countries is larger than in the previous growth regressions because of the lack of need to have internationally comparable GDP data for country of origin. Descriptive statistics are found in Appendix Table 13.

41 Immigrants educated in their home country necessarily come to the U.S. at an older age than comparable immigrants educated in the U.S., suggesting that there might be differential selectivity and motivation of these two groups. But the key issue for identifying the impact of cognitive skills is that any selectivity in migration is the same across countries (which would then be captured by α_4), or at least is not correlated with differences in home-country cognitive skills.

the same country of origin (independent of where they were educated); in our more elaborate specifications with country-of-origin fixed effects, this parameter is not identified.

The first two columns of Table 5 report the estimates of the impact of cognitive skills from stratified samples for the two groups of immigrants. Test scores are normalized to mean zero and a standard deviation of one, so that the estimates indicate the proportionate increase in earnings from a one standard deviation increase in scores. Other things equal, there is essentially no relationship of U.S. earnings to scores of their country of origin, either quantitatively or statistically, for the 50,597 immigrants educated entirely in the U.S. On the other hand, one standard deviation greater performance in country-specific average test scores translates into a statistically significant earnings increase of approximately 16 % for the 258,977 immigrants educated in their country of origin.

This estimate is surprisingly close to recent estimates for cognitive skills of U.S. workers, which indicate 10-15 % returns to a standard deviation of test scores for young workers and 19 % across the full age range of workers.[42] The closeness to the various estimates is surprising given that just average country scores as opposed to individual specific scores are used in the estimation here, although the averaging of scores does eliminate the measurement error found in individual test data.

Column 3 combines the samples and fully estimates Eq. (4). These estimates indicate a significant impact of test scores emanating from schooling in the immigrant's country of origin (δ_O). In contrast, the estimate of (home-country) test score for U.S.-educated immigrants (δ) is statistically insignificant, although the point estimate is noticeably greater than zero. Column 4 demonstrates that this latter effect comes entirely from the influence of immigrants from Mexico (who constitute 37 % of all immigrants to the U.S.). The estimation for immigrants from Mexico is prone to classification error, because many Mexican families tend to move back and forth from Mexico—thus making assignment to U.S. or Mexican schooling prone to error.[43] Excluding Mexican immigrants, $\widehat{\delta_O}$ is highly significant with a point estimate of 0.13, while the coefficient for U.S.-educated immigrants falls to -0.026 and remains statistically insignificant.

The prior estimates indicate that the estimation strategy might be sensitive to variations in immigration patterns across the 64 sampled countries. For example, in addition to the complications for Mexican immigrants, the immigrants from other countries might vary by where they come in the ability distribution of the home country and the like. For this reason, the remaining columns of Table 5 contain country-of-origin fixed effects. Thus, immigrants educated entirely abroad in their home country are compared directly to immigrants *from the same country* educated entirely in the U.S. This eliminates any potential bias emanating from features specific to the country of origin, be they specific selectivity of the immigrant population or country-specific cultural traits. The only remaining assumption required for identification of our parameter of interest is that any potential difference between the early-

42 Murnane et al. (2000) provide evidence from the High School and Beyond and the National Longitudinal Survey of the High School Class of 1972. Their estimates suggest some variation with males obtaining a 15 % increase and females a 10 % increase per standard deviation of test performance. Lazear (2003), relying on a somewhat younger sample from NELS88, provides a single estimate of 12 %. These estimates are also very close to those in Mulligan (1999), who finds 11 % for the normalized AFQT score in the NLSY data. Hanushek and Zhang (2009) estimate a return of 19 % from the International Adult Literacy Survey, which samples workers aged 16–65.

43 The assignment of individuals to U.S. schooling is based on census data indicating immigration before age 6. The assignment of individuals to schooling all in country of origin is based on age of immigration greater than years of schooling plus six. A person who moves back and forth during the schooling years could be erroneously classified as all U.S. or no U.S. schooling, even though they are really in the partial treatment category (which is excluded from the difference-in-differences estimation).

Table 5 Difference-in-differences estimates of returns to country-of-origin cognitive skills for U.S. immigrants

Sample:	U.S. educated[a]	Educated in country of origin[b]	All immigrants	W/o Mexico	All immigrants	W/o Mexico	Growth sample[c]	Only English speaking countries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cognitive skills × educated in country of origin			0.0873	0.1324	0.1375	0.1398	0.1670	0.1616
			(2.02)	*(3.31)*	*(3.16)*	*(4.13)*	*(3.77)*	*(3.57)*
Cognitive skills	0.0050	0.1582	0.0634	−0.0258	Not identified	Not identified	Not identified	Not identified
	(0.14)	*(2.37)*	*(1.06)*	*(1.42)*				
Educated in country of origin			−0.1385	−0.1011	−0.1298	−0.0626	−0.1309	−0.0206
			(3.95)	*(3.03)*	*(2.98)*	*(2.07)*	*(2.58)*	*(0.83)*
Years of schooling	0.1155	0.0673	0.0700	0.0863	0.0579	0.0856	0.0553	0.0992
	(14.08)	*(7.35)*	*(7.43)*	*(13.47)*	*(4.14)*	*(17.40)*	*(4.06)*	*(15.41)*

Table 5 continued

Sample:	U.S. educated[a] (1)	Educated in country of origin[b] (2)	All immigrants (3)	W/o Mexico (4)	All immigrants (5)	W/o Mexico (6)	Growth sample[c] (7)	Only English speaking countries (8)
Potential experience	0.0372	0.0235	0.0243	0.0215	0.0241	0.0227	0.0233	0.0205
	(19.71)	*(4.12)*	*(5.67)*	*(4.75)*	*(7.68)*	*(5.57)*	*(6.50)*	*(2.80)*
Potential experience squared	−0.00064	−0.00035	−0.00036	−0.0004	−0.00039	−0.0004	−0.0004	−0.0004
	(13.02)	*(4.79)*	*(6.10)*	*(4.90)*	*(7.40)*	*(5.87)*	*(6.46)*	*(3.25)*
Fixed effects for country of origin	no	no	no	no	yes	yes	yes	yes
Observations	50,597	258,977	309,574	187,506	309,574	187,506	273,213	72,091
No. of countries	64	64	64	63	64	63	47	12
R^2	0.157	0.170	0.180	0.132	0.196	0.150	0.202	0.156

Notes Dependent variable: log(annual earnings). Cognitive skills refer to average test score of country of origin (centered at zero). Sample: All immigrants identified by country of birth not in school whose age is greater than 25, who are employed, and who earned more than $1,000 in 1999. Immigrants who had obtained some but not all of their education in the U.S. were excluded from the sample. Immigrants from all countries of origin for which there are cognitive-skill scores, except for the following countries (areas) which could not be identified because of census restrictions on release of data for small cells: Swaziland, Slovenia, Macau-China, Luxembourg, Liechtenstein, Estonia, Botswana, Bahrain, Tunisia, and Iceland. Israel could not be identified separately from Palestine; both were assigned the Israeli score. Robust absolute values of t-statistics in parentheses with clustering by country of origin. Source: Authors' calculations from 2000 Census IPUMS data

[a] U.S. educated immigrants are identified as immigrating to the U.S. before the beginning year of schooling

[b] Immigrants educated in their country of origin are identified as immigrating to the U.S. after the final year of schooling

[c] The economic growth sample relies on the data for immigrants from the 50 countries in the basic growth regressions

immigrated U-S.-educated and the late-immigrated home-educated group of immigrants from each country (as captured by the *ORIGIN* indicator) does not vary across countries in a way associated with country-of-origin test scores.

Column 5 displays the primary estimation across all sampled countries with country-specific fixed effects. The estimated impact of cognitive skills is a 14 % increase in earnings from each standard deviation increase in origin-country test scores (when educated there). This estimate is highly significant. Further, the point estimate is virtually unchanged by excluding the Mexican immigrants (col. 6). The standard error is reduced by clearer assignment to treatment category (when Mexicans excluded), even though the sample is substantially reduced.

The final two columns investigate the sensitivity of these estimates to sample definition. First, our estimation of growth models used the 50 countries for which we could obtain the relevant economic data for GDP growth. Restricting this analysis to that smaller sample yields a slight increase in the magnitude of $\widehat{\delta}_O$ to 17 %, while it remains statistically significant. Second, because immigrants from non-English speaking countries may have lower earnings because of language difficulties, the final column shows estimates that come entirely from countries where English is the primary or official language.[44] Again, even for this sample of just 12 countries, variations in cognitive skills across countries have a strongly significant impact on earnings of 16 %.

The remaining rows of Table 5 provide estimates of the complete set of parameters. While there is some variance across samples in the estimate of the effect on earnings of being educated entirely in the country of origin, this appears to reduce average earnings by 6–13 % with the exception of English-speaking immigrants, who appear to suffer no significant average earnings loss compared to people educated entirely in the U.S. The estimated "Mincer" parameters (α_1, α_2, and α_3) appear within the range of typical estimates for the general population (see Heckman et al. 2008). Results remain qualitatively the same when indicators for decade of immigration and for gender are added to the model.[45] These last specifications address in part concerns that unmeasured differences between late immigrants (those receiving schooling in their home country) and early immigrants (those receiving schooling in the U.S.) might be driving the results.

These difference-in-differences estimates provide support for two conclusions about the potential causal impacts of cognitive skills. First, they contrast individuals receiving the treatment of home-country schooling to immigrants from the same country, all within the same labor market. Thus, they cannot be driven by differences in underlying economic institutions around the globe that are correlated with differences in cognitive skills. Second, they pinpoint the impact of schooling differences across countries, as distinct from family or cultural differences in attitudes, motivation, child rearing, and the like. In sum, the estimates, which are highly stable across different estimation samples, provide evidence that the economic impact is a causal one, and not purely associational.

It is very difficult, however, to compare magnitudes of coefficients from the immigrant earnings models to the growth regressions. These estimates are restricted to the private returns and leave out any of the externalities implicit in the estimated growth models. Thus, while

44 Data on English language come from the CIA World Factbook. Countries were coded as English speaking if the CIA World Factbook listed English as an official language or as the most widely spoken language in the country. See https://www.cia.gov/library/publications/the-world-factbook/.

45 When analyzed separately by gender, the results hold strongly for males whereas results for females—while pointing in the same direction—mostly do not reach statistical significance, as is common in labor-market analyses.

the estimated earnings impacts of cognitive skills are remarkably close to existing panel-data estimates, they do not translate easily into aggregate growth estimates.

The estimates do, however, provide direct support for the production view of schooling as contrasted with the signaling or screening view. Since the analysis of Spence (1973), questions have been raised about whether schools simply act as devices to select more able students as opposed to providing them with new knowledge and skills. Thus, an individual may get more schooling simply to signal high ability to the labor market. The difficulty has been that the labor-market implications for the returns to school attainment are the same in the alternative models—those with more schooling have more skills and thus receive higher wages. This fact has led to a variety of alternative ways to identify production versus signaling (cf. Weiss 1995; Riley 2001). One salient approach is reliance on what happens during schooling as opposed to the market returns to school attainment to identify the differences. The previous results provide just such evidence, because they show that the quality of different schools and the cognitive skills related to different schooling have direct payoffs within the same labor market. Therefore, these estimates yield even stronger evidence that policies to improve schools have social payoffs, as opposed to being limited to private payoffs as would be implied in the screening model.

7 Skill improvement and improved growth

The prior analyses have relied upon the average test score for each nation in order to characterize differences in skills of their labor forces. As noted, most of the variation in test scores occurs between countries, but the existence of some systematic change for countries suggests the possibility of an alternative difference-in-differences approach that uses the time-series evidence on performance within each country to identify the impact of skills on growth. Specifically, countries that improve the skills of their population—no matter how it is done—by the underlying model should see commensurate improvements in their rate of growth. This estimation removes any country-specific fixed effects affecting growth rates—such as basic economic institutions, cultural factors, political environment, and the like—and focuses on whether a country that alters the cognitive skills of its population is observed to receive an economic return.

While others have investigated turning points in growth, our focus is low-frequency changes such as those that might result from evolutionary schooling policies and that alter the path of economic growth.[46] Policies affecting the skill composition of the labor force necessarily unfold over lengthy periods and are not seen as sharp changes in outcomes.

To characterize the longitudinal patterns of test scores, we regress separate test scores by year, age group, and subject on a time variable (as well as age-group and subject indicators) and use the time coefficient as the measure of change in cognitive skills for each nation (see Appendix B for details). The amount of noise in each test observation, particularly with our common scaling, implies that such trends are also estimated with considerable noise. We therefore trust the rough cross-country pattern more than the specific point estimates of changes in each country. To put limits on the amount of noise affecting our analyses, we rely

[46] Relevant studies include Hausmann et al. (2005) who look at episodes of “growth accelerations”; Jones and Olken (2008) who consider patterns of 10-year periods of acceleration and collapse; and Barro and Ursúa (2008) who identify events of major declines in consumption that have potential implications for long-run growth. The identified periods are generally characterized by financial crisis, political instability, or war.

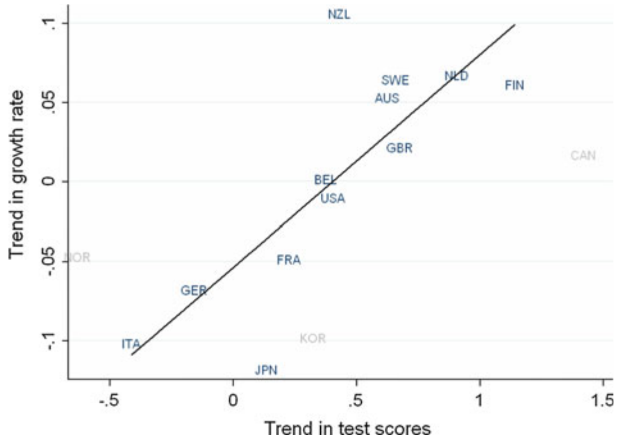

Fig. 2 Trends in Growth Rates vs. Trends in Test Scores. *Notes*: Scatter plot of trend in the growth rate of GDP per capita from 1975 to 2000 against trend in test scores, which is equivalent to the first column of Table 6. Three countries without test scores before 1972 in light gray; regression line refers to the remaining twelve countries. See Appendix B for details

on the sample of OECD countries that have test observations both before 1985 and up to 2003.[47]

As is evident from Fig. 2 (see also Appendix Fig. 5), substantial changes in test performance—both positive and negative—have occurred for OECD countries.[48] The rapid growth in performance of such countries as Canada, Finland, and the Netherlands contrasts sharply with the declining scores in Germany, Italy, and Norway. For our purposes, however, we are not interested in test scores for the school-aged population but instead in the skills of the relevant portions of the labor force. Thus, we need to assume that the currently observed trends in performance reflect long-run patterns of skill change and specifically those holding during the earlier time periods.

In a parallel manner, we estimate a time trend for annual growth rates in each country using the Penn World Tables data. The annual growth rate series for each country contains considerable noise, largely reflecting short-run business cycle phenomena or financial crises, and the trend estimation is designed to extract long-run changes in growth.[49]

The consistency of changes in test performance and changes in growth rates is evident in Fig. 2. When we split countries by being above or below the median change in growth rates and above or below the median change in cognitive skills, all countries fall into either the positive or negative quadrants on both measures. The largest outliers from the trend line are

[47] In fact, all countries except Canada, Korea, and Norway have test scores dating back at least to 1971.

[48] A comparison of the country rankings of projected skill levels for 1975 and 2000 yields a Spearman rank correlation is 0.78—again reinforcing the validity of country average scores for the main growth analysis.

[49] Descriptive statistics are found in Appendix Table 14. We also tried alternative measures of growth-rate changes, including the difference between the average growth rate in the first five years and in the last five years; trend growth using IMF data in national currencies; and IMF national currency data for the period 1975–2004. Using IMF national currency data is consistent with Nuxoll (1994) and Hanousek et al. (2008) who argue that using national accounts data is superior to relying on the price and exchange-rate adjustments in the basic Penn World Tables data when looking at growth rates. In our investigation of these options, the estimates of the impact of changes in test scores remain statistically significant and quantitatively very similar across alternatives and compared to the estimates reported in Table 6.

Table 6 Changes in cognitive skills and changes in growth paths

	(1)	(2)	(3)	(4)	(5)[a]	(6)[b]	(7)	(8)	(9)
Trend in cognitive skills	0.084	0.073	0.074	0.074	0.080	0.117			0.073
	(3.10)	*(3.21)*	*(3.07)*	*(3.04)*	*(3.34)*	*(6.90)*			*(2.97)*
Dummy for cognitive-skill trend above median							0.117	0.103	
							(5.98)	*(4.87)*	
Average annual growth rate in GDP per capita 1975–2000		−0.030	−0.035	−0.028	−0.039	−0.085		−0.004	−0.031
		(2.73)	*(1.61)*	*(1.69)*	*(2.32)*	*(5.26)*		*(0.21)*	*(2.55)*
Initial GDP per capita			−0.002					0.005	
			(0.27)					*(0.80)*	
Change in years of schooling 1975–2000				−0.004					
				(0.21)					
Trend in cognitive skills in 1999–2009									0.0004
									(0.03)
No. of countries	15	15	15	15	15	12	15	15	15
R^2 (adj.)	0.380	0.586	0.551	0.550	0.582	0.891	0.713	0.735	0.548

Notes Dependent variable: trend in the growth rate of GDP per capita from 1975 to 2000. Regressions include a constant. Sample: OECD countries with test-score data both before 1985 and up to 2003. Test scores are average of math and science, primary through end of secondary school. Absolute t-statistics in parentheses

[a] WLS with inverse of standard error with which the trend in test scores was estimated as weights

[b] Excluding countries without test scores before 1972 (Canada, Korea, and Norway)

precisely the countries that have less historical test score data (Canada, Korea, and Norway) and that thus have poorer trend data.

We provide estimates of simple models of the change in growth rates over the 1975–2000 period in Table 6. By focusing on changes in test scores and in growth rates, these specifications are essentially equivalent to panel estimates with country fixed effects that eliminate time-invariant factors of cultural, institutions, or other potential influences. For the 15 OECD countries, 38 % of the variance in growth-rate changes can be explained by test-score changes.[50] If we add measures for the average growth rate in each country and the initial GDP per capita (col. 2–3), the change in achievement scores remains statistically significant at near the same level as found in the simple regressions of column 1. The same is true when the change in quantitative educational attainment is added to the model (col. 4). Importantly, the change in educational attainment is orthogonal to the change in growth rates (either with controls for the test-score trend or without), reinforcing the introductory skepticism about the efficacy of past reliance on school attainment measures of human capital. Likewise, results are hardly affected if we weigh each observation by the inverse of the standard error with

[50] Results are fully consistent when adding six non-OECD countries with available data to the analysis (disregarding three non-OECD countries that prove extreme outliers on the estimated test-score trend at more than two and a half standard deviations above/below the OECD sample mean). While in the extended 21-country sample, OECD countries on average have lower growth trends, the effect of test-score trends on growth trends does not differ significantly between the OECD and non-OECD countries, and the main effect of the test-score trend is unaltered from the OECD-sample analysis (detailed results available from the authors upon request).

which the trend in test scores was estimated, in order to down weight those that are more noisily estimated (col. 5).

If, however, we restrict the analysis to those countries with test scores spanning a range of more than three decades, from at least 1971 to 2003, both the coefficient estimate and the explained variance grow in size (col. 6), as suggested in Fig. 2. In the sample without the three countries with limited time series information (Canada, Korea, and Norway), the test-score trend alone accounts for 64 % of the variation in growth trends. Alternative specifications look simply at whether the test-score trend is above or below the OECD median (col. 7–8). In all cases, the impact of changes in test scores on changes in growth rates remains very stable and is always statistically significant.

The underlying identifying assumption of these analyses is that the observed test-score trend captures a prior trend and is not affected by the partly overlapping growth trend. One way to test the validity of this identifying assumption is to use the most recently available test score data to estimate the trend in test scores subsequent to the period over which the growth trend is estimated. To do so, we estimate the trend in test scores from the 24 available test observations in 1999–2009 (details available on request). When entering the test-score trend in 1999–2009 to our regression, it is totally unrelated to the prior growth trend and does not affect the result on the prior test-score trend (col. 9). In fact, the test-score trend in 1999–2009 is not correlated with the prior long-term test-score trend (correlation coefficient −0.302, p-value 0.274) or with the growth trend in 1975–2000 (correlation coefficient −0.293, p-value 0.289).[51] This result corroborates the assumption that the identifying test-score variation is not itself caused by variation in growth. Furthermore, analyses presented in Appendix D (Table 16) use the most recent Penn World Table data (version 7.0) to show that results are strengthened when relating the initial test-score trend to the growth trend in 1985–2007 (rather than 1975–2000).

Still, this analysis requires backward extrapolation of the test-score data in order to capture changes for workers in the labor force. Thus, it cannot be considered definitive. We can, however, relate these estimates to the prior growth models. If we assumed that the observed trend in test scores had been going on since the oldest person in the current labor force went to school, an annual increase in test scores by 1 % of a standard deviation would translate into an annual increase in the growth rate by 0.07–0.12 percentage points. However, if we more realistically thought that any change in test scores began at the beginning of our observation period, then the impact of student improvements on the average labor force is much less, and the projected change in growth rates would be commensurately reduced. Back-of-the-envelope calculations suggest that in such a setting, the estimates based on the trend analysis in Table 6 are close to the steady-state estimates in Table 1.

In conclusion, the positive relationship between improving cognitive skills and improving growth rates provides another set of surprisingly consistent results based on a different approach to identifying the causal impact of cognitive skills—a focus on changes within each country that removes country-specific fixed effects. While it requires large extrapolations of changes to cover existing workers, the results are remarkably compatible with the underlying growth model—showing growth rates changing in a manner consistent with changes in cognitive skills.

[51] Results are qualitatively the same when using just the PISA tests in 2000–2009.

8 Rocket scientists or basic education for all?

While addressing the range of potential schooling policy options is clearly beyond the scope of this paper, our new data series allows us to extend the growth analysis to illuminate one important issue—whether to concentrate attention at the lowest or at the highest achievers. Some argue in favor of elitist school systems which focus on the top performers as potential future managers of the economy and drivers of innovation. Others favor more egalitarian school systems to ensure well-educated masses that will be capable of implementing established technologies. In other words, should education policy focus on forming a small group of "rocket scientists," or are approaches such as the Education for All initiative (UNESCO 2005) more promising in spurring growth?

To capture these differences in the distributional patterns of the test-score performance in different countries, we use the microdata from each of the international assessments to calculate measures of the share of students in each country who reach at least basic skills as well as those who reach superior performance levels (see Appendix B). We use performance of at least 400 test-score points on our transformed international scale—one standard deviation below the OECD mean—as our threshold of basic literacy and numeracy.[52] The international median of this share of students is 86 % in our sample, ranging from 18 % in Peru to 97 % in the Netherlands and Japan. As our threshold for superior performance, we take 600 points or one standard deviation above the OECD mean.[53] This level is reached by an international median of only 5 %, although it ranges from below 0.1 % in Colombia and Morocco to 18 % in Singapore and Korea and 22 % in Taiwan.[54] (As shown in Appendix Fig. 4, these differences represent more than simple mean displacement.)

As seen in the first three columns of Table 7, both measures of the test-score distribution are significantly related to economic growth, either when entered individually or jointly.[55] Both the basic-skill and the top-performing dimensions of educational performance appear separately important for growth. From the estimates in column 3, a ten percentage point increase in the share of students reaching basic literacy is associated with 0.3 percentage points higher annual growth, and a ten percentage point increase in the share of top-performing students is associated with 1.3 percentage points higher annual growth. However, it may be much more feasible to increase the basic-literacy share than to increase the top-performing share by the same amount, as suggested by the fact that the international standard deviations of

[52] The PISA 2003 science test uses the threshold of 400 points as the lowest bound for a basic level of science literacy (Organisation for Economic Co-operation and Development 2004, p. 292), and on the math test this corresponds to the middle of the level 1 range (358 to 420 test-score points), which denotes that 15-year-old students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. For example, given the exchange rate between dollars and euros, a student at level 1 can calculate the euro-equivalent of a given number of dollars.

[53] A score of 600 points is near the threshold of the level 5 range of performance on the PISA 2003 math test, which denotes that 15-year-old students can develop and work with models for complex situations, identifying constraints and specifying assumptions; they can reflect on their answers and can formulate and communicate their interpretations and reasoning.

[54] The distributions depicted in Fig. 4 reveal that such distributional measures capture much more of the overall distribution than a simple measure such as the standard deviation in national test scores. The standard deviation in test scores does not enter our basic model significantly (see Castelló and Doménech 2002 for related analyses using measures of educational inequality based on years of schooling).

[55] In the joint model, the two measures are separately significant even though they are highly correlated across countries with a simple correlation of 0.73. The mean test score used in previous models is more highly correlated with the basic literacy share ($r = 0.96$) than with the top-performing share ($r = 0.85$). If the mean test score is added to column 3, the basic-literacy share becomes insignificant, but in a specification with just the mean, mean and top-performing shares both remain significant.

Table 7 Rocket scientists or basic education for all?

	(1)	(2)	(3)	(4)[a]	(5)[b]	(6)[b]	(7)[b]
Share of students reaching basic literacy	4.717		2.732	1.002	3.460	5.150	5.869
	(6.64)		*(3.61)*	*(1.33)*	*(3.81)*	*(2.87)*	*(3.33)*
Share of top-performing students		19.347	12.880	11.735	8.460	4.226	−1.530
		(2.653)	*(4.35)*	*(4.18)*	*(2.37)*	*(0.65)*	*(0.22)*
Share of students reaching basic literacy × initial GDP per capita					0.376		
					(1.25)		
Share of top-performing students × Initial GDP per capita					−2.148		−1.649
					(2.11)		*(2.07)*
Share of students reaching basic literacy × Share of top-performing students						42.357	53.538
						(1.48)	*(1.91)*
No. of countries	50	50	50	45	50	50	50
R^2 (adj.)	0.610	0.646	0.719	0.823	0.734	0.727	0.746

Notes Dependent variable: average annual growth rate in GDP per capita, 1960–2000. Control variables: GDP per capita 1960, years of schooling 1960, and a constant. Shares are based on average test scores in math and science, primary through end of secondary school, all years. Absolute t-statistics in parentheses

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

[b] All interacted variables are centered on zero

these two shares are 0.215 and 0.054, respectively. Thus, increasing each share by roughly half a standard deviation (10 percentage points basic-literacy share and 2.5 percentage points top-performing share) yields a similar growth effect of roughly 0.3 percentage points.

The impact of having more top performers is only slightly reduced by introducing the measures of economic institutions, fertility, and tropical geography (col. 4). On the other hand, the separate influence of basic literacy levels falls quantitatively and becomes statistically insignificant in the expanded model (for the 45 countries with complete data), in line with an interpretation where part of the effect of basic literacy comes through improved institutions (Glaeser et al. 2004).

The effect of the basic-literacy share does not vary significantly with the initial level of development, but the effect of the top-performing share is significantly larger in countries that have more scope to catch up to the initially most productive countries (col. 5). These results appear consistent with a mixture of the basic models of human capital and growth mentioned earlier. The accumulation of skills as a standard production factor, emphasized by augmented neoclassical growth models (e.g., Mankiw et al. 1992), is probably best captured by the basic-literacy term, which has positive effects that are similar in size across all countries. But, the larger growth effect of high-level skills in countries farther from the technological frontier is most consistent with technological diffusion models (e.g., Nelson and Phelps 1966). From this perspective, countries need high-skilled human capital for an imitation strategy, and the process of economic convergence is accelerated in countries with larger shares of high-performing students.[56] Obvious cases are East Asian countries such as Taiwan, Singapore, and Korea that all have particularly large shares of high-performers, started from relatively low levels, and have shown outstanding growth performances, but the results of column 5 are nonetheless robust to the inclusion of an East Asian dummy, or a full set of regional dummies.

[56] For an alternative model of imitation and innovation that emphasizes the innovation margin, see Vandenbussche et al. (2006) and Aghion et al. (2009). These studies, however, focus just on developed countries and miss the role of rocket scientists in the transmission of technologies to developing countries. Moreover, Hanushek and Woessmann (2011b) show that differences in basic skills are more important than differences in advanced skills in explaining growth differences among just the OECD countries.

A particularly informative extension considers the interaction of the top-performing and basic-literacy shares (col. 6 and 7). This complementarity between basic skills and top-level skills suggests that in order to be able to implement the imitation and innovation strategies developed by scientists, countries need a workforce with at least basic skills.[57]

Many countries have focused on either basic skills or engineers and scientists. In terms of growth, our estimates suggest that developing basic skills and highly talented people reinforce each other. Moreover, achieving basic literacy for all may well be a precondition for identifying those who can reach "rocket scientist" status. In other words, tournaments among a large pool of students with basic skills may be an efficient way to obtain a large share of high-performers.

9 Conclusions

A myriad of empirical estimates of cross-country growth models exist. The general criticism of these is that they provide little confidence that the models satisfactorily identify the causal impact of their included determinants of growth. And, a related criticism is that they then cannot provide any real policy guidance.

We have focused on the role of educational achievement, or cognitive skills, in determining economic growth and have taken the quest for policy guidance seriously. We have investigated a set of models that approach identification from different vantage points. Individually, as discussed, these approaches do require some strong maintained hypotheses, but importantly each of the approaches is subject to different questions and would fail for very different reasons. While there remain other threats to identification that cannot be ruled out in our samples, the alternative analytical perspectives narrow the range of possible opposing explanations for the stylized facts based on reverse causation, economic and social institutions, and cultural influences.

The clearest result here is the consistency of the alternative estimates of the cognitive skills-growth relationship—both in terms of quantitative impacts and statistical significance. The remarkable stability of the models in the face of alternative specifications, varying samples, and alternative measures of cognitive skills implies a robustness uncommon to most cross-country growth modeling. In terms of previous questions about the fragility of any estimates of human capital and growth, these estimates underscore a simple finding that prior results suffered from critical measurement issues.

The stylized fact of a very strong relationship between cognitive skills and growth does not address all concerns. For policy advice, it is important to know whether the estimated relationship is causal or a mere association reflecting omitted variables, poor achievement measurement, or restricted models of growth. With the limited sample of nations—each with alternative political, cultural, and economic institutions—it is clearly very difficult to rule out all hypotheses about other possible reasons for the association between cognitive skills and growth. Our approach involves adapting common microeconometric approaches to the very different circumstances of long-term economic growth. We estimate instrumental-variable models using institutional characteristics of each country's school system to rule out that the skill-growth connection just reflects cultural differences across countries. Difference-in-differences approaches for immigrants from different countries on the U.S. labor market to rule out the possibility that test scores simply reflect cultural factors or economic institutions of

[57] The issue of skill complementarity in production has been addressed in explaining the pattern of earnings inequality. The U.S. analysis of Autor et al. (2006, 2008) suggests that high-skilled workers and low-skilled workers are complements, a result that helps explain income variations across the educational spectrum.

the home countries. And longitudinal models to eliminate the level effects which may be interrelated with country-specific institutions and cultures. Conclusions about the impact of cognitive skills are remarkably consistent across the very different estimation methods.

Our analysis is reinforced by Ciccone and Papaioannou (2009) who find that countries with a more skilled labor force (using the Hanushek and Kimko test measures) experienced faster growth in skill-intensive industries during the 1980s and 1990s. This evidence, which is derived from within-country analysis of development outcomes, strengthens our interpretation that more skilled people contribute to a more rapid adoption of new technologies and production processes—a central element of both endogenous growth models that stress innovation and ideas (Romer 1990) and of models of technological diffusion and growth (Nelson and Phelps 1966). By using country and industry fixed effects, it also excludes a variety of concerns about endogeneity due to variations in country institutions and cultures which tend to affect sectors uniformly.

The simple conclusion from the combined evidence is that differences in cognitive skills lead to economically significant differences in economic growth. Moreover, since the tests concentrate on the impact of schools, the evidence suggests that school policy can, if effective in raising cognitive skills, be an important force in economic development. Thus, it would be inappropriate to interpret the test differences as a simple reflection of ability or family differences, factors that might be very impervious to policy changes.

Finally, what are the economic rewards that are suggested by the models for improving skills? Almost all of the alternative specifications and modeling approaches suggest that one standard deviation higher cognitive skills of a country's workforce is associated with approximately two percentage points higher annual growth in per capita GDP. This magnitude is clearly substantial, particularly when compared to regional growth rates that average between 1.4 and 4.5 % over the 1960-2000 period (Appendix Table 8). On the other hand, it is implausible to expect a country to improve by one standard deviation—bringing, say, Mexico up to the OECD average—over any reasonable time horizon. It is plausible to think of getting schooling improvements that would lift a country's average by 1/4 standard deviation (25 points on a PISA scale). This kind of improvement has, for example, been observed by Poland during the past decade and by Finland over the past two to three decades. In an endogenous growth world, where long-run growth rates can change permanently with improved skills and human capital, the estimates would imply a boast in growth rates of 1/2 %. Such a change would lead to enormous changes in the economic outcomes for a country. Hanushek and Woessmann (2011b) simulate the impact of such a change on GDP and find that the present value of improvements over the next 80 years would be almost three times the value of current GDP for the country.[58] Put in terms of the present value of future GDP, this amounts to a 6.2 % increase in GDP over the 80-year period.

An alternative perspective is that of the neoclassical model where there would be a period of higher growth because of the improved skills but the economy would then return to the prior steady-state rate of growth at a higher level of income. The prior growth model estimates provide a means of estimating the impacts of this.[59] In the same scenario as above, the present value would be slightly over two times current GDP or 4.3 % of the present value of GDP over the 80 year period. The relative closeness of the endogenous and neoclassical estimates reflects the fact that convergence is, similar to that in Mankiw et al. (1992), relatively slow

[58] These simulations assume that the past growth patterns hold into the future; that base growth in per capita GDP is 1.5 %; that the reform of schools takes 20 years to complete; that the impact depends on the average quality of the labor force; and that future values are discounted at 3 %.

[59] For these simulations we rely upon the estimates in Table 1, column 9, that have the log of initial GDP per capita as an explanatory variable.

so that it takes a substantial time before the economy reverts to the old steady-state path. These simulations also underscore that the estimated impacts are still large even if part of our estimates were produced by omitted variables that are correlated with cognitive skills. If, for example, the true causal impact of skill differences were only half as large as suggested by our estimates, we would still be left with extraordinarily large policy opportunities.

By itself, finding a potential role for schools does not point to any clear policies. Indeed, that discussion would enter into a variety of controversial areas and would lead us far afield. Nonetheless, our aggregate data provide direct evidence that both providing broad basic education—education for all—*and* pushing significant numbers to very high achievement levels have economic payoffs.

Appendix A: Regional data

See Table 8.

Appendix B: Measures of cognitive skills

A key element of our work is developing a measure that can equate knowledge of people across countries. In many ways this is an extension of notions of human capital that have been developed over the past half century. But it is a specific refinement that, while important in a variety of applications within nations, becomes a necessity when comparing different countries. Within a country, human capital is often proxied by quantity of schooling. This is partly necessitated by commonly available data but partly justified on the idea that differences in knowledge between levels of schooling are greater than those within levels of schooling.

Until recent publicity, most people were unaware of the international student testing that could provide direct comparisons of student knowledge across countries. In fact, international assessments of student achievement, aimed largely at math and science, were begun over four decades ago. Although national participation has been voluntary, recent expansions to all OECD countries and more have led to increasingly valid and reliable indicators of cognitive skills.

Internationally comparable student achievement tests have been conducted since the First International Mathematics Study (FIMS), which tested in 1964. The latest international studies used in our analyses are the 2003 cycles of the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). From FIMS to the latest TIMSS and PISA, a total of 12 international student achievement tests (ISATs) were conducted.[60] Although varying across the individual assessments, testing covers math, science, and reading for three age/grade groups: primary education (age 9/10), lower secondary education (age 13 to 15), and the final year of secondary education (generally grade 12 or 13).

Given this 3×3 grade-by-subject matrix, Table 9 summarizes the specific ISATs that have been conducted in three periods of time: late 1960s/early 1970s (1964-72), 1980s (1982-91), and late 1990s/early 2000s (1995-2003). Several features of the emerging pattern are worth

60 In this study, we do not include the two tests conducted by the International Assessment of Educational Progress (IAEP) in 1988 and 1991, because they used the U.S. NAEP test as their testing instrument, which is geared to the U.S. curriculum and may thus introduce bias to the international testing. By contrast, the tests included here are not associated with the curriculum in any particular country, but have been devised in an international cooperative process between all participating countries.

Table 8 Income, education, and growth across world regions

Region[a]	No.countries[b]	GDP per capita 1960 (US$)	Growth of GDP per capita 1960–2000 (%)	GDP per capita 2000 (US$)	Years of schooling 1960	Test score	All Penn World Tables Countries	
							No. countries[c]	GDP per capita 1960 (US$)
Asia	11	1,891	4.5	13,571	4.0	479.8	15	1,642
Sub-Saharan Africa	3	2,304	1.4	3,792	3.3	360.0	40	1,482
Middle East and North Africa	8	2,599	2.7	8,415	2.7	412.4	10	2,487
Southern Europe	5	4,030	3.4	14,943	5.6	466.4	5	4,030
Latin America	7	4,152	1.8	8,063	4.7	388.3	24	3,276
Central Europe	7	8,859	2.6	24,163	8.3	505.3	7	8,859
Northern Europe	5	8,962	2.6	25,185	8.0	497.3	5	8,962
Commonwealth OECD	4	11,251	2.1	26,147	9.5	500.3	4	11,251
Note: Asia w/oJapan	*10*	*1,614*	*4.5*	*12,460*	*3.5*	*474.7*	*14*	*1,427*

[a] The country observations contained in the eight regions are: Asia (11): China, Hong Kong, India, Indonesia, Japan, Rep. of Korea, Malaysia, Philippines, Singapore, Taiwan, Thailand; Sub-Saharan Africa (3): Ghana, South Africa, Zimbabwe; Middle East and North Africa (8): Cyprus, Egypt, Iran, Israel, Jordan, Morocco, Tunisia, Turkey; Southern Europe (5): Greece, Italy, Portugal, Romania, Spain; Latin America (7): Argentina, Brazil, Chile, Colombia, Mexico, Peru, Uruguay; Central Europe (7): Austria, Belgium, France, Ireland, Netherlands, Switzerland, United Kingdom; Northern Europe (5): Denmark, Finland, Iceland, Norway, Sweden; Commonwealth OECD members (4): Australia, Canada, New Zealand, USA.

[b] Sample of all countries by region with internationally comparable data on GDP that ever participated in an international student achievement test; see Appendix B for details.

[c] Sample of all countries in Penn World Tables with data on GDP per capita in 1960 by region.

Sources GDP: own calculations based on Penn World Tables (Heston et al. (2002)); years of schooling: own calculations based on Cohen and Soto (2007); test score: own calculations based on international student achievement tests; see Appendix B for details.

Table 9 International tests by period, subject, and age group

	Math	Science	Reading
1964–72			
Primary		FISS	
Lower secondary	FIMS	FISS	FIRS
Final secondary	FIMS	FISS	
1982–91			
Primary		SISS	SIRS
Lower secondary	SIMS	SISS	SIRS
Final secondary	SIMS	SISS	
1995–2003			
Primary	TIMSS	TIMSS	PIRLS
	TIMSS 2003	TIMSS 2003	
Lower secondary	TIMSS	TIMSS	
	TIMSS-Repeat	TIMSS-Repeat	
	PISA 2000/02	PISA 2000/02	
	TIMSS 2003	TIMSS 2003	PISA 2000/02
	PISA 2003	PISA 2003	PISA 2003
Final secondary	TIMSS	TIMSS	

Notes For abbreviations and details, see Table 10

noting. First, math and science have been tested at all three grade levels, while reading has never been tested in the final grade of secondary school. Second, all subjects are available in all periods, although coverage is more extensive in math and science than in reading. Third, in each period, the lower secondary level has been tested in all three subjects; thus, there is no primary or final-secondary study that would add a subject not already tested at lower secondary in the period. Fourth, each cell available in 1964-91 has at least one counterpart in 1995-2003.

Table 10 provides additional detail on each ISAT. A total of 77 countries have participated in at least one of the ISATs in math or science, but several of the countries have participated at only one or a few points in time. Even within the same assessment, countries do not always participate at all grade levels. The largest number of countries tends to have participated at the lower secondary level.

To obtain a common measure of cognitive skills, we want to draw on as much internationally comparable information as possible. This raises the issue whether the different ISATs with their different participating countries, student samples, and perspectives on what should be tested (see Neidorf et al. 2006) are measuring a common dimension of cognitive skills. For example, the TIMSS tests are related to elements of the school curricula common to participating countries, while the PISA tests are designed to be applied assessments of real-world problems, irrespective of specific curricula. However, the fact is that the TIMSS tests with their curricular focus and the PISA tests with their applied focus are highly correlated at the country level. For example, the correlation between the TIMSS 2003 tests of 8th graders and the PISA 2003 tests of 15-year-olds across the 19 countries participating in both is as high as 0.87 in math and 0.97 in science. It is also 0.86 in both math and science across the 21 countries participating both in the TIMSS 1999 tests and the PISA

Table 10 The international student achievement tests

	Abbreviation	Study	Year	Subject	Age[a,b]	Countries[c]	Organization[d]	Scale[e]
1	FIMS	First International Mathematics Study	1964	Math	13,FS	11	IEA	PC
2	FISS	First International Science Study	1970–71	Science	10,14,FS	14,16,16	IEA	PC
3	FIRS	First International Reading Study	1970–72	Reading	13	12	IEA	PC
4	SIMS	Second International Mathematics Study	1980–82	Math	13,FS	17,12	IEA	PC
5	SISS	Second International Science Study	1983–84	Science	10,13,FS	15,17,13	IEA	PC
6	SIRS	Second International Reading Study	1990–91	Reading	9,13	26,30	IEA	IRT
7	TIMSS	Third International Mathematics and Science Study	1994–95	Math/Science	9(3+4), 13(7+8),FS	25,39,21	IEA	IRT
8	TIMSS-Repeat	TIMSS-Repeat	1999	Math/Science	13(8)	38	IEA	IRT
9	PISA 2000/02	Programme for International Student Assessment	2000+02	Reading/ Math/Science	15	31+10	OECD	IRT
10	PIRLS	Progress in International Reading Literacy Study	2001	Reading	9(4)	34	IEA	IRT
11	TIMSS 2003	Trends in International Mathematics and Science Study	2003	Math/Science	9(4),13(8)	24,45	IEA	IRT
12	PISA 2003	Programme for International Student Assessment	2003	Reading/ Math/Science	15	40	OECD	IRT

[a] Grade in parentheses where grade level was target population.
[b] FS = final year of secondary education (differs across countries).
[c] Number of participating countries that yielded internationally comparable performance data.
[d] Conducting organization: International Association for the Evaluation of Educational Achievement (IEA); Organisation for Economic Co-operation and Development (OECD).
[e] Test scale: percent-correct formal (PC); item-response-theory proficiency scale (IRT).

2000/02 tests. Thus, ISATs with very different foci and perspectives tend, nonetheless, to be highly related, lending support to our approach of aggregating different ISATs for each country.

The general idea behind our approach to aggregation is that of empirical calibration. We rely upon information about the overall distribution of scores on each ISAT to compare national responses. This contrasts with the psychometric approach to scaling that calibrates tests through common elements on each test. In reality, the international testing situations are separate events with no general attempt to provide common scaling across tests and across the full time period.

The fact that the scales of their test-score results are not directly equated across tests is a major drawback in comparative uses of the various ISATs. They do not use the same test questions; nor do they even use the same technique and scale of mapping answers into test scores.[61] The early tests mainly used aggregate scores in "percent correct" format, but with questions of varying difficulty in the different tests, these scores will not be comparable across tests. The later tests use a more sophisticated scale, constructed using Item Response Theory (IRT). Among other things, IRT weights different questions by their revealed difficulty and then maps answers onto a pre-set scale set to yield a given international mean and standard deviation among the participating countries. However, the questions on which the mapping is based are not the same in the different tests. Even more, the set of participating countries varies considerably across tests, making the separately developed scales incomparable across ISATs.

Therefore, to compare performance on the ISATs across tests and thus over time, we have to project the performance of different countries on different tests onto a common metric. For that, we have to develop a common metric both for the *level* and for the *variation* of test performance.

Comparable level. To make the level of ISATs comparable, we need information on test performance that is comparable over time. Such information is available in the United States in the form of the National Assessment of Educational Progress (NAEP), which tests the math, science, and reading performance of nationally representative samples of 9-, 13-, and 17-year-old U.S. students in an intertemporally comparable way since 1969. This is the only available information on educational performance that is consistently available for comparisons over time. The United States is also the only country that participated in every ISAT. Given the time-series evidence on the performance of U.S. students, we can thus scale the level of each ISAT relative to the known intertemporally comparable test performance of the United States. Figure 3 shows the available NAEP results in the three subjects and age groups.[62] Despite some notable changes, the performance of U.S. students has been relatively flat over the period 1969-1999.

We start by calculating the U.S. performance difference between 1999 and any earlier point in time and express it in standard deviations (s.d.) of the international PISA 2000 study:

$$U_{a,s,t}^{US} = \left(NAEP_{a,s,t}^{US} - NAEP_{a,s,1999}^{US}\right) \frac{SD_s^{US,PISA}}{SD_{a,s}^{US,NAEP}} \tag{A1}$$

[61] Recent testing in both TIMSS and PISA has involved overlapping test items that permit test calibration, but these do not provide any benchmarks across the two testing regimes or links with earlier testing.

[62] Note that changes in NAEP testing make it difficult to use this methodology for the more recent PISA and TIMSS assessments. For example, the science tests were revised in 2009, and the new scale that was employed makes the data incomparable to prior years. On the other hand, recent PISA and TIMSS assessments have been designed to provide comparability over time of the subject surveys.

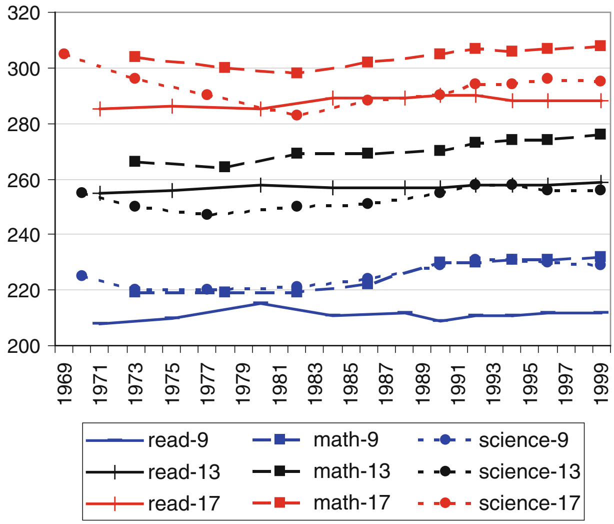

Fig. 3 Student Achievement in the United States over Time: The National Assessment of Educational Progress (NAEP) Source: U.S. Department of Education, Institute of Education Sciences (2008)

where U is the standardized performance difference of U.S. students at age a in subject s at time t relative to 1999, $NAEP$ is the age-, subject-, and time-specific NAEP test score, $SD^{US,PISA}$ is the subject-specific s.d. of U.S. students on the PISA test, and $SD^{US,NAEP}$ is the age- and subject-specific s.d. of the U.S. NAEP test.[63] NAEP scores are available at 2–4 year intervals over the period; values for non-NAEP years are obtained by linear interpolation between available years.

This alone does not yet yield a common scale for all the countries on the different tests. While we know for each participating country whether it performed above or below the respective U.S. performance on each specific test, we need to make the international variation in test scores comparable across the different ISATs to determine "how much" above or below.

Comparable variation. Developing a common metric for the variation of test scores in the different ISATs is harder to achieve than for the level. There is no explicit external information available on trends in the cross-country performance variation, and the diversity of original tests and participating countries precludes a direct comparison across tests. One way to achieve comparability, though, would be to have a group of countries across which it is reasonable to assume relative constancy in the size of the cross-country variation in test scores and whose members participated in sufficient number in the different tests. This group could only include relatively stable countries with relatively stable education systems over time, which should not have experienced major changes in overall enrollment across the ISATs.

Thus, we suggest two criteria for a group of countries to serve as a standardization benchmark for performance variation over time. First, the countries have been member states of the relatively homogenous and economically advanced group of OECD countries in the whole

63 The s.d. of the NAEP tests in reading for 1984-1996 and in math and science in 1977/78-1996 are reported in Department of Education, Institute of Education Sciences (2008). Since no s.d. information is available for the earlier and the 1999 NAEP tests, and since the available s.d. are relatively stable over time, we take a simple mean of the available s.d. in each subject at each age level over time. PISA tested only 15-year-olds, but has the same three subjects as the NAEP test.

period of ISAT observations, that is, since 1964. Second, the countries should have had a substantial enrollment in secondary education already in 1964. Given data constraints, we implement this by dropping all countries where more than half of the 2001 population aged 45-54 (the cohort roughly in secondary school in the first ISAT) did not attain upper secondary education (OECD 2003a). There are 13 countries that meet both of these measures of stability, which we term the "OECD Standardization Group" (OSG) of countries.[64]

Under the assumption that the cross-country variation among the OSG countries did not vary substantially since 1964, we can use the OSG countries to develop a comparable scale for the variation on the different ISATs. We do so by projecting the s.d. among those of the OSG countries that participated in any particular ISAT from the subject-specific PISA test onto the particular ISAT. That is, we transform the original test score O of country i (specific for each age a and subject s) at time t into a transformed test score X according to:

$$X^i_{a,s,t} = \left(O^i_{a,s,t} - \overline{O^{OSG}_{a,s,t}}\right) \frac{SD^{OSG}_{s,PISA}}{SD^{OSG}_{a,s,t}} \tag{A2}$$

The test score X has the following distributional characteristics for each ISAT. First, it has a mean of zero among the OSG (attained by subtracting the OSG mean $\overline{O^{OSG}}$ from each country's original test score). Second, it has a between-country s.d. among the OSG that is the same as the s.d. of the very same countries on the PISA test in the specific subject (attained by dividing through the s.d. among the OSG countries in the specific test and multiplying by the s.d. of these same countries in the relevant PISA test). In effect, this rescaled test score now has a metric whose variation is comparable across tests.

Performance on a common metric. Finally, we use the time-series evidence on educational performance in the U.S. derived above to put a common level to the intertemporally comparable metric for the different ISATs. This is achieved in the standardized test score I:

$$I^i_{a,s,t} = X^i_{a,s,t} - X^{US}_{a,s,t} + O^{US}_{s,PISA} + U^{US}_{a,s,t} \tag{A3}$$

which adjusts the variation-adjusted test score X so that the U.S. performance level on each test equals the U.S. performance on the PISA test in the specific subject plus the age- and subject-specific adjustment factor U based on NAEP as derived in Eq. (A1) above.

Equation (A3) yields measures of the performance of the participating countries in each ISAT on a common scale that is comparable across ISATs. In effect, the internationally and intertemporally standardized test score I projects the PISA scale onto all other tests.

While we are reasonably confident about the comparisons of the standardized scores within the OECD countries which are fully tested in recent years, we are less certain about countries that are far from the measured OECD performance. In particular, countries far off the scale of the original test scores—e.g., two s.d. below the mean—may not be well represented because the tests may be too hard and thus not very informative for them. Our linear transformations are susceptible to considerable noise for these countries.

Our 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. Table 11 reports the basic

64 The OSG countries are: Austria, Belgium, Canada, Denmark, France, Germany, Iceland, Japan, Norway, Sweden, Switzerland, the United Kingdom, and the United States. The Netherlands also meets both criteria, but does not have internationally comparable PISA 2000 data which we require for our standardization.

Table 11 International data on cognitive skills

Code	Country	Growth sample[a]	Cognitive[b]	Lowsec[c]	Basic[d]	Top[e]
ALB	Albania	0	3.785	3.785	0.424	0.013
ARG	Argentina	1	3.920	3.920	0.492	0.027
ARM	Armenia	0	4.429	4.490	0.745	0.008
AUS	Australia	1	5.094	5.138	0.938	0.112
AUT	Austria	1	5.089	5.090	0.931	0.097
BEL	Belgium	1	5.041	5.072	0.931	0.094
BGR	Bulgaria	0	4.789	4.789	0.765	0.083
BHR	Bahrain	0	4.114	4.114	0.608	0.003
BRA	Brazil	1	3.638	3.638	0.338	0.011
BWA	Botswana	0	3.575	3.575	0.374	0.000
CAN	Canada	1	5.038	5.125	0.948	0.083
CHE	Switzerland	1	5.142	5.102	0.919	0.134
CHL	Chile	1	4.049	3.945	0.625	0.013
CHN	China	1	4.939	4.939	0.935	0.083
COL	Colombia	1	4.152	4.152	0.644	0.000
CYP	Cyprus	1	4.542	4.413	0.825	0.011
CZE	Czech Republic	0	5.108	5.177	0.931	0.122
DNK	Denmark	1	4.962	4.869	0.888	0.088
EGY	Egypt	1	4.030	4.030	0.577	0.010
ESP	Spain	1	4.829	4.829	0.859	0.079
EST	Estonia	0	5.192	5.192	0.973	0.095
FIN	Finland	1	5.126	5.173	0.958	0.124
FRA	France	1	5.040	4.972	0.926	0.085
GBR	United Kingdom	1	4.950	4.995	0.929	0.088
GER	Germany	0	4.956	4.959	0.906	0.105
GHA	Ghana	1	3.603	3.252	0.403	0.010
GRC	Greece	1	4.608	4.618	0.798	0.042
HKG	Hong Kong	1	5.195	5.265	0.944	0.123
HUN	Hungary	0	5.045	5.134	0.941	0.103
IDN	Indonesia	1	3.880	3.880	0.467	0.008
IND	India	1	4.281	4.165	0.922	0.013
IRL	Ireland	1	4.995	5.040	0.914	0.094
IRN	Iran	1	4.219	4.262	0.727	0.006
ISL	Iceland	1	4.936	4.945	0.908	0.074
ISR	Israel	1	4.686	4.660	0.826	0.053
ITA	Italy	1	4.758	4.693	0.875	0.054
JOR	Jordan	1	4.264	4.264	0.662	0.044
JPN	Japan	1	5.310	5.398	0.967	0.168
KOR	Korea, Republic of	1	5.338	5.401	0.962	0.178
KWT	Kuwait	0	4.046	4.223	0.575	0.000
LBN	Lebanon	0	3.950	3.950	0.595	0.002

Table 11 continued

Code	Country	Growth sample[a]	Cognitive[b]	Lowsec[c]	basic[d]	Top[e]
LIE	Liechtenstein	0	5.128	5.128	0.860	0.198
LTU	Lithuania	0	4.779	4.694	0.891	0.030
LUX	Luxembourg	0	4.641	4.641	0.776	0.067
LVA	Latvia	0	4.803	4.779	0.869	0.050
MAC	Macao-China	0	5.260	5.260	0.919	0.204
MAR	Morocco	1	3.327	3.243	0.344	0.001
MDA	Moldova	0	4.530	4.419	0.787	0.029
MEX	Mexico	1	3.998	3.998	0.489	0.009
MKD	Macedonia	0	4.151	4.151	0.609	0.028
MYS	Malaysia	1	4.838	4.838	0.864	0.065
NGA	Nigeria	0	4.154	4.163	0.671	0.001
NLD	Netherlands	1	5.115	5.149	0.965	0.092
NOR	Norway	1	4.830	4.855	0.894	0.056
NZL	New Zealand	1	4.978	5.009	0.910	0.106
PER	Peru	1	3.125	3.125	0.182	0.002
PHL	Philippines	1	3.647	3.502	0.485	0.006
POL	Poland	0	4.846	4.861	0.838	0.099
PRT	Portugal	1	4.564	4.592	0.803	0.032
PSE	Palestine	0	4.062	4.062	0.571	0.008
ROM	Romania	1	4.562	4.562	0.780	0.046
RUS	Russian Federation	0	4.922	4.906	0.884	0.081
SAU	Saudi Arabia	0	3.663	3.663	0.331	0.000
SGP	Singapore	1	5.330	5.512	0.945	0.177
SRB	Serbia	0	4.447	4.447	0.718	0.024
SVK	Slovak Rep.	0	5.052	5.052	0.906	0.112
SVN	Slovenia	0	4.993	5.076	0.939	0.061
SWE	Sweden	1	5.013	4.948	0.939	0.088
SWZ	Swaziland	0	4.398	4.398	0.801	0.004
THA	Thailand	1	4.565	4.556	0.851	0.019
TUN	Tunisia	1	3.795	3.889	0.458	0.003
TUR	Turkey	1	4.128	4.128	0.582	0.039
TWN	Taiwan (Chinese Taipei)	1	5.452	5.599	0.958	0.219
URY	Uruguay	1	4.300	4.300	0.615	0.049
USA	United States	1	4.903	4.911	0.918	0.073
ZAF	South Africa	1	3.089	2.683	0.353	0.005
ZWE	Zimbabwe	1	4.107	4.107	0.684	0.010

Notes A data file is available at www.cesifo.de/woessmann#data or http://hanushek.stanford.edu/download
[a]Indicator of whether country is in the main sample of 50 countries contained in the growth regressions, for which internationally comparable GDP data are available
[b]Average test score in math and science, primary through end of secondary school, all years (scaled to PISA scale divided by 100)
[c]Average test score in math and science, only lower secondary, all years (scaled to PISA scale divided by 100)
[d]Share of students reaching basic literacy (based on average test scores in math and science, primary through end of secondary school, all years)
[e]Share of top-performing students (based on average test scores in math and science, primary through end of secondary school, all years)

combined measure for the 77 countries that have ever participated in any of the math and science tests.[65] The sample for our growth regressions contains 50 of these countries.[66]

Distributional measures. Apart from the mean scores, we also analyze the distribution of test scores in each country by accessing the microdata of all ISATs.[67] The kernel density plots for math achievement on the 2003 PISA in Fig. 4 show that countries vary significantly in their patterns of test-score distributions. The depicted selected examples of developed countries reveals that it is possible to achieve relatively high median performance both with a relatively equal spread (Finland) and with a relatively unequal spread (Belgium) in the test scores at the student level. The same is true for countries with low average performance such as the depicted developing countries, where Brazil has a long right tail in contrast to Indonesia which shows a much greater density around its median.

To depict both ends of the distribution, we aim to calculate both the share of students reaching a basic level of literacy in the different subjects equivalent to 400 test-score points on the PISA scale (one student-level s.d. below the OECD mean) and the share of students reaching a top performance level equivalent to 600 test-score points on the PISA scale.

To do so, we use the above transformation to translate these two thresholds into the specific metric of each ISAT. Using the microdata of each ISAT, we then calculate the share of students in each country reaching the thresholds in the overall distribution of the ISAT. The information from the different ISATs is again combined by taking a simple average of the shares across tests.

Trends over time. The standardized performance information over a long period of time also allows us to derive longitudinal patterns of test scores for countries that participated both in early and recent ISATs. Given the amplification of noise in first-differenced data and the limitations of our rescaling method for poorly performing countries mentioned above, we perform the trend estimation only for the sample of 15 OECD countries that participated both in an ISAT before 1985 (i.e., on FIMS, FISS, FIRS, SIMS, or SISS) and up to 2003, spanning a period of nearly 20 years. (Twelve countries participated in a test before 1971, spanning a period of over 30 years.)

To estimate the trend in test performance, for each country we regress performance on the different ISATs, expressed on the standardized test metric developed above, on dummies for the subjects, dummies for the age groups, and on the year the test was conducted. The unit of observation in these country-specific regressions is each subject-by-age-by-year occasion of an ISAT, using all available tests, subjects, and age groups (see Appendix Table 10). To account for heteroscedasticity and for the fact that the signal-to-noise ratio will be larger the smaller the number of OSG countries that participated in a test, we weight the regression by

65 The sources of the underlying international test data are: Beaton et al. (1996a, 1996b), Lee and Barro (2001), Martin et al. (1997, 2000, 2004), Mullis et al. (1997, 1998, 2000, 2003, 2004), OECD (2001, 2003b, 2004), and own calculations based on the microdata of the early tests, and own calculations based on the microdata of the early tests.

66 Twenty-five of the total of 77 countries with cognitive-skill data are not included in the growth database due to lack of data on economic output or because they drop out of the sample for a standard exclusion criterion in growth analyses (15 former communist countries, 3 countries for which oil production is the dominant industry, 2 small countries, 3 newly created countries, 2 further countries lacking early output data). In addition, two strong outliers are excluded in most models (see above). There are four countries with cognitive-skill data that have a few years of economic data missing at the beginning or end of the 1960–2000 period. Data for Tunisia start in 1961, and data for Cyprus (1996), Singapore (1996), and Taiwan (1998) end slightly earlier than in 2000. These countries were included in the growth regressions by estimating average annual growth over the available 36-to-39-year period.

67 Unfortunately, the microdata from the FIMS test do not seem to be available in an accessible way any more, so that the distributional measures only draw on the remaining ISATs.

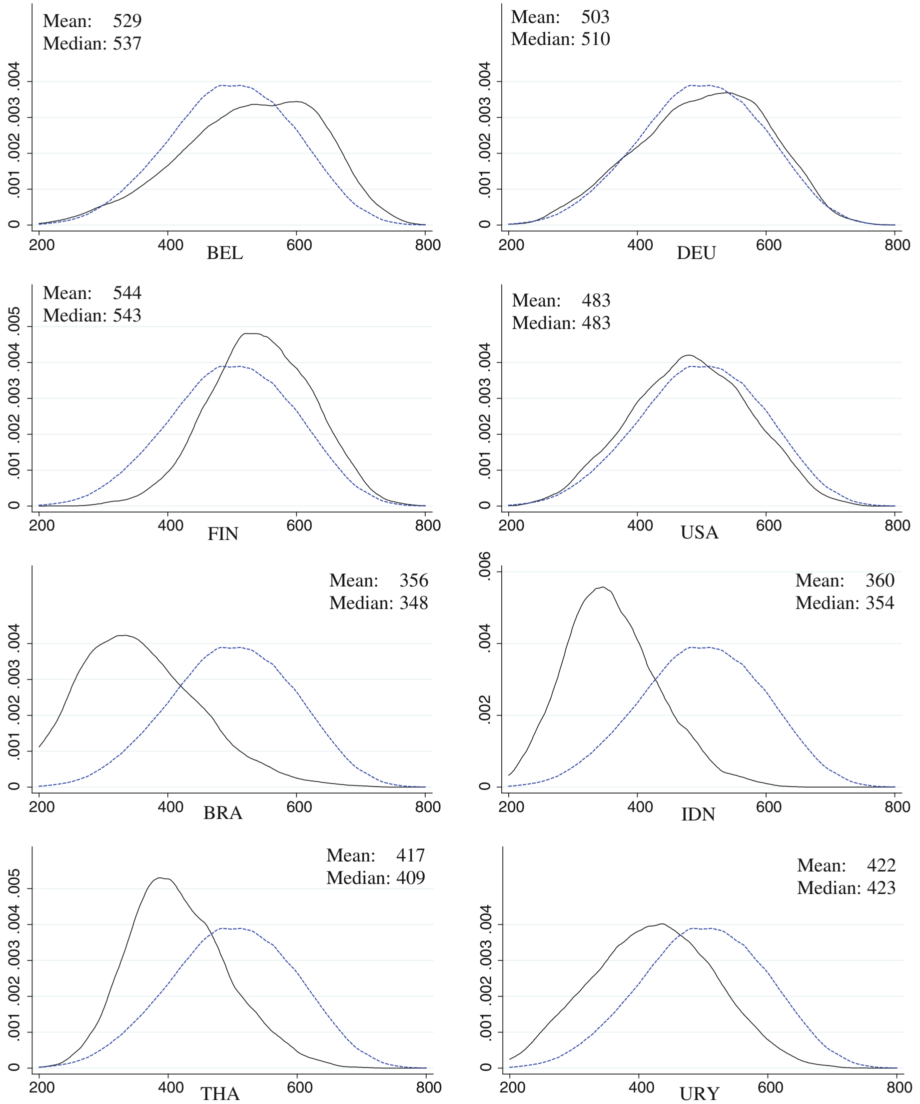

Fig. 4 Selected Examples of the Distribution of Student Performance. *Notes*: Kernel densities of student performance on the PISA 2003 math test. Bold solid line: specified country; thin dotted line: OECD countries. Codes: Belgium (BEL), Germany (DEU), Finland (FIN), United States of America (USA), Brazil (BRA), Indonesia (IDN), Thailand (THA), Uruguay (URY)

the square root of the number of OSG countries participating in each test. The coefficient on the year variable provides us with the time trend that we are interested in. The patterns captured by these country-specific regressions are shown in Fig. 5 that simply extrapolates scores for the range of 1975-2000 with scores anchored by the PISA 2000 score.

One possible worry with combining the different tests into a single measure and with estimating performance changes over time is that enrollment shares have changed to different extents over time, especially at the secondary level. To test the extent to which this affects our

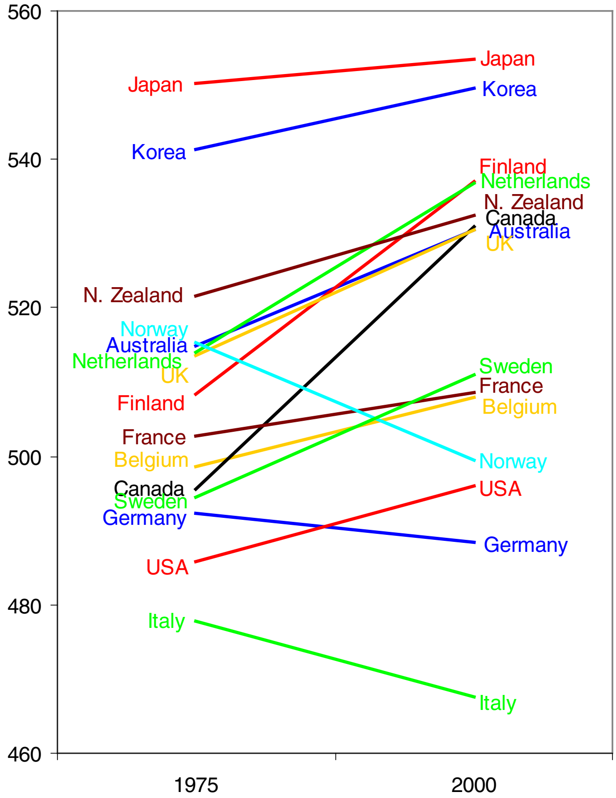

Fig. 5 Trends in Test Scores Notes: Depiction based on PISA 2000 performance and a backward induction based on the coefficient on a time variable from a regression of all available international test scores (by year, age group, and subject) on the time variable and dummies for age group and subject. See Appendix B for details

cognitive-skill measures, we calculated the correlation between our measure of trend in test scores and changes in enrollment rates. It turns out that the two are orthogonal to each other, diluting worries that differential changes in enrollment bias the results reported in this paper.[68]

Comparisons with Cognitive Skills Measure of Hanushek and Kimko (2000). The skill measures developed in Hanushek and Kimko (2000) fail to account for the unequal variances of the tests over time, but instead assume a constant variance.[69] Our measure is highly correlated with the Hanushek-Kimko measure ($r = 0.83$), but the important question is the relationship with growth. For the 30 countries in common from the two data sets, we estimate the growth models with the alternative measures of cognitive skills. While both versions of the test-score measure enter the model strongly and significantly, statistical precision is considerably higher with the new measure ($t = 7.43$ vs. $t = 4.02$), as is the explanatory power of the model (adj. $R^2 = 0.80$ vs. adj. $R^2 = 0.61$). The content of signal

[68] As noted previously, direct estimation of the impacts of school selectivity and of test exclusions on our growth models confirm that potential testing problems do not bias our growth estimation (Hanushek and Woessmann 2011c).

[69] Hanushek and Kimko (2000) actually have alternative measures. Two of their three measures assume a constant mean for all of the tests, similar to what is also done in Lee and Barro (2001).

relative to noise in the test-score measure thus seems to be considerably raised in the new measure.

Data download

The data by country may be downloaded in Excel or Stata formats at:
http://hanushek.stanford.edu/download or www.cesifo.de/woessmann#data.

Data sources

Beaton, Albert E., Ina V.S. Mullis, Michael O. Martin, Eugenio J. Gonzalez, Dana L. Kelly, and Teresa A. Smith. 1996. *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

Beaton, Albert E., Michael O. Martin, Ina V.S. Mullis, Eugenio J. Gonzalez, Teresa A. Smith, and Dana L. Kelly. 1996. *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

Lee, Jong-Wha, and Robert J. Barro. 2001. Schooling Quality in a Cross-Section of Countries. *Economica* 68, no. 272: 465-488.

Martin, Michael O., Ina V.S. Mullis, Albert E. Beaton, Eugenio J. Gonzalez, Teresa A. Smith, and Dana L. Kelly. 1997. *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

Martin, Michael O., Ina V.S. Mullis, Eugenio J. Gonzalez, Kelvin D. Gregory, Teresa A. Smith, Steven J. Chrostowski, Robert A. Garden, and Kathleen M. O'Connor. 2000. *TIMSS 1999 International Science Report:Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: Boston College.

Martin, Michael O., Ina V.S. Mullis, Eugenio J. Gonzalez, and Steven J. Chrostowski. 2004. *TIMSS 2003 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College.

Mullis, Ina V.S., Michael O. Martin, Albert E. Beaton, Eugenio J. Gonzalez, Dana L. Kelly, and Teresa A. Smith. 1997. *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

—. 1998. *Mathematics and Science Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

Mullis, Ina V.S., Michael O. Martin, Eugenio J. Gonzalez, Kelvin D. Gregory, Robert A. Garden, Kathleen M. O'Connor, Steven J. Chrostowski, and Teresa A. Smith. 2000. *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: Boston College.

Mullis, Ina V.S., Michael O. Martin, Eugenio J. Gonzalez, and Ann M. Kennedy. 2003. *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary School in 35 Countries*. Chestnut Hill, MA: International Study Center, Boston College.

Mullis, Ina V.S., Michael O. Martin, Eugenio J. Gonzalez, and Steven J. Chrostowski. 2004. *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College.

Organisation for Economic Co-operation and Development (OECD). 2001. *Knowledge and Skills for Life: First Results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.

—. 2003a. *Education at a Glance: OECD Indicators 2003*. Paris: OECD.

—. 2003b. *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000*. Paris: OECD.

—. 2004. *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD.

U.S. Department of Education, Institute of Education Sciences. 2008. *National Assessment of Educational Progress–The Nation's Report Card*. Website:http://nces.ed.gov/nationsreportcard/aboutnaep.asp.

Appendix C: Descriptive statistics

See Appendix Tables 12, 13, 14 and Figure 6.

Table 12 Descriptive statistics for the growth models

	Observations	Mean	Std. Dev.	Min	Max
Average annual growth rate in GDP per capita 1960–2000	50	2.903	1.387	0.967	6.871
Cognitive skills (all grades)	50	4.546	0.611	3.089	5.452
Cognitive skills (lower secondary)	50	4.535	0.671	2.683	5.599
Share of students reaching basic literacy	50	0.761	0.215	0.182	0.967
Share of top-performing students	50	0.062	0.054	0.000	0.219
GDP per capita 1960	50	4,991	3,676	685	14,877
Years of schooling 1960	50	5.447	2.877	0.611	10.963
Years of schooling, average 1960–2000	50	7.425	2.654	2.098	11.845

Table 12 continued

	Observations	Mean	Std. Dev.	Min	Max
External exit exam system	43	0.661	0.467	0	1
Private enrollment share	19	0.186	0.206	0	0.720
Centralization (share) of decisions on organization of instruction	17	0.104	0.117	0	0.380

Notes Descriptive statistics for variables used in Tables 1, 2, 3, 4, and 7. See main text for data sources

Table 13 Descriptive statistics for the U.S.-immigrant models

	Observations	Mean	Std. dev.	Min	Max
Annual earnings	309,574	33,243	40,983	1,000	385,000
Cognitive skills	309,574	4.334	0.535	3.089	5.452
Educated in country of origin	309,574	0.837	0.370	0	1
Years of schooling	309,574	11.558	5.006	0	20
Potential experience	309,574	24.841	11.966	0	87

Notes Descriptive statistics for variables used in Table 5. See main text for data sources

Table 14 Descriptive statistics for the changes-in-growth-paths models

	Observations	Mean	Std. Dev.	Min	Max
Trend in growth rate of GDP per capita 1975–2000	15	−0.007	0.071	−0.118	0.106
Trend in cognitive skills	15	0.409	0.546	−0.630	1.420
Average annual growth rate in GDP per capita 1975–2000	15	2.318	1.106	0.855	5.978
GDP per capita 1975	15	13,884	3,217	3,720	18,175
Change in years of schooling 1975–2000	15	1.994	0.895	0.899	4.376

Notes Descriptive statistics for variables used in Table 6. See main text for data sources

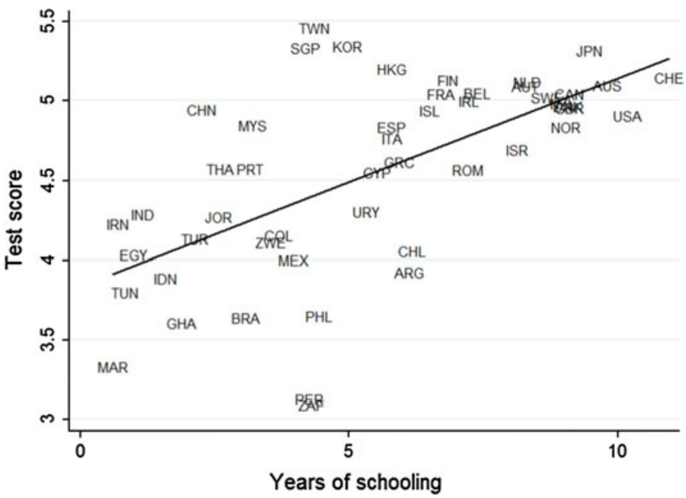

Fig. 6 Years of Schooling and Test Scores. *Notes*: Scatter plot of years of schooling in 1960 against cognitive skills (average test scores in math and science, primary through end of secondary school, all years)

Appendix D: Impact of data updates and expansion of observed growth period

The analysis in this paper concentrates on measurement issues surrounding cognitive skills. But these are not the only measurement issues that have been raised in the context of empirical growth analysis. This appendix provides an analysis of the impact of other data measurement issues in the area of data on school attainment and on economic growth. It also considers the most recent wave of cognitive skills data. These updates and revisions do not substantially change the results presented in our main analyses.

The measurement of school attainment has been discussed at various times. Barro and Lee (1993) developed the initial international database for school attainment. This was criticized by Cohen and Soto (2007), who produced the dataset that is the basis of our estimation. Most recently, Barro and Lee (2010) have produced a new dataset of their own. Using the latest series of school attainment developed by Barro and Lee (2010) (data version 1.0, 03/10, accessed on May 17, 2010) has no effect on the estimates of the impact of cognitive skills and only slightly increases the attainment coefficient. As shown in Table 15, column 2, the estimated cognitive skills coefficient is 1.92, as opposed to 1.98 in our base estimates (column 1).

A second area of concern is the underlying economic data themselves. It has recently been argued that changes in the Penn World Tables (PWT), particularly the 6.2 revision, lead to significant alternations in standard estimates of growth models (Ciccone and Jarocinski (2010); Appleton et al. (2011)).[70] To assess the impact of these data, we compare the most recent version of the PWT (version 7.0, released on June 3, 2011; Heston et al. (2011)) to our estimates based on version 6.1. As is apparent from column 3, the estimate on cognitive skills is hardly affected when estimating the same model with the same growth period (1960-2000) using the new economic data; in fact, the point estimate is slightly higher.

The new version of the PWT allows us to expand the growth period to 2007. Again, results are strongly confirmed in this 47-year growth period (column 4), with the point estimate closer to the original estimate again. We use 2007 as the endpoint of the considered growth period rather than 2009 (which is the latest year available in the new PWT) because we do not want the long-run growth analysis being affected by the global recession that started at the end of 2008 (as clearly visible in the PWT data, where average growth rates in our sample drop from around 4 % in the preceding years to 1.6 in 2007-2008 and to -2.5 in 2008-2009). Still, our result on cognitive skills is confirmed also in the 1960-2009 growth period (not shown), with a point estimate of 1.76 ($t = 7.32$).

The expanded PWT data also allow us to perform an analysis where the observation period of test scores strictly pre-dates the observation period of economic growth. For a sample of 25 countries, we have test-score data observed between 1964 and 1984. Column 5 uses these test-score data to predict economic growth in 1985-2007. Again, the significant effect of cognitive skills on growth is confirmed, with a point estimate substantially larger than in the base model. (Again, results are very similar for the growth period expanded to 2009.)

While we are reluctant to perform analyses on shorter growth periods because these are prone to country-specific shocks and business-cycle fluctuations, we can expand our sample of countries with test scores pre-dating the observed growth period to 37 countries when we use the tests conducted until 1995 to predict growth in 1995-2009 or in 2000-2009. Again, results confirm a strong estimate on cognitive skills with a point estimate larger than in

[70] As noted in section 7 above, calculation of growth rates relying on national account data rather than on purchasing power parity incomes does not significantly affect our estimates.

Table 15 Growth regressions with updated data series on school attainment and economic growth

PWT version	6.1	6.1	7.0	7.0	7.0
Years of schooling data	Cohen and Soto (2007)	Barro and Lee (2010)	Barro and Lee (2010)	Barro and Lee (2010)	Barro and Lee (2010)
Growth period	1960–2000	1960–2000	1960–2000	1960–2007	1985–2007
Test scores	All years	All years	All years	All years	Until 1984
	(1)	(2)	(3)	(4)	(5)
Cognitive skills	1.980	1.921	2.133	1.881	3.593
	(9.12)	*(9.25)*	*(9.01)*	*(7.78)*	*(2.56)*
Initial years of schooling	0.026	0.079	0.018	0.018	−0.079
	(0.34)	*(1.09)*	*(0.23)*	*(0.22)*	*(0.41)*
Initial GDP per capita	−0.302	−0.324	−0.219	−0.212	−0.123
	(5.54)	*(7.01)*	*(5.59)*	*(5.29)*	*(2.20)*
No. of countries	50	50	50	50	25
R^2 (adj.)	0.728	0.734	0.667	0.610	0.318

Notes Dependent variable: average annual growth rate in GDP per capita. Regressions include a constant. Test scores are average of math and science, primary through end of secondary school. Absolute t-statistics in parentheses

Table 16 Changes in cognitive skills and changes in growth paths with updated data series on economic growth

Period of trend in growth	1975–2000 (1)	1975–2007 (2)	1985–2007 (3)
Trend in cognitive skills	0.072	0.072	0.115
	(2.61)	*(2.92)*	*(2.94)*
Average annual growth rate	−0.017	−0.022	−0.067
in GDP per capita over period	*(1.30)*	*(1.69)*	*(3.16)*
No. of countries	15	15	15
R^2 (adj.)	0.316	0.387	0.516

Notes Dependent variable: trend in the growth rate of GDP percapita over the period shown in the header, based on PWT version 7.0. Regressions include a constant. Sample: OECD countries with test-score data both before 1985 and up to 2003. Test scores are average of math and science, primary through end of secondary school. Absolute t-statistics in parentheses

the base model (not shown), although the precise point estimate is sensitive to excluding individual countries when considering this shorter growth period.

We have also experimented with the latest wave of test-score data, the 2009 wave of the PISA study. Our test-score measure derived from the tests conducted in 1964-2003 is strongly correlated with the PISA 2009 data (correlation coefficient of 0.94 for the 37 countries available in both datasets). This corroborates the assumption of relative stability underlying our main analysis. Also, using the PISA 2009 as an alternative measure of cognitive skills in the growth regressions fully confirms our results, with a highly significant point estimate extremely close to our main analysis (1.96).

Finally, Table 16 shows that our analysis in Section 7 on trends in skills and in growth is strongly confirmed with the new 7.0 version of the PWT. Considering the revised data over the same 1975-2000 period (col. 1) hardly affects the result on cognitive skills from col. 2 of Table 6. Similarly, extending the growth period with the newly available data to 2007 (or to 2009, not shown) confirms the previous result (col. 2). Similarly, when using the growth period 1985-2007 (col. 3), so that the test-score trend partly predates the growth rate trend, even strengthens the result of a significant positive relation of trends in test scores with trends in growth rates.

Acknowledgments We thank Mark Bils, Pete Klenow, Giovanni Peri, Paul Romer, participants at seminars at Stanford University, the London School of Economics, the Universities of Stockholm, Mannheim, Heidelberg, Uppsala, and Rochester, Simon Fraser University, the European School of Management and Technology in Berlin, the City University of Hong Kong, the World Bank, the NBER growth conference in San Francisco, and the Econometric Society World Congress in Shanghai, and anonymous referees for helpful comments and suggestions. We are grateful to the International Association for the Evaluation of Educational Achievement (IEA) for providing us with the micro data for the FISS, SIMS, and SISS studies. Jason Grissom and Trey Miller provided important research assistance. Support has come from CESifo, the Packard Humanities Institute, the W. Glenn Campbell and Rita Ricardo-Campbell National Fellowship of the Hoover Institution, and the Pact for Research and Innovation of the Leibniz Association.

References

Acemoglu, D., Johnson, S., & Robinson, A. R. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review, 91*(5), 1369–1401.

Acemoglu, D., Johnson, S., & Robinson, A. R. (2005). Institutions as a fundamental cause of long-run growth. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (pp. 385–472). Amsterdam: North Holland.

Aghion, P., & Howitt, P. (1998). *Endogenous growth theory*. Cambridge, MA: MIT Press.

Aghion, P., Boustan, L., Hoxby, C. M., & Vandenbussche, J. (2009). The causal impact of education on economic growth: Evidence from the U.S. (mimeo) Department of Economics: Harvard University (March).

Andrews, D. R., Moreira, M. J., & Stock, J. H. (2007). Performance of conditional Wald tests in IV regression with weak instruments. *Journal of Econometrics, 139*(1), 116–132.

Appleton, S., Atherton, P., & Bleaney, M. (2011). Growth regressions and data revisions in Penn World Tables. *Journal of Economic Studies, 38*(3), 301–312.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2006). The polarization of the U.S. labor market. *American Economic Review, 96*(2), 189–194.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2008). Trends in U.S. wage inequality: Revising the revisionists. *Review of Economics and Statistics, 90*(2), 300–323.

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

Barro, R. J. (1997). *Determinants of economic growth: A cross-country empirical study*. Cambridge, MA: MIT Press.

Barro, R. J. (2001). Human capital and growth. *American Economic Review, 91*(2), 12–17.

Barro, R. J., & Lee, J.-W. (1993). International comparisons of educational attainment. *Journal of Monetary Economics, 32*(3), 363–394.

Barro, R. J., & Lee, J.-W. (2010). A new data set of educational attainment in the world, 1950-2010. NBER Working Paper 15902. Cambridge, MA: National Bureau of Economic Research (April).

Barro, R. J., & McCleary, R. M. (2003). Religion and economic growth across countries. *American Sociological Review, 68*(5), 760–781.

Barro, R. J., & Ursúa José, F. (2008). Macroeconomic crises since 1870. *Brookings Papers on Economic Activity, 1*, 336–350.

Bils, M., & Klenow, P. J. (2000). Does schooling cause growth? *American Economic Review, 90*(5), 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.

Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature, 39*(4), 1137–1176.

Castelló, A., & Doménech, R. (2002). Human capital inequality and economic growth: Some new evidence. *Economic Journal, 112*(478), C187–C200.

Ciccone, A., & Jarocinski, M. (2010). Determinants of economic growth: Will data tell? *American Economic Journal: Macroeconomics, 2*(4), 222–246.

Ciccone, A., & Papaioannou, E. (2009). Human capital, the structure of production, and growth. *Review of Economics and Statistics, 91*(1), 66–82.

Cingano, F., Leonardi, M., Messina, J., & Pica, G. (2010). The effects of employment protection legislation and financial market imperfections on investment: evidence from a firm-level panel of EU countries. *Economic Policy, 25*(61), 117–163.

Cohen, D., & Soto, M. (2007). Growth and human capital: Good data, good results. *Journal of Economic Growth, 12*(1), 51–76.

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

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review, 97*(2), 31–47.

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education.* (pp. 697–812). Amsterdam: Elsevier.

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), 7–55.

Ehrlich, I. (2007). The mystery of human capital as engine of growth, or why the US became the economic superpower in the 20th Century. NBER Working Paper 12868. Cambridge, MA: National Bureau of Economic Research (January).

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.
Glenn, C. L., & De Groof, J. (2002). *Finding the right balance: Freedom, autonomy and accountability in education, Vol. II*. The Netherlands: Lemma Publishers.
Goldin, C., & Katz, L. F. (2008). *The race between education and technology*. Cambridge, MA: Harvard University Press.
Gundlach, E., Woessmann, L., & Gmelin, J. (2001). The decline of schooling productivity in OECD countries. *Economic Journal, 111*(May), C135–C147.
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.
Hanousek, J., Hajkova, D., & Filer, R. K. (2008). A rise by any other name? Sensitivity of growth regressions to data source. *Journal of Macroeconomics, 30*(3), 1188–1206.
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. A., & Kimko, D. D. (2000). Schooling, labor force quality, and the growth of nations. *American Economic Review, 90*(5), 1184–1208.
Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46*(3), 607–668.
Hanushek, E. A., & Woessmann, L. (2011a). 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. (2011b). How much do educational outcomes matter in OECD countries? *Economic Policy, 26*(67), 427–491.
Hanushek, E. A., & Woessmann, L. (2011c). Sample selectivity and the validity of international student achievement tests in economic research. *Economic Letters, 110*(2), 79–82.
Hanushek, E. A., & Woessmann, L. (2012). Schooling, educational achievement, and the Latin American growth puzzle. *Journal of Development Economics* (forthcoming).
Hanushek, E. A., & Zhang, L. (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital, 3*(2), 107–143.
Hausmann, R., Pritchett, L., & Rodrik, D. (2005). Growth accelerations. *Journal of Economic Growth, 10*(4), 303–329.
Heckman, J. J., Lochner, L. J., & Todd, P. E. (2008). Earnings functions and rates of return. *Journal of Human Capital, 2*(1), 1–31.
Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411–482.
Hendricks, L. (2002). How important is human capital for development? Evidence from immigrant earnings. *American Economic Review, 92*(1), 198–219.
Heston, A., Summers, R., & Aten, B. (2002). Penn World Table Version 6.1. Center for International Comparisons at the University of Pennsylvania (CICUP). Philadelphia: University of Pennsylvania.
Heston, A., Summers, R., & Aten, B. (2011). Penn World Table Version 7.0. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania Philadelphia: University of Pennsylvania (June 3 update).
Jones, B. F., & Olken, B. A. (2008). The anatomy of start-stop growth. *Review of Economics and Statistics, 90*(3), 582–587.
Katz, L. F., & Autor, D. H. (1999). Changes in the wage structure and earnings inequality. In A. Orley & D. Card (Eds.), *Handbook of Labor Economics.* (pp. 1463–1558). Amsterdam: Elsevier.
Krueger, A. B., & Lindahl, M. (2001). Education for growth: Why and for whom? *Journal of Economic Literature, 39*(4), 1101–1136.
Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review, 10*(3), 179–214.
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. *American Economic Review, 82*(4), 942–963.
Lucas, R. E., Jr. (1988). On the mechanics of economic development. *Journal of Monetary Economics, 22*(1), 3–42.
Mankiw, N. G., Romer, D., & Weil, D. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics, 107*(2), 407–437.
Mincer, J. (1974). *Schooling, experience, and earnings*. New York: NBER.
Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica, 71*(4), 1027–1048.

Mulligan, C. B. (1999). Galton versus the human capital approach to inheritance. *Journal of Political Economy 107*(6), pt. 2 (December): S184–S224.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management, 19*(4), 547–568.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics, 77*(2), 251–266.

Neidorf, T. S., Binkley, M., Gattis, K., & Nohara, D. (2006). *Comparing mathematics content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 assessments*. Washington: National Center for Education Statistics.

Nelson, R. R., & Phelps, E. (1966). Investment in humans, technology diffusion and economic growth. *American Economic Review, 56*(2), 69–75.

Nicoletti, G., & Scarpetta, S. (2003). Regulation, productivity, and growth: OECD evidence. *Economic Policy, 18*(1), 10–72.

Nuxoll, D. A. (1994). Differences in relative prices and international differences in growth rates. *American Economic Review, 84*(5), 1423–1436.

Organisation for Economic Co-operation and Development. (1998). *Education at a glance: OECD Indicators*. Paris: Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD.

Pritchett, L. (2006). Does learning to add up add up? The returns to schooling in aggregate data. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education.* (pp. 635–695). Amsterdam: North Holland.

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

Riley, J. G. (2001). Silver signals: Twenty-five years of screening and signaling. *Journal of Economic Literature, 39*(2), 432–478.

Romer, P. (1990). Endogenous technological change. *Journal of Political Economy 99*(5), pt. II: S71–S102.

Sachs, J. D., & Warner, A. M. (1995). Economic reform and the process of global integration. *Brookings Papers on Economic Activity* (1),1–96.

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.

Schoellman, T. (2012). Education quality and development accounting. *Review of Economic Studies, 79*(1), 388–417.

Spence, A. M (1973). Job market signalling. *Quarterly Journal of Economics, 87*(3), 355–374.

Topel, R. (1999). Labor markets and economic growth. In A. Orley & D. Card (Eds.), *Handbook of Labor Economics.* (pp. 2943–2984). Amsterdam: Elsevier.

UNESCO. (1998). *World education report, 1998: Teachers and teaching in a changing world*. Paris: UNESCO.

UNESCO. (2005). *Education for all: The quality imperative, EFA Global Monitoring Report*. Paris: UNESCO.

Vandenbussche, J., Aghion, P., & Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *Journal of Economic Growth, 11*(2), 97–127.

Weiss, A. (1995). Human capital vs. signalling explanations of wages. *Journal of Economic Perspectives, 9*(4), 133–154.

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. *Economic Journal, 120*(546), F229–F255.

Woessmann, L. (2003a). 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. (2003b). Schooling resources, educational institutions, and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics, 65*(2), 117–170.

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