Educational Outcomes

SUMMARY

Existing growth research provides little explanation for the very large differences in long-run growth performance across OECD countries. We show that cognitive skills can account for growth differences within the OECD, whereas a range of economic institutions and quantitative measures of tertiary education cannot. Under the growth model estimates and plausible projection parameters, school improvements falling within currently observed performance levels yield very large gains. The present value of OECD aggregate gains through 2090 could be as much as $275 trillion, or 13.8% of the discounted value of future GDP for plausible policy changes. Extensive sensitivity analyses indicate that, while different model frameworks and alternative parameter choices make a difference, the economic impact of improved educational outcomes remains enormous. Interestingly, the quantitative difference between an endogenous and neoclassical model framework – with improved skills affecting the long-run growth rate versus just the steady-state income level – matters less than academic discussions suggest. We close by discussing evidence on which education policy reforms may be able to bring about the simulated improvements in educational outcomes.

– Eric A. Hanushek and Ludger Woessmann

Economic Policy July 2011 Printed in Great Britain

How much do educational outcomes matter in OECD countries?

Eric A. Hanushek and Ludger Woessmann

Hoover Institution, Stanford University, NBER and CESifo; University of Munich, Ifo Institute for Economic Research, CESifo and IZA

1. INTRODUCTION

Despite its surge over the past two decades, research in the economics of growth – both theoretical and empirical – has produced surprisingly few resilient results about policies that might promote long-run growth in developed countries (cf. Aghion and Howitt, 2006). Most of the robust results that exist refer either to the importance of basic economic institutions, with important policy implications for developing countries, or to policies that affect short- to medium-term growth in developed countries. Here we present evidence that improved human capital, measured by cognitive skills, has the potential for substantial improvements in the long-run economic well-being of OECD countries.

This paper was presented at the 52nd Panel Meeting of *Economic Policy* in Rome. We thank four anonymous referees, Luigi Pistaferri, Fabiano Schivardi and the other panel participants for their comments. Woessmann gratefully acknowledges the support and hospitality provided by the W. Glenn Campbell and Rita Ricardo-Campbell National Fellowship of the Hoover Institution, Stanford University, as well as support by the Pact for Research and Innovation of the Leibniz Association. Hanushek was supported by the Packard Humanities Institute.
The Managing Editor in charge of this paper was Tullio Jappelli.

Economic Policy July 2011 pp. 427–491 Printed in Great Britain

The immense variation in the long-run growth experiences of developed countries has largely escaped notice. For example, from 1960 to 2000, GDP per capita grew on average by less than 1.5% per year in New Zealand and Switzerland, but by more than 4% per year in Ireland, Japan and South Korea. As a consequence, the average Korean was about 10 times as well off in 2000 as in 1960, and the average Irish and Japanese about 5 times. By contrast, the average New Zealander and Swiss were only 1.6–1.8 times as well off than 40 years before. These stark differences are directly visible when comparing the three fastest-growing and the three slowest-growing countries highlighted (together with the United States) in Figure 1, which plots GDP per capita in 1960 and 2000: Korea surpassed several other OECD countries, including Mexico; Japan and Ireland went from 40–45% to 131–140% of New Zealand's income; and Ireland caught up to Switzerland from an initial level of 35% of its GDP per capita.

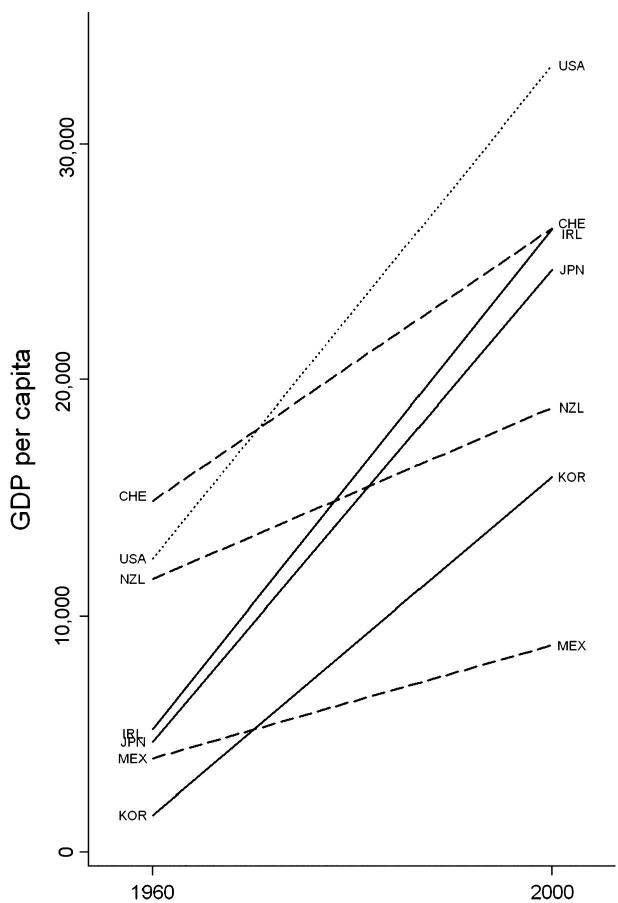

Figure 1. GDP per capita in fastest- and slowest-growing OECD countries in 1960 and 2000

Notes: GDP per capita in constant international dollars. See Table 1 for data for all OECD countries.

Source: Authors' depiction based on data from Heston *et al.* (2002).

Following prior work on the difference between developing and developed countries, this paper focuses on the role of human capital as measured by cognitive skills in explaining these long-run growth differences among OECD countries. Our analysis of growth differences relies largely on the 24 OECD countries with consistent data on cognitive skills and economic growth, although we also provide some relevant comparisons with the expanded sample of 50 countries that incorporates non-OECD countries.[1] Within the OECD, long-run growth is closely related to cognitive skills.

Perhaps the leading candidate for being a more fundamental explanation of growth than human capital is the quality of a country's economic institutions such as having secure property rights or an open economy along with regulations of labour and product markets, bureaucratic burdens and the like. But, we show that these institutions do not help us understand differences in long-run growth rates among OECD countries.

We then turn to considering how educational policy differs across countries. Policy choices are most readily seen in very different levels of tertiary education, but they also show up in varying levels of achievement. For example, while Australia currently has 87% of a cohort entering tertiary education, Norway has 71%, and Italy has 51% (Organization for Economic Co-operation and Development, 2010). We separately measure basic and top skills, based on the micro data of the international achievement tests, and then consider the implications both of high skills and of tertiary schooling on growth. The results suggest that basic skills have substantial growth pay-offs in OECD countries and that, if anything, the return to top skills is lower, not higher, in the OECD. Further, we do not find a specific role of tertiary attainment for OECD growth once direct measures of skills are taken into consideration.

The analysis of growth regressions allows us to project the economic value of education reform alternatives for each OECD country. We evaluate the economic outcomes of a series of plausible policy programmes including improving student performance by 25 PISA points (1/4 standard deviation); bringing all countries in the OECD up to the level of Finland; and bringing all students in OECD countries to minimum proficiency (400 points on the PISA tests). As discussed below, history indicates that at least the first two reforms fall within international experience.

The present value of the reform efforts varies by country, depending on current economic and educational performance. However, the simulation exercises suggest that the aggregate gains across all OECD countries range from $90 trillion to $275 trillion for the different policy alternatives. These gains, for example, far exceed the level of stimulus funds in the recent global recession. We provide detailed sensitivity analyses of the simulations with respect to a range of alternative specifications and

[1] With the limited country samples, there is a distinct trade-off between incorporating the added observations from the full world sample and restricting the economic relationships to being the same across all countries. Throughout the analysis, we provide information on the similarities and differences of developed and developing countries.

parameter choices but find that the qualitative picture of huge returns to improving skills remains.

Knowledge of the potential gains from improving schools does not, however, indicate what should be done to obtain these results. In fact, school improvement has been high on the policy agenda of a large number of OECD countries, but the results of actions have many times fallen short of expectations. Experience suggests that simple increases in school resources do not consistently improve outcomes. Teacher quality is very important, but measuring and regulating quality is exceedingly difficult – suggesting that indirect policies are essential. Emerging research results suggest that there are general policies related to the institutional structure of schools that can promote significantly higher achievement. Institutional elements – involving choice and competition, decentralization and school autonomy, performance pay and outcome accountability – positively alter the incentives in schools and, according to existing research, promote higher achievement.

2. CONCEPTUAL FRAMEWORK

Economists have considered the process of economic growth for much of the last 100 years, but most studies remained as theory with little empirical work. Over the past two decades, economists linked analysis much more closely to empirical observations and in the process rediscovered the importance of growth. Our analysis, mirroring much of the recent empirical work, concentrates on the role of human capital.

Prior theoretical and empirical work has pursued a variety of specifications of the underlying growth process (see the reviews in Hanushek and Woessmann, 2008, 2010a). Nonetheless, the restricted variation of experiences across countries plus general data limitations have made it difficult to distinguish among the competing models of growth – and such is the case here.

We model a country's growth rate as a function of the skills of workers and other factors 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.

$$growth = \alpha_1\, human\ capital + \alpha_2\, other\ factors + \varepsilon \qquad (1)$$

This formulation suggests that nations with more human capital tend to continue to make greater productivity gains than nations with less human capital, although we consider the possibility that the induced growth in productivity disappears over time.[2]

[2] In terms of the major theoretical distinctions, our formulations combine key elements of the competing models. The fact that the rate of technological change and productivity improvement is directly related to the stock of human capital of the nation makes it an endogenous growth model. At the same time, by including the initial level of income among the control variables, our model does allow for conditional convergence, a leading feature of the 'augmented neoclassical' approach, the commonly suggested alternative view.

The empirical macroeconomic literature focusing on cross-country differences in economic growth has overwhelmingly employed measures related to school attainment, or years of schooling, to test the human capital aspects of growth models. It has tended to find a significant positive association between quantitative measures of schooling and economic growth. To give an idea of the robustness of this association, an extensive empirical analysis by Sala-i-Martin *et al.* (2004) of 67 explanatory variables in growth regressions on a sample of 88 countries found that primary schooling was the most robust influence factor (after an East Asian dummy) on growth in GDP per capita in 1960–96.

Nevertheless, we believe that these formulations introduce substantial bias into the estimation. Average years of schooling is a particularly incomplete and potentially misleading measure of education for comparing the impacts of human capital on the economies of different countries. It implicitly assumes 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 South Africa is assumed to create the same increase in productive human capital as a year of schooling in Korea. Additionally, formulations relying on this measure assume that formal schooling is the primary (sole) source of education and that variations in non-school factors have negligible effects on education outcomes and skills. This neglect of cross-country differences in the quality of education and in the strength of family, health and other influences is probably the major drawback of such a quantitative measure of schooling.

To see this, consider a standard version of an education production function as employed in a very extensive literature (for a review see Hanushek, 2002), where skills are expressed as a function of a range of factors:

$$\begin{aligned} \textit{human capital} = {} & \beta_1 \textit{family inputs} + \beta_2 \textit{schooling inputs} \\ & + \beta_3 \textit{individual ability} + \beta_4 \textit{other factors} + v \end{aligned} \qquad (2)$$

In general, human capital combines school attainment and quality with other relevant factors including education in the family, labour market experience, health, and so forth.

Thus, while school attainment has been convenient in empirical work because of its ready availability across countries, its use ignores differences in school quality in addition to other important determinants of people's skills. A more satisfying alternative is to incorporate variations in cognitive skills, which can be determined by results of international assessments of mathematics, science and reading achievement, as a direct measure of the human capital input into empirical analyses of economic growth. The focus on cognitive skills has a number of potential advantages. First, it captures variations in the knowledge and ability that schools strive to produce and thus relates the putative outputs of schooling to subsequent economic success. Second, by emphasizing total outcomes of education, it incorporates 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), it opens the investigation of the importance of different policies designed to affect the quality aspects of schools. Fourth, it is practical because of the extensive development of consistent and reliable cross-country assessments.

Our analysis relies on the measures of cognitive skills developed in Hanushek and Woessmann (2009). Between 1964 and 2003, twelve different international tests of maths, science or reading were administered to a voluntarily participating group of countries (for a review see Hanushek and Woessmann, 2011). 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). The assessments are designed to identify a common set of expected skills, which are then tested in the local language. Each test is newly constructed, until recently with no effort to link to any of the other tests. Hanushek and Woessmann (2009) describe the construction of consistent measures at the national level across countries through empirical calibration of the different tests. By transforming the means and variances of the original country scores (partly based on external longitudinal test score information available for the United States), each is placed into a common distribution of outcomes. Each age group and subject is normalized to the PISA standard of mean 500 and individual standard deviation of 100 across OECD countries, and then all available test scores are aggregated to the country level.

We interpret the test scores as an index of the human capital of the populations (and workforce) of each country. This interpretation of our averages over different cohorts is reasonable if a country's scores have been stable across time, implying that estimates from the current school-aged population provide an estimate of the older working population. If scores (and skills) change over time, some measurement error is clearly introduced. Scores have in fact changed some (Hanushek and Woessmann, 2009), but within our period of observations it still appears that the differences in levels dominate any intertemporal score changes. Nonetheless, any measurement error in this case will tend to bias downward the estimates of the impact of cognitive skills on growth, so that our estimates of economic implications will be conservative.

The data on GDP per capita and its growth for our analyses come from the Penn World Tables (Heston *et al.*, 2002). Quantitative educational attainment data are taken from the latest version of the Barro and Lee (2010) database (data version 1.0, 3/10, accessed on May 17, 2010). Additional measures of specific control variables will be discussed in the relevant sections below.

Table 1 provides basic descriptive statistics on the combined measure of educational performance and the underlying economic data. We have already discussed

Table 1. Income, growth, and schooling in OECD countries, 1960–2000

		Growth			GDP per capita			Years of schooling			Test scores		
		1960–2000	1960–1980	1980–2000	1960	1980	2000	1960	1980	2000	Mean	Basic	Top
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Australia	AUS	2.22	2.41	2.03	10,618	17,092	25,535	9.3	11.5	11.8	509.4	93.8	11.2
Austria	AUT	2.96	3.86	2.07	7,365	15,706	23,681	4.2	7.3	9.0	508.9	93.1	9.7
Belgium	BEL	2.84	3.78	1.91	7,759	16,303	23,784	6.9	8.5	10.2	504.1	93.1	9.4
Canada	CAN	2.40	3.06	1.75	10,419	19,022	26,922	8.3	9.7	11.0	503.8	94.8	8.3
Denmark	DEN	2.25	2.61	1.90	10,917	18,282	26,627	7.9	9.3	9.7	496.2	88.8	8.8
Finland	FIN	2.95	3.73	2.17	7,438	15,484	23,798	5.7	8.3	8.2	512.6	95.8	12.4
France	FRA	2.65	3.68	1.63	7,860	16,201	22,371	4.2	6.0	9.5	504.0	92.6	8.5
Greece	GRE	3.19	5.34	1.09	4,159	11,767	14,625	7.4	7.1	8.9	460.8	79.8	4.2
Iceland	ICE	2.84	4.07	1.61	8,106	18,017	24,794	6.2	7.9	9.6	493.6	90.8	7.4
Ireland	IRE	4.14	3.30	4.98	5,208	9,978	26,379	8.0	9.9	10.9	499.5	91.4	9.4
Italy	ITA	2.95	4.08	1.83	6,817	15,161	21,794	4.9	6.5	9.1	475.8	87.5	5.4
Japan	JAP	4.26	6.24	2.31	4,657	15,631	24,672	8.0	9.3	10.9	531.0	96.7	16.8
Korea, Rep.	KOR	5.95	5.78	6.13	1,571	4,830	15,881	4.3	8.3	11.1	533.8	96.2	17.8
Mexico	MEX	2.00	3.30	0.71	3,970	7,603	8,766	2.8	4.9	7.6	399.8	48.9	0.9
Netherlands	NET	2.44	2.82	2.06	9,263	16,164	24,313	6.3	9.4	10.7	511.5	96.5	9.2
New Zealand	NEW	1.23	1.07	1.38	11,555	14,304	18,824	10.2	11.8	12.2	497.8	91.0	10.6
Norway	NOR	3.02	3.62	2.42	8,239	16,772	27,044	7.6	9.1	11.3	483.0	89.4	5.6
Portugal	POR	3.92	4.95	2.89	3,434	9,024	15,955	3.2	5.5	7.6	456.4	80.3	3.2
Spain	SPA	3.43	4.59	2.27	4,693	11,520	18,055	3.3	6.2	9.3	482.9	85.9	7.9
Sweden	SWE	2.15	2.69	1.61	10,112	17,179	23,662	7.3	9.4	11.1	501.3	93.9	8.8
Switzerland	SWI	1.45	2.05	0.85	14,877	22,320	26,422	7.6	10.3	9.6	514.2	91.9	13.4
Turkey	TUR	2.35	2.38	2.32	2,700	4,325	6,838	1.6	3.6	6.0	412.8	58.2	3.9
United Kingdom	GBR	2.10	1.98	2.21	9,682	14,340	22,188	7.0	8.1	9.0	495.0	92.9	8.8
United States	USA	2.50	2.75	2.25	12,414	21,337	33,308	9.2	12.2	13.0	490.3	91.8	7.3
Mean		2.84	3.51	2.18	7,660	14,515	21,927	6.3	8.3	9.9	490.8	88.1	8.7
Minimum		1.23	1.07	0.71	1,571	4,325	6,838	1.6	3.6	6.0	399.8	48.9	0.9
Maximum		5.95	6.24	6.13	14,877	22,320	33,308	10.2	12.2	13.0	533.8	96.7	17.8

Notes: Sample: 24 OECD countries with income and education data for the whole period. Growth = average annual growth rate in GDP per capita (in percent). GDP per capita: in constant international dollars. Years of schooling: population aged 15 years and above. Test scores: average of maths and science, all available tests (1964–2003); basic/top: percentage of students over 400/600.

Sources: Heston *et al.* (2002); Barro and Lee (2010); Hanushek and Woessmann (2009).

the wide variation in growth rates across OECD countries. What is also clear from Table 1 is that both school attainment and test scores vary widely, suggesting directly that any impact of these human capital measures on growth differences should be easily detected.

3. BASIC GROWTH MODELS FOR OECD COUNTRIES

Our cross-country regressions follow a growing literature which, over the past ten years, demonstrates that consideration of cognitive skills dramatically alters the assessment of the role of education and knowledge in economic development. Analyzing growth in 1960–90 for a sample of 31 countries with available data (including 18 OECD countries), Hanushek and Kimko (2000) first showed a statistically and economically significant positive relationship between cognitive skills and economic growth. This relationship has been subsequently confirmed in a range of studies with different focuses (for a complete review see Hanushek and Woessmann, 2008, 2011). Most recently, Hanushek and Woessmann (2009) extend the empirical analysis to incorporate 50 countries that have participated in one or more international testing occasions between 1964 and 2003 and have aggregate economic data for the period 1960–2000. We use that database for our analysis focused on OECD countries.

As a starting point for our analyses, we replicate the basic analysis, only replacing the extended version of the Cohen and Soto (2007) data on years of schooling by the newly available latest version of the Barro and Lee (2010) database on years of schooling. Our sample contains the 24 OECD countries with available data. From the total of 30 OECD countries, the sample misses four countries – the Czech Republic, Hungary, Poland and the Slovak Republic – because their communist history prevents them from having internationally comparable economic data during the period of analysis. In addition, Germany drops out because of missing economic and test score data for the Eastern parts before 1990, and Luxembourg is left out as a small country with a population of less than one million, as is common practice (see Mankiw *et al.*, 1992).

Table 2 presents the basic results on the association between educational outcomes and long-run economic growth in the sample of OECD countries. The inclusion of initial GDP per capita in all specifications simply reflects the fact that it is easier to grow when one is farther from the technology frontier, because one just must imitate others rather than invent new things.

When the cognitive-skill data are ignored (column 1), years of schooling in 1960 are significantly associated with average annual growth rates in real GDP per capita in 1960–2000. However, once our test-score measure of human capital is included, we see that cognitive skills are highly significant while years of schooling becomes statistically insignificant and drops to close to zero. Furthermore, the OECD-sample growth variance explained by the model increases from 56% to 83% when measuring human capital by cognitive skills rather than years of schooling. Note that in the

Table 2. Basic results on educational outcomes and long-run economic growth in OECD countries

				1960–1980	1980–2000	Avg. years[a]	GDP>10,000[b]	Growth<5%[c]	Full country sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cognitive skills		1.864	1.968	1.737	2.350	1.735	1.398	1.525		1.978
		(5.83)	*(6.72)*	*(2.95)*	*(3.91)*	*(5.26)*	*(2.89)*	*(4.33)*		*(7.98)*
Initial years of schooling	0.173	0.046		0.024	0.147	0.096	0.029	0.056	0.396	0.080
	(2.09)	*(0.82)*		*(0.23)*	*(1.50)*	*(1.40)*	*(0.50)*	*(1.06)*	*(2.89)*	*(1.07)*
Initial GDP per capita	−0.293	−0.303	−0.285	−0.338	−0.250	−0.311	−0.302	−0.273	−0.405	−0.313
	(5.20)	*(8.61)*	*(10.33)*	*(5.23)*	*(6.18)*	*(9.55)*	*(8.58)*	*(7.41)*	*(4.67)*	*(5.61)*
OECD									1.437	0.859
									(1.68)	*(0.32)*
OECD × Cognitive skills										−0.203
										(0.36)
OECD × Initial years of schooling									−0.110	
									(0.65)	
No. of countries	24	24	24	24	24	24	22	23	50	50
F (OECD and interaction)									2.37	0.10
R^2 (adj.)	0.559	0.828	0.831	0.630	0.658	0.839	0.835	0.738	0.297	0.723

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960–2000 (unless noted otherwise). Sample: OECD countries (unless noted otherwise). Regressions include a constant. *t*-statistics in parentheses.

[a]Years of schooling averaged between 1960 and 2000.

[b]Sample excludes Mexico and Turkey.

[c]Sample excludes Korea.

Source: Authors' calculations based on data in Heston *et al.* (2002), Hanushek and Woessmann (2009), and Barro and Lee (2010).

OECD sample, the bivariate association with initial per-capita GDP already accounts for 49% of the variance in subsequent growth, making the relative increase in understanding non-convergent growth through cognitive skills substantial.

The estimated coefficient on cognitive skills implies that an increase of one standard deviation in educational achievement (i.e., 100 test-score points on the PISA scale) yields an average annual growth rate over 40 years that is 1.86 percentage points higher. This historical experience suggests a very powerful response to improvements in educational outcomes, particularly when compared to the average 2.2% annual growth within the OECD over the past two decades.

Figure 2 depicts the fundamental association graphically, plotting growth in real per-capita GDP between 1960 and 2000 against average test scores after allowing for differences in initial GDP per capita and average years of schooling. With the slight exceptions of New Zealand (below the regression line) and the United States (above) – to which we return below – the OECD countries align closely along the regression line that depicts the positive association between cognitive skills and economic growth.

Column 3 of Table 2 reports the same model excluding years of schooling, whose effect could not be significantly differentiated from zero. The point estimate on cognitive skills, as well as the adjusted R^2, increases slightly in this reduced model.

Columns 4 and 5 break down the analysis into the 20-year sub-periods of 1960–80 and 1980–2000. The positive association of growth with cognitive skills is

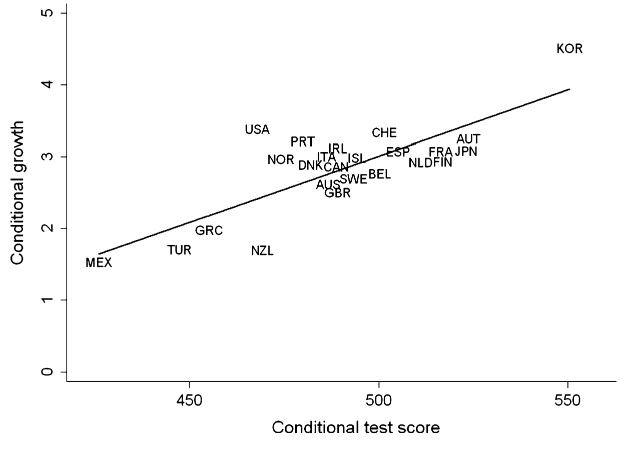

Figure 2. Educational performance and economic growth across OECD countries

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, average test scores on international student achievement tests, and average years of schooling in 1960 (mean of the unconditional variables added to each axis).

Source: Authors' depiction based on the regression analysis reported in column 2 of Table 2.

clearly visible in both sub-periods, with the point estimate slightly larger in the later period.

To reduce concerns of reverse causality between economic growth and quantitative schooling investments, the basic model uses school attainment in 1960 before the growth period, but combines that with average test scores over the entire period. However, the results are hardly affected by using average years of attainment across the period (column 6).

Because of the limited sample size, we want to ensure that results are not driven by individual outliers. Column 7 excludes Mexico and Turkey, two countries at the bottom of the sample of OECD countries today in terms of measures of GDP per capita, socioeconomic background, and educational spending. While the point estimate on cognitive skills is slightly reduced, the association remains strong and statistically highly significant. The same is true when excluding Korea (column 8), a country with extraordinary conditional test scores and growth experience (see Figure 2).

The final columns provide tests for differences in the education-growth nexus between OECD and non-OECD countries. The sample now is the full sample of 50 countries with data on test scores and growth. In the model without cognitive skills (column 9), the significant association between years of schooling and growth does not differ significantly between OECD and non-OECD countries. However, the OECD dummy is marginally significant at the 10% level and positive, indicating a remaining growth advantage of OECD countries unexplained by the model. But once cognitive skills are included in column 10, neither the OECD dummy nor its interaction with cognitive skills are statistically significant, indicating that the OECD countries actually fit well within the rest of the world on this association.

4. CAN INSTITUTIONS EXPLAIN DIFFERENCES IN RICH-COUNTRY GROWTH?

Most economists believe that fundamental economic institutions such as clear property rights and openness to international trade are important for a well-functioning economy, and by implication for economic growth, a position strongly supported in the review by Acemoglu *et al.* (2005). But, because these factors are embedded in most OECD economies, attention to institutions has also moved to regulations and restrictions in labour, capital and product markets. We investigate both areas and find little reason to believe that either affects long-run growth of OECD countries.

4.1. Property-rights and free-trade institutions

Two measures have been most consistently found in past growth analyses to be associated with growth and income differences in global country samples. The first is a measure of security of property rights – an index of the protection against expropriation risk, averaged over 1985–95, from Political Risk Services, a private company which assesses the risk that investments will be expropriated in different countries. The second is the measure of openness proposed by Sachs and Warner (1995, 1997) that reflects the fraction of years between 1960 and 1998 that a country is 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.

Without considering cognitive skills, protection against expropriation is significantly associated with economic growth across the OECD countries between 1960 and 2000 (column 1 of Table 3). However, when cognitive skills are included in the model, the coefficient on the institutional variable drops substantially in size and becomes statistically insignificant, whereas the coefficient on cognitive skills remains highly significant and close to our previous models without institutional controls (column 2).[3]

Table 3. Economic institutions versus educational outcomes in OECD-country long-run growth models

	OECD sample			Full country sample	
	(1)	(2)	(3)	(4)	(5)
Cognitive skills		1.527	1.388	1.265	1.266
		(3.39)	*(3.05)*	*(4.76)*	*(4.96)*
Initial years of schooling	0.135	0.058	0.067	0.067	0.047
	(2.11)	*(1.01)*	*(1.18)*	*(1.02)*	*(0.74)*
Initial GDP per capita	−0.381	−0.328	−0.322	−0.373	−0.304
	(7.83)	*(7.75)*	*(7.64)*	*(8.10)*	*(5.65)*
Protection against expropriation	0.729	0.224	0.081	0.332	0.492
	(3.95)	*(1.06)*	*(0.36)*	*(2.36)*	*(3.25)*
Openness			0.750	0.485	0.645
			(1.56)	*(1.34)*	*(1.83)*
OECD					2.589
					(1.26)
OECD × Protection against expropriation					−0.372
					(1.54)
No. of countries	24	24	23	48	48
F (Protection and Openness)			1.77	6.17	
F (OECD and interaction)					2.76
R^2 (adj.)	0.740	0.830	0.841	0.789	0.806

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960–2000. Regressions include a constant. *t*-statistics in parentheses. 'Protection against expropriation' is scaled from 0 to 10, with higher values corresponding to higher security of property rights. 'Openness' is scaled as a fraction of years that a country is classified as having an economy open to international trade.

Source: Authors' calculations based on data in Heston *et al.* (2002), Barro and Lee (2010), Hanushek and Woessmann (2009), Acemoglu *et al.* (2001), and an updated version of Sachs and Warner (1995).

[3] This result supports the view expressed in Glaeser *et al.* (2004) that human capital may be the more basic source of growth than institutions. Acemoglu *et al.* (2001) base their identification of institutional effects on the development of instruments derived from historical colonization patterns. An underlying argument is that colonists could bring with them a set of fundamental institutions that can be useful in explaining current institutions. Glaeser *et al.* (2004) argue that the colonists not only brought knowledge of institutions but also human capital – and that the human capital might be more fundamental to the growth process.

Conceptually, it is an open question whether any connection between human capital and institutions stems from human capital causing better institutions or the opposite. But at least within the group of OECD countries, that part of institutional variation that is not related to cognitive skills is not related to long-run growth, whereas that part of skill variation that is not related to institutions remains a strong predictor of long-run growth.

Column 3 adds the measure of openness to the model. Property-rights security and openness to trade are individually and jointly insignificant in predicting long-run OECD growth, whereas cognitive skills remain strongly significant.[4] Again, the insignificance in the institutional measures does not mean that institutions are unimportant for long-run growth. Rather, they point to the fact that the OECD countries share broadly similar institutions, so that this kind of institutional variation is unlikely to account for much of the substantial variation in long-run growth in this rich-country sample. Most OECD countries are coded as open throughout the period of observation. The exceptions are Mexico, New Zealand and Turkey that had substantial periods of being more closed, but the differences between the openness of these and the remainder of the OECD explain little of OECD growth differences.

The fundamental result is seen in the final two columns that are based on the full sample of 50 countries with available data. Property-rights and free-trade institutions help us understand long-run growth differences between rich and poor countries (column 4), but they do not contribute to our understanding of long-run growth differences within the group of rich countries (column 5). By contrast, the significant effect of cognitive skills on long-run growth in the OECD sample is robust to the inclusion of the institutional measures.

4.2. Regulation of product and labour markets

A substantial literature stresses the importance of how OECD countries regulate their product and labour markets. For example, Nicoletti and Scarpetta (2003) show that short-run growth experiences across OECD countries are related to product market regulations, and Cingano *et al.* (2010) find that employment protection legislation is associated with firm-level investment and other firm outcomes across sectors and across firms with different financial constraints.

But, do these factors enter into the long-run growth experiences of countries? Product market regulations affecting sectoral productivity may lead to structural change and international specialization, thereby reducing any net effects on aggregate growth rates. Similarly, differences in employment protection legislation may lead to differential growth experiences in booms versus recessions that cancel out over the business cycle.

[4] In the larger sample of worldwide experiences, there are indications that good institutions complement high cognitive skills, but identifying such interactions within the OECD sample is impossible (see Hanushek and Woessmann, 2008).

To investigate the long-run growth implications, we add a rich set of regulatory measures to our growth models. Specifically, we employ the latest versions of far-ranging indicators of regulations of both product and labour markets: product market regulations (PMR) from Wölfl *et al.* (2009), who update previous versions since Nicoletti *et al.* (2000); and employment protection legislation (EPL) from Venn (2009), who updates previous versions since OECD (1999).[5]

The results are unambiguous and telling: not a single of the large battery of measures that depict product and labour market regulations comes close to being significantly related to the variation in long-run growth experiences across OECD countries (Table 4). At the same time, regulatory practices do not affect the result that educational outcomes are a powerful predictor of long-run growth differences among OECD countries.

The first six columns of Table 4 employ aggregate and sub-indicators of product market regulation, and extensive robustness checks confirm the basic result (see details in Hanushek and Woessmann, 2010b). For example, the results hold for all underlying sub-indices, including indicators of public ownership, involvement in business operations, regulatory and administrative opacity, administrative burdens on start-ups, barriers to competition, and explicit and other barriers to trade and investment. The available indicators of product market regulation refer to 1998, the first year for which they are available. While earlier measures would be preferable, these should still capture the most basic overall patterns, under the assumption that main institutional variation is in the cross-section. Further, results are unaffected by using the available indicators for 2003 or 2008 instead, or by taking the average over the three observations, indicating that lack of results is not driven by simple measurement error. Finally, to align the regulatory measures more closely with the period of growth observations, we performed all regressions for growth between 1980 and 2000 without affecting the results.

The final five columns of Table 4 focus on labour market regulation. Columns 7 and 8 add the two versions of the aggregate employment protection index suggested by the OECD to the model. The first version combines regulations of regular employment contracts and of temporary contracts, and the second version adds sub-indices of additional regulation of collective dismissal to this. Neither measure enters the model significantly or affects the estimate on cognitive skills. The same is true when using separately the three sub-indicators of protection of permanent workers against (individual) dismissal, strictness of regulation on temporary forms of employment, and specific requirements for collective dismissal (columns 9–11).

The indicators of employment regulation are measured as averages of the annual values between 1985 and 2000. Results are similar when the growth period is restricted to 1980–2000, which aligns more closely to the period of observation of the

[5] For details on the measures of product market regulation and employment protection, see www.oecd.org/eco/pmr and www.oecd.org/employment/protection, respectively.

Table 4. Regulation of product and labour markets versus educational outcomes in OECD-country long-run growth models

	Product market regulation (PMR)						Employment protection regulation (EPR)				
	Overall PMR indicator	Administrative regulation	Domestic economic regulation	State control	Barriers to entrepreneurship	Barriers to trade and investment	EP version 1	EP version 2	Dismissal regular contracts	Regulation temporary contracts	Regulation collective dismissal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Cognitive skills	1.843	1.897	1.840	1.843	1.859	1.864	1.827	1.788	1.899	1.788	1.913
	(5.67)	*(5.92)*	*(5.60)*	*(5.66)*	*(5.75)*	*(5.67)*	*(5.41)*	*(5.06)*	*(5.39)*	*(5.00)*	*(5.95)*
Initial years of schooling	0.026	0.015	0.036	0.033	0.025	0.047	0.027	0.025	0.034	0.036	0.071
	(0.41)	*(0.23)*	*(0.60)*	*(0.55)*	*(0.40)*	*(0.75)*	*(0.41)*	*(0.37)*	*(0.48)*	*(0.58)*	*(1.21)*
Initial GDP per capita	−0.309	−0.310	−0.309	−0.311	−0.308	−0.303	−0.310	−0.307	−0.308	−0.307	−0.324
	(8.43)	*(8.69)*	*(8.27)*	*(8.30)*	*(8.52)*	*(8.36)*	*(8.09)*	*(8.25)*	*(7.89)*	*(8.22)*	*(8.52)*
Product market regulation	−0.157	−0.178	−0.080	−0.084	−0.157	0.004					
	(0.68)	*(1.06)*	*(0.55)*	*(0.67)*	*(0.76)*	*(0.02)*					
Employment protection legislation							−0.081	−0.109	−0.051	−0.047	0.139
							(0.62)	*(0.63)*	*(0.34)*	*(0.58)*	*(1.42)*
No. of countries	24	24	24	24	24	24	23	23	23	23	23
R^2 (adj.)	0.824	0.829	0.822	0.824	0.825	0.819	0.822	0.822	0.819	0.821	0.836

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960–2000. Sample: OECD countries. Regressions include a constant. *t*-statistics in parentheses. Each indicator of product market regulation and employment protection regulation is scaled from 0 to 6, with higher values reflecting more restrictions.

Source: Authors' calculations based on data in Heston *et al.* (2002), Hanushek and Woessmann (2009), Barro and Lee (2010), Venn (2009), and Wölfl *et al.* (2009).

regulatory measures (not shown). Results are also robust to the new, third version of the aggregate OECD employment protection index, available only in 2008, which adds the maximum time to make a claim of unfair dismissal, authorization and reporting requirements for temporary work agencies, and regulations requiring equal treatment of regular and agency workers at the user firm as three new sub-indicators.

The robustness checks (further detailed in Hanushek and Woessmann, 2010b) consistently document no convincing evidence that institutional or regulatory differences can account for differences in long-run growth among rich countries. Instead, cognitive skills emerge as the one strong policy factor underlying growth differences in the OECD.

5. DIFFERENT LEVELS OF SKILLS AND EDUCATION

An important and recurring policy question is which level of skills and education is most decisive for OECD growth. We analyze several dimensions of this: whether returns to average skills differ across countries; whether basic skills are more or less important than top skills; and whether there is a specific role of tertiary attainment.

5.1. Differential returns to average skills

First, the graphical plot of Figure 2 suggests that there is no obvious non-linearity in the test-score growth association across OECD countries. Thus, a squared test-score term does not enter the model significantly, and an exponential test-score specification does not improve the fit of the model (not shown).

Second, there is no obvious difference in the effect of average test scores between countries with initially low versus high income. While the small sample size does not allow for extensive interaction models, an interaction of test scores with an indicator for above-median initial GDP per capita does not enter the model significantly (not shown).

Third, from quantile regression estimates in 5% steps of the effect of average test scores for percentiles of the growth distribution, it is evident that the effect is relatively constant across the whole distribution of growth residuals (see details in Hanushek and Woessmann, 2010b). In fact, all quantile regression point estimates fall within standard confidence intervals around the OLS estimate.

5.2. Basic versus top skills

A leading policy question refers to effects of different ranges of the skill distribution. Should developed countries focus, in an egalitarian way, on decent basic skills for the whole population or, in an elitist way, on nurturing top scientists and engineers?

At a conceptual level, Vandenbussche *et al.* (2006) assume that the innovation process is more intensive in high-skill labour than the imitation process. They then present an endogenous growth model with innovation and imitation where high-skill labour has a greater growth-enhancing effect for countries closer to the technological frontier and countries further from the technological frontier get greater value from what they call 'unskilled human capital'. While the untested underlying assumption seems reasonable, there are also reasonable arguments to be made for an opposite view: The innovation literature suggests many innovations emerge from lucky coincidences, while almost by definition purposeful imitation processes require skilled scientists.

A different conceptual extension starts from the perspective of a high-skilled scientist. If this scientist were to work in a country that produces at the technological frontier, his only option would be to use his skill in the innovation of new technologies. If, by contrast, he were to work in a country that produces far below the technological frontier, he would still have the option to employ his skills in such innovative activities, but he would also have the additional option of employing his skills in imitating the more productive technologies at the technological frontier. This scientist will tend to work in the activity that promises the higher benefits, implying that the return to high-skill labour may well be larger when below the technological frontier than at it. While concentrations of high-skilled labour and spillovers across them may still be important, the alternative perspectives do introduce questions about the underlying assumptions.

We revisit these issues with two analyses. Here we define skills in terms of being at the top or bottom of the distribution of cognitive skills. In the next section, we follow Vandenbussche *et al.* (2006) and conceptualize the difference between high- and low-skill labour as school attainment at the tertiary versus non-tertiary level.

We start by replicating Hanushek and Woessmann (2009) who incorporate both the share of students who reach a basic skill level (a score of at least 400 on the tests) and the share of students who reach top-level skills (at least 600) in a growth regression of the full 50-country sample of OECD and non-OECD countries (column 1 of Table 5).[6] Both skill dimensions enter the model significantly, but the point estimate on the top-skill dimension is substantially higher. A 10 percentage point increase in the basic-skill share is associated with 0.3 percentage points higher annual growth; a 10 percentage point increase in the top-skill share is associated with 1.3 percentage points higher annual growth. Note that this does not necessarily provide an estimate of the relative importance of the two skill dimensions, as it may be much more feasible to increase the basic share than to increase the top share by the same amount. This might be suggested by the fact that the international standard deviation of the basic skill percentage is about four times as large as that of the top-skill level. It might also suggest that further improvements for countries already at the top of the distribution – say Finland, Korea and Japan – might

[6] These scores are one standard deviation below and above the OECD mean, respectively. The OECD also identifies five levels of skills on each of its tests (see, for example, OECD, 2007). The 400 falls in the range of level 1 while the 600 falls in the range of level 5.

Table 5. Basic versus high-level skills

Sample:	Full	OECD	Full	Full	OECD	Full	Full	OECD	Full
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Share of students reaching basic literacy	2.644	4.434	2.146	4.674	5.272	4.549			
	(3.51)	*(3.52)*	*(2.58)*	*(6.71)*	*(5.65)*	*(5.54)*			
Share of top-performing students	12.602	3.234	16.536				18.904	10.933	21.499
	(4.35)	*(0.99)*	*(4.90)*				*(7.45)*	*(3.61)*	*(7.19)*
Initial years of schooling	0.066	0.035	0.070	0.129	0.039	0.131	0.116	0.086	0.122
	(0.87)	*(0.60)*	*(0.94)*	*(1.49)*	*(0.68)*	*(1.47)*	*(1.40)*	*(1.23)*	*(1.53)*
Initial GDP per capita	−0.305	−0.305	−0.317	−0.326	−0.314	−0.340	−0.297	−0.275	−0.314
	(6.43)	*(8.24)*	*(5.63)*	*(5.85)*	*(8.71)*	*(5.05)*	*(5.62)*	*(6.09)*	*(5.32)*
OECD			−0.659			0.150			1.018
			(0.44)			*(0.10)*			*(2.17)*
OECD × basic literacy			2.074			0.024			
			(0.94)			*(0.01)*			
OECD × top-performing			−13.422						−10.733
			(2.08)						*(2.11)*
No. of countries	50	24	50	50	24	50	50	24	50
F (OECD and interaction)			1.62			0.10			2.66
R^2 (adj.)	0.724	0.822	0.734	0.616	0.822	0.600	0.655	0.720	0.679

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960–2000. Regressions include a constant. *t*-statistics in parentheses.

Source: Authors' calculations based on data in Heston *et al.* (2002), Barro and Lee (2010), and Hanushek and Woessmann (2009).

be more difficult than improvement in countries lower down in the distribution. Existing variations in advanced performance do nonetheless suggest the feasibility of improvement (Hanushek *et al.*, 2010).

When estimating the same model on the OECD sample, though, the point estimate on the top-skill share is only a fourth of the one estimated in the full-country sample and loses statistical significance (column 2). By contrast, the point estimate on the basic-skill share is slightly larger than in the full-country sample, and remains highly significant. The specification of column 3 shows that the difference in the estimate on the top-skill share between OECD and non-OECD countries is statistically significant. Of course, the measures of the two skill dimensions are highly collinear (their correlation is 0.73 in the full sample and 0.70 in the OECD sample), limiting precision in the joint specification. Results in columns 4–9, however, reveal that the pattern of results is similar when entering one of the two measures at a time.

5.3. Non-tertiary versus tertiary schooling

Vandenbussche *et al.* (2006) suggest that countries close to the technological frontier should emphasize tertiary education. To investigate this, we make use of the Barro and Lee (2010) database which provides average years of schooling separately at the primary, secondary and tertiary level. With little meaningful variation in the completion of primary education across the bulk of OECD countries, we combine the two basic levels of schooling into one category of non-tertiary schooling.

In the full-country sample, the coefficients on non-tertiary and tertiary schooling are both close to zero when cognitive skills are controlled for (column 1 of Table 6). By contrast, in the OECD sample, the point estimate on years of tertiary schooling becomes larger (column 2), and reaches marginal significance (at the 10% level) when years of non-tertiary schooling are not included in the model (column 3). However, results in column 4 indicate that this is completely driven by the United States. Once the United States is excluded, the coefficient on tertiary education is much smaller and again insignificant. The United States is well known for its extensive tertiary education system, and Figure 2 already indicated that the United States has the strongest positive residual in the growth model. While this might be an indication of growth-enhancing effects of its high-quality higher-education system, the lack of robustness in the sample without the United States suggests that it might rather be an indication of the high-skilled immigrant population that it attracts, of a particular set of economic institutions (not captured by our institutional measures), or of any other idiosyncrasy of the US economy. Moreover, once measures of cognitive skills are added (columns 5–7), tertiary education is insignificant even in the sample that includes the United States.

Table 6. Non-tertiary versus tertiary schooling

Sample:	Full	OECD	OECD	OECD w/o USA	OECD	OECD	OECD w/o USA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cognitive skills	1.923 *(9.12)*	1.888 *(6.09)*	1.912 *(6.83)*	2.043 *(8.14)*			
Share of students reaching basic literacy					5.458 *(7.13)*		
Share of top-performing students						11.597 *(3.93)*	12.855 *(4.44)*
Years of non-tertiary schooling	0.076 *(0.94)*	0.012 *(0.059)*					
Years of tertiary schooling	0.198 *(0.16)*	1.291 *(1.58)*	1.344 *(1.77)*	0.543 *(0.74)*	1.685 *(2.31)*	1.014 *(0.96)*	0.149 *(0.13)*
Initial GDP per capita	−0.325 *(6.81)*	−0.323 *(8.85)*	−0.320 *(9.78)*	−0.323 *(11.22)*	−0.344 *(10.53)*	−0.264 *(6.08)*	−0.263 *(6.37)*
No. of countries	50	24	24	23	24	24	23
R^2 (adj.)	0.728	0.839	0.847	0.886	0.856	0.712	0.750

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960–2000. Regressions include a constant. *t*-statistics in parentheses.

Source: Authors' calculations based on data in Heston *et al.* (2002), Barro and Lee (2010), and Hanushek and Woessmann (2009).

While distinguishing the effects of different dimensions of the skill and schooling distribution is difficult in these small samples, some basic patterns prove clear. First, the significant effect of cognitive skills is extremely robust to consideration of any quantitative measure of different levels of school attainment. Second, the finding of Vandenbussche *et al.* (2006) of a particular effect of tertiary attainment in rich countries is not robust once the focus is on long-run growth experiences and educational outcome measures are taken into account. Of course, this does not mean that learning beyond the secondary level does not matter. Rather, in the spirit of a life-cycle interpretation where early skills facilitate the development of subsequent skills (Cunha *et al.*, 2006), it means that outcome measures of learning in school are a good predictor for the accumulation of further skills in life and for the capacity to deploy these skills effectively.

6. CALCULATING THE ECONOMIC VALUE OF EDUCATION REFORMS

The results indicate that educational outcomes have strong effects on long-run economic growth of OECD countries but do not tell us directly the economic value of any improvements in educational outcomes. In particular, the growth rate effects do not map linearly into the economic value of any education reform in a country, not least because different time lags are involved between successful reform in the education system today and the improvement of skills in the national workforce. Here, we perform simulation analyses that use the previous estimates to project the economic impact of different scenarios of school improvements in the OECD member states.

We consider three specific reform scenarios that reflect plausible policy goals. First, improving average student performance by 1/4 standard deviation, or 25 PISA points, in each country; second, bringing all OECD countries up to the level of the PISA top performer, Finland; and third, bringing all students in OECD countries to minimum proficiency, defined as 400 points on the PISA test scale.

6.1. Issues of causation

Of course, considerable controversy surrounds cross-country growth regressions and any causal interpretation of them (see, for example, Levine and Renelt, 1992; Bils and Klenow, 2000). An obvious concern is that countries that grow faster have the resources to invest in schools so that growth could cause higher scores. We implicitly address this below when we find a lack of relationship across countries in the amount spent on schools and the observed test scores.

In other work, we have considered a series of analyses aimed at eliminating many of the other natural concerns about the identification of the causal impacts of cognitive skills (Hanushek and Woessmann, 2009).[7] Each of the analyses points to the plausibility of a causal interpretation of the basic models. In addition, in specifications with country and industry fixed effects, Ciccone and Papaioannou (2009) find that countries with a more skilled labour force (using the Hanushek and Kimko, 2000, test measures) experienced faster growth in skill-intensive industries during the 1980s and 1990s. This cross-sector analysis reinforces the other causality discussions as it tends to rule out bias from omitted cultural effects which should be uniform across sectors. Nonetheless, with the limited international variations, it is difficult to demonstrate identification conclusively. Therefore, below we consider a variety of sensitivity analyses designed to show the impact on outcomes when part of the estimated effects is non-causal.

One final issue related to causation is important. Because we consider economic outcomes far into the future, the precise form of the underlying growth model potentially makes a noticeable difference. We begin with what is commonly referred to as an 'endogenous growth' model where improved skills yield a permanently higher growth rate. We subsequently consider a 'neoclassical growth' model where skills affect the level of income but effects on growth are transitory and the economy returns to its previous long-run growth rate. Our previously estimated models actually support analysis of this range of underlying models, and applying the alternatives gives another idea of the bounds on future economic effects from school improvement.

6.2. The projection model

The economic projections involve several components. First, we calculate the time path of the annual growth rate engendered by education reform designed to move students from their current performance to a given new level. This pattern of economic outcomes represents the confluence of three separate dynamic processes: (1) Changes in schools lead to the progressive improvement in student achievement until students fully reach the new steady-state level of achievement; (2) students with better skills move into the labour force and the average skills of workers increase as new, higher achieving workers replace retiring workers; and (3) the economy responds to the progressive improvement of the average skill level of the workforce. Second, based on the pattern of predicted growth rates, we model the future expansion of

[7] To rule out simple reverse causation, Hanushek and Woessmann (2009) separate the timing of the analysis by estimating the effect of scores on tests conducted until the early 1980s on economic growth in 1980–2000, finding an even larger effect. Three further direct tests of causality devised to rule out certain alternative explanations based on unobserved country-specific cultures and institutions confirm the results. The first one considers the earnings of immigrants to the United States and finds that the international test scores for their home country significantly explain US earnings but only for those educated in their home country and not for those educated in the United States. A second analysis takes out level considerations and shows that changes in test scores over time are systematically related to changes in growth rates over time. A third causality analysis uses institutional features of school systems as instrumental variables for test performance, thereby employing only that part of the variation in test outcomes emanating from such country differences as use of central exams, decentralized decision making, and the share of privately operated schools. These results support a causal interpretation and also suggest that schooling can be a policy instrument contributing to economic outcomes.

GDP with and without the education reform. Third, based on these projections, we calculate the total value of the reform by aggregating the discounted values of the annual differences between the GDP with reform and the GDP without reform.

6.2.1. Framework for baseline projections. Implementing the projection model requires a number of parameter assumptions and simplifications, most of which are subsequently subjected to a sensitivity analysis.

The simulation does not adopt any specific reform package but instead focuses just on the ultimate change in achievement. For the purposes here, reforms are assumed to take 20 years to complete, and the path of increased achievement during the reform period is taken as linear. For example, an average improvement of 25 points on PISA is assumed to reflect a gain in the student population of 1.25 points per year. This might be realistic, for example, when the reform relies upon a process of upgrading the skills of teachers – either by training for existing teachers or by changing the workforce through replacement of existing teachers. This linear path dictates the quality of new cohorts of workers at each point in time.

The expected work life is assumed to be 40 years, which implies that each new cohort of workers is 2.5% of the workforce. Thus, even after an educational reform is fully implemented, it takes 40 years until the full labour force is at the new skill level.

The benchmark here considers all economic returns that arise during the lifetime of a child that is born at the beginning of the reform in 2010. According to the most recent data (that refer to 2006), a simple average of male and female life expectancy at birth over all OECD countries is 79 years (OECD, 2009b).[8] Therefore, the baseline calculations take a time horizon until 2090, considering all future returns that accrue until then, but neglecting any returns that accrue after 2090.

The simulations rely on the estimates of growth relationships derived from the 24 OECD countries with complete data. As indicated in column 2 of Table 2, the coefficient estimate is 1.864, suggesting, for example, that a 50 point higher average PISA score (i.e., one-half standard deviation higher) would be associated with 0.93% higher annual growth in the long run in the endogenous growth projections.

The value of improvement in economic outcomes from added growth also depends, of course, on the path of economies that would be obtained without educational improvement. The baseline analysis here takes the annual growth of OECD economies in the absence of education reform to be 1.5%. This is simply the average annual growth rate of potential GDP per worker of the OECD area over the past two decades: 1.5% in 1987–96 and 1.4% in 1997–2006 (OECD, 2009a).

Finally, more immediate benefits are both more valuable and more certain than those far in the future. In order to incorporate this, the entire stream is converted

[8] Note that these life expectancy numbers are based on age-specific mortality rates prevalent in 2006, and as such do not include the effect of any future decline in age-specific mortality rates. Life expectancy at birth has increased by an average of more than 10 years since 1960.

into a present discounted value. In simplest terms, the present discounted value is the current dollar amount that would be equivalent to the future stream of returns calculated from the growth model. Specifically, if that amount of funds was invested today, it would be possible to reproduce the projected stream of economic growth benefits from the principal amount and associated investment returns. Thus, this calculation of present discounted value allows a relevant comparison for any other current policy actions.

In doing so, the discount rate at which to adjust future benefits becomes an important parameter. A standard value of the social discount rate used in long-term projections on the sustainability of pension systems and public finance is 3% (e.g., Börsch-Supan, 2000; Hagist *et al.*, 2005), a precedent that is followed here.[9] By contrast, the influential Stern Review report that estimates the cost of climate change uses a discount rate of only 1.4%, thereby giving a much higher value to future costs and benefits (Stern, 2007). In our robustness analyses, we will also consider such alternative discount rates.

A number of untested assumptions go into our projections. First, they assume that skills play the same role in the future as they have in the past, so that the evidence of past results provides a direct way to project the future. Second, while the statistical analysis did not look at how economies adjust to improved skills, the calculations assume that the experience of other countries with greater cognitive skills provide the relevant insight into how the new skills will be absorbed into the economy. Third, the projection of simultaneous improvement across countries presumes that all countries can grow faster without detracting from (or benefiting) growth in other countries. In other words, the higher levels of human capital in each country allow it to innovate, to improve its production, and to import new technologies without detracting from the growth prospects for other countries.[10] Further, the estimates ignore any other aspects of interactions such as migration of skilled labour across borders. (Of course, one way that a country could improve its human capital would be by arranging for its youth to obtain schooling in another country with better schools – as long as the more educated youth return to their home country to work). Fourth, all countries are assumed to have a stationary population with a constant age distribution. Fifth, the projections are the gross benefits of reform, and they equal net benefits only if we assume reform is costless – an assumption

[9] As a practical value for the social discount rate in cost-benefit analysis (derived from an optimal growth rate model), Moore *et al.* (2004) suggest using a time-declining scale of discount rates for intergenerational projects that do not crowd out private investment, starting with 3.5% for years 0–50, 2.5% for years 50–100, 1.5% for years 100–200, 0.5% for years 200–300, and 0% years over 300. (The proper starting value is actually 3.3% based on the parameter values they assume for the growth rate in per capita consumption (2.3%), the social marginal utility of consumption with respect to per capita consumption (1), and the utility discount rate (1%)).

[10] Rather than being negative, spillovers of one country's human capital investments on other countries could also be positive. For example, if one country pushes out the world technological frontier by improving its human capital, other countries can gain from this by imitation and reach a higher productivity level. No attempt is made to consider how technological change occurs and the impact on wages and earnings. Obviously, different patterns of productivity improvements will play out differently in the labour market as seen in the United States over time (Goldin and Katz, 2008).

discussed explicitly in Section 7.4 below. Finally, all calculations are in real (inflation-adjusted) terms – 2010 dollars under purchasing power parity.

6.2.2. The phases of reform. The economic impact of the reform varies across four phases that are defined by the average quality of the labour force.

(a) *Phase 1 (2010–2030):* During the 20 years of the education reform programme, the additional growth in GDP per capita due to the reform in year t is given by:

$$\Delta^t = growth\ coefficient * \Delta PISA * \frac{1}{working\ life} * \frac{t-2010}{20} + \Delta^{t-1} \tag{3}$$

where the *growth coefficient* comes from the regression estimations presented in the previous sections and $\Delta PISA$ is the increase in the average PISA test score due to the reform. The working life term indicates that each cohort of new, higher achieving students is only a fraction of the total labour force.

(b) *Phase 2 (2031–2050):* The education reform is now fully enacted, and achievement of all subsequent students remains at the new level. But for the length of a work life from the start of reform, which in the baseline simulations is assumed to last 40 years, there are still workers with initial levels of skills that are being replaced in retirement by higher achieving workers. During this phase, the additional growth in GDP per capita in year t due to the reform is given by:

$$\Delta^t = growth\ coefficient * \Delta PISA * \frac{1}{working\ life} + \Delta^{t-1} \tag{4}$$

(c) *Phase 3 (2051–2070):* During this phase, the first 20 labour-market cohorts – which only partially profited from the education reform – are replaced by cohorts that profited from the fully enacted education reform:

$$\Delta^t = growth\ coefficient * \Delta PISA * \frac{1}{working\ life} - \left(\Delta^{t-40} - \Delta^{t-41}\right) + \Delta^{t-1} \tag{5}$$

(d) *Phase 4 (after 2070):* Finally, the whole workforce has gone through the reformed education system. The annual growth rate is now increased by the constant long-run growth effect Δ:

$$\Delta = growth\ coefficient * \Delta PISA \tag{6}$$

6.2.3. The level of GDP with and without reform. *Without reform*, the economy grows at the constant growth rate of potential GDP:

$$GDP^t_{no\ reform} = GDP^{t-1}_{no\ reform} * (1 + potential\ growth) \tag{7}$$

With reform, the annual growth rate is additionally increased by the growth effect Δ^t:

$$GDP^t_{reform} = GDP^{t-1}_{reform} * (1 + potential\ growth + \Delta^t) \quad (8)$$

In the neoclassical specification, an additional term ensures that the growth rate is negatively affected by the (log) level of GDP reached in the previous period. As a consequence, the annual growth rate without and with reform will converge to the same rate of potential growth in the long run.

6.2.4. Cumulative effects of the reform. The total value of any reform is given by the sum of the discounted values of the annual differences between the GDP with reform and the GDP without reform:

$$Total\ value\ of\ the\ reform = \sum_{t=2010}^{t=2090} \left(GDP^t_{reform} - GDP^t_{no\ reform} \right) * (1 + discount\ rate)^{-(t-2010)} \quad (9)$$

In the baseline scenario, the eighty year time horizon over which future returns will be considered is the lifetime of a child born at the beginning of the reform, which takes the projections to the year 2090.

6.3. Baseline reform projections

We start with the results of the baseline projection model in an endogenous-growth framework for the three education reform scenarios.

6.3.1. Scenario I: Increase average performance by 25 PISA points. A simple starting point is to consider the economic impact on OECD countries of a 0.25 standard deviation improvement, equivalent to a 25 point increase on PISA scores. The reform policy is begun in 2010 and on average yields 25 point higher scores in 2030 that remain permanently at that level for all subsequent students.[11] The goal of having all OECD countries boost their average PISA scores by 25 points over the next 20 years is less than the most rapidly improving education system in the OECD, Poland, achieved between 2000 and 2006 alone (Organization for Eco-

[11] All calculations of PISA scores underlying the following simulations refer to the average performance in maths and science (in line with the underlying growth model), averaged over the three PISA cycles 2000, 2003, and 2006 (see, e.g., OECD, 2007). All underlying measures of gross domestic product (GDP) are in US dollars, measured in purchasing power parities (PPP), expressed in prices of 2010. The GDP measures were calculated from the most recent measure of GDP in current prices and current PPPs available for all countries (2007, extracted from http://stats.oecd.org on 10 August 2009), projected to 2010 using OECD estimates of annual changes in potential GDP and in GDP deflators (Organization for Economic Co-operation and Development, 2009a).

nomic Co-operation and Development, 2007). Such achievement change was also observed in three German states – Saxony, Brandenburg and Saxony-Anhalt – over the same period (PISA-Konsortium Deutschland, 2008). And, over a longer period more like the 20 years of the scenario, both Finland and Canada appear to have showed score growth of this magnitude (Hanushek and Woessmann, 2009). Thus, there is clear support that improvement is possible (see also Mourshed *et al*., 2010).

A policy like this is uniform across countries, so the relative improvement is the same for all countries.[12] Figure 3 provides a summary of the marginal impact on GDP for each year into the future. While there are no impacts initially until higher-achieving students start becoming more significant in the labour market, GDP will be more than 3% higher than what would be expected without improvements in human capital as early as 2041. (The figure also shows a 95% confidence

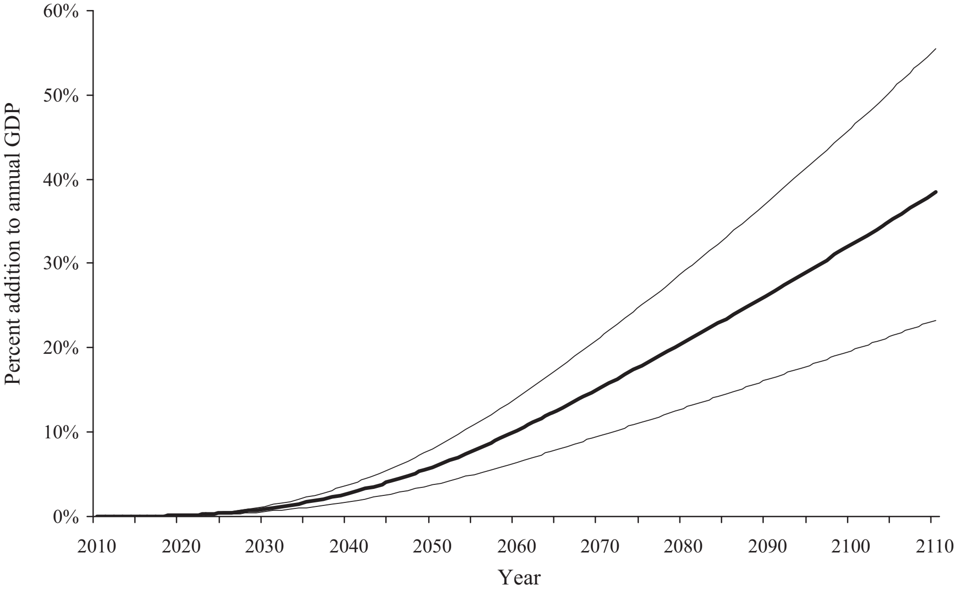

Figure 3. Improvement in annual GDP with Scenario I (increase everyone by ¼ std. dev.)

Notes: GDP with reform relative to GDP without reform in each year after the reform starts. Main line: point estimate of Scenario I. Thin lines: 95% confidence interval of the point estimate of the growth regression.

Source: Authors' calculations.

[12] Note that the calculations also assume that the top ranked countries can feasibly improve their scores. The relatively flat performance of countries such as Japan and Korea that have been at the top for a number of years raises the question about whether there is room for further improvement or whether there is some sort of ceiling effect in the existing tests. As an alternative, the next scenario will only assume improvements that do not go beyond the current top performer.

bound of 1.9–4.1% higher GDP, based on the relevant 95% confidence bounds for the regression coefficient in column 2 of Table 2). By the end of expected life in 2090 for the person born in 2010, GDP per capita would be expected to be over 26% above the 'education as usual' level.

The magnitude of such a change is best understood with an example. In the absence of changes in educational policy, France would be expected to have a GDP (in 2010 US dollars) of $3,606 billion in 2041. If on the other hand it achieved the improvement in cognitive skills that took it from an average PISA score of 505 to 530, total GDP would be expected to be $3,715 in 2041, or $108 billion higher. These calculations illustrate a simple point: While 3% may at first seem like a small change, it is a very large number when applied to the entire GDP of any of the OECD countries.

These calculations are by themselves misleading, because the impacts of improved cognitive skills continue to occur far into the future. The 3.0% improvement in 2041 rises to a 5.9% improvement in 2050, 15.3% in 2070, and 26.3% in 2090. These dynamic improvements in the economy yield ongoing gains to society, and the appropriate summary of the impact of educational improvements accumulates the value of these annual gains.

Importantly, after all people in the labour force have obtained the new and improved education (in 2070), annual growth will be 0.47 percentage points higher. This implies that each country that achieves the average improvement of ¼ standard deviation of achievement will have a cumulative impact on the economy through 2090 that is equal to 288% of current year GDP. The first column of Table 7 provides these discounted values of all of the future increases through 2090 for each OECD country. The dollar value for each country varies by the level of GDP in 2010 – but the total impact across the OECD is $123 trillion in present value.

Because these are put into present value terms, they can be compared to current economic values. For example, these calculations indicate that the value of improvements through long-run growth far outstrips the cost of the recent worldwide recession and are an order of magnitude larger than the worldwide fiscal stimulus efforts.

6.3.2. Scenario II: Bring each country to Finland average level (of 546 PISA points). The success of Finland on the PISA tests is well known. In the second scenario, the performance of Finnish students is taken as a benchmark for the performance levels that are possible. The economic impact calculated is found from projecting the impact on growth for each OECD country under the assumption that it could bring itself in twenty years to the top of the rankings as identified by Finland – an average PISA score of 546.

The impact on different economies varies by the size of reform (in addition to the size of the economy itself). Finland, for example, under this scenario would

Table 7. Baseline projections of the economic value of three education reform scenarios

	Scenario I: Increase avg. performance by ¼ std. dev.	Scenario II: Bring each country to Finnish level of 546 points on PISA				Scenario III: Bring all to minimum of 400 points on PISA			
	Value of reform ($ billion)	Value of reform ($ billion)	In % of current GDP	Long-run growth increase (p.p.)	Note: Increase in PISA score	Value of reform ($ billion)	In % of current GDP	Long-run growth increase (p.p.)	Note: Share of students below minimum skills (in %)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Australia	2,631	2,092	229	0.38	20.1	2,430	266	0.43	9.8
Austria	969	1,545	460	0.72	38.4	1,308	390	0.62	13.9
Belgium	1,208	1,586	379	0.60	32.2	1,816	434	0.68	15.3
Canada	4,051	2,728	194	0.32	17.2	3,075	219	0.36	8.1
Czech Republic	830	1,177	409	0.64	34.5	1,054	366	0.58	13.1
Denmark	608	1,231	584	0.88	47.5	908	430	0.67	15.2
Finland	594	0	0	0.00	0.0	255	124	0.21	4.7
France	6,557	11,349	499	0.77	41.3	9,844	433	0.68	15.3
Germany	8,822	17,245	564	0.86	46.0	15,166	496	0.77	17.3
Greece	1,047	4,253	1172	1.59	85.2	2,943	811	1.17	26.5
Hungary	603	1,323	633	0.95	51.0	972	465	0.72	16.3
Iceland	36	66	530	0.81	43.6	46	371	0.59	13.3
Ireland	585	995	490	0.76	40.6	664	327	0.52	11.8
Italy	5,526	19,353	1010	1.41	75.6	13,503	705	1.04	23.5
Japan	13,280	2,871	62	0.11	5.7	10,382	226	0.37	8.3
Korea, Rep.	4,120	756	53	0.09	4.8	2,544	178	0.30	6.7
Luxembourg	126	421	963	1.36	72.7	289	662	0.99	22.3
Mexico	4,753	39,363	2389	2.68	143.9	29,557	1794	2.19	49.5
Netherlands	2,032	1,344	191	0.31	16.9	1,779	253	0.41	9.3
New Zealand	361	275	220	0.36	19.4	385	308	0.49	11.2
Norway	844	1,975	675	1.00	53.9	1,391	476	0.74	16.6
Poland	2,119	5,320	724	1.07	57.2	3,766	513	0.79	17.8
Portugal	742	2,860	1112	1.52	81.7	1,878	730	1.07	24.2
Slovak Republic	343	787	661	0.99	52.9	549	461	0.72	16.2
Spain	4,496	12,332	791	1.15	61.7	8,237	529	0.81	18.3

Table 7. (Continued)

	Scenario I: Increase avg. performance by ¼ std. dev.	Scenario II: Bring each country to Finnish level of 546 points on PISA				Scenario III: Bring all to minimum of 400 points on PISA			
	Value of reform ($ billion)	Value of reform ($ billion)	In % of current GDP	Long-run growth increase (p.p.)	Note: Increase in PISA score	Value of reform ($ billion)	In % of current GDP	Long-run growth increase (p.p.)	Note: Share of students below minimum skills (in %)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sweden	1,080	1,761	470	0.73	39.2	1,406	375	0.59	13.4
Switzerland	1,003	1,159	333	0.53	28.6	1,263	363	0.58	13.0
Turkey	3,043	19,450	1844	2.24	120.1	15,089	1430	1.85	41.8
United Kingdom	6,862	7,892	332	0.53	28.5	7,669	322	0.52	11.7
United States	43,835	111,923	737	1.08	58.1	86,167	567	0.86	19.4
OECD	123,108	275,429	645	0.93	49.8	226,333	530	0.80	18.0

Notes: Discounted value of future increases in GDP until 2090, expressed in billion dollars (PPP) and as a percentage of current GDP. 'Long-run growth increase' refers to increase in annual growth rate (in percentage points) once the whole labour force has reached higher level of educational performance. 'Increase in PISA score' refers to the ultimate increase in educational performance due to reform scenario II. 'Share of students below minimum skills' refers to the share of students in each country performing below the minimum skill level of 400 PISA points. See text for reform parameters.

Source: Authors' calculations.

neither change its schools nor see any long-term economic changes. On the other side, Mexico and Turkey would require enormous changes in their educational achievement, and, if the changes were feasible, would see their economies completely transformed. As discussed above, there is experience with school improvements of 25 or more points (including Finland itself), but there are no cases matching these extremes.

Columns 2–5 of Table 7 present the country-by-country impacts of these changes. On average, the OECD countries would see a nearly 50 point increase in performance (one-half standard deviation). While the change in Japan or Korea amounts to about 5 points, the change in Mexico is 144 points – an almost inconceivable change given current knowledge of how to transform schools or cognitive skills in general. (Again, the calculations assume that adjustment is complete within 20 years. An alternative view would be that a number of countries would actually require longer for a reform programme to yield such large changes. We model the implications in the robustness analyses below.)

The present value for OECD improvements under this scenario is $275 trillion, or more than six times the current GDP of the OECD countries. The United States itself, which currently falls over 50 points behind Finland, would by historical growth patterns see a present value of improved GDP of over $112 trillion, or some 40% of the OECD total – reflecting both the size of the country and its distance behind Finland. Germany would see a $17 trillion improvement, or more than five times current GDP.

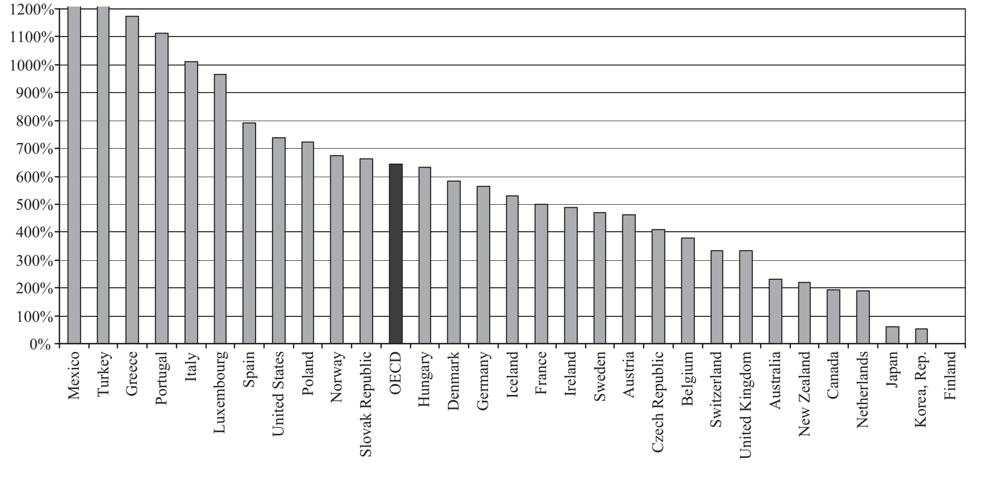

Figure 4. Present value of Scenario II (bring each country to Finnish level) in percent of current GDP

Notes: Discounted value of future increases in GDP until 2090 due to a reform that brings each country to the Finnish average level of 546 PISA points, expressed as a percentage of current GDP. Value is 2389% for Mexico and 1844% for Turkey.

Source: Authors' depiction based on the projection analysis reported in column 3 of Table 7.

The rankings of countries according to increases compared to current GDP are shown in Figure 4. One interpretation of this figure is the amount of economic leverage from educational improvements that is possible for different OECD countries.

6.3.3. Scenario III: Bring everyone up to minimum skill level of 400 PISA points. The final scenario considered is a 'compensatory' improvement in education where all students are brought up to a minimal skill level – which is defined here as obtaining a score of 400 on the PISA tests, or one standard deviation below the OECD average. While the previous simulations could be thought of as displaying the results of shifting the entire achievement distribution, this scenario considers the implications of bringing up the bottom of the distribution.

In order to understand the implications of changing just one portion of the achievement distribution, we employ the alternative estimation of the underlying growth models of column 2 of Table 5. Specifically, instead of relying on just average cognitive skills in the growth models, the proportion of the population with scores less than 400 and the proportion with scores over 600 are included in the growth models.

For these calculations, all OECD countries including Finland have room for improvement. On average, 18% of students in the OECD countries score below 400. As might be expected from the average scores, the required improvements are largest in Mexico and Turkey (see column 9 of Table 7).

Columns 6–8 of Table 7 display the economic outcomes according to historical growth patterns of bringing all OECD students up to minimum competence levels. The overall OECD change would be an average annual growth rate that was 0.8% higher after reform was accomplished and after the full labour force had received the improved education. The total improvements for the OECD countries from achieving universal minimum proficiency would have a present value of $226 trillion. Again, there is a wide range of outcomes including relatively small improvements of 219% of current GDP for Canada as compared to nine OECD countries that would experience a benefit more than five times their current GDP.

The range of outcomes is depicted in Figure 5, which ranks countries by the benefits compared to current GDP. Even Finland could by these calculations more than double its current GDP by bringing the relatively modest proportion of low performers (4.7%) up to scores of 400. Note also that the effects of these policies on the separate countries differ from the previous scenario, reflecting the differences in the underlying distribution of student performance.

6.4. An alternative neoclassical growth framework

The projections so far assume that higher educational achievement allows a country to keep on growing at a higher rate in the long run. Such a specification captures the basic ideas of endogenous growth theory, where a better-educated workforce

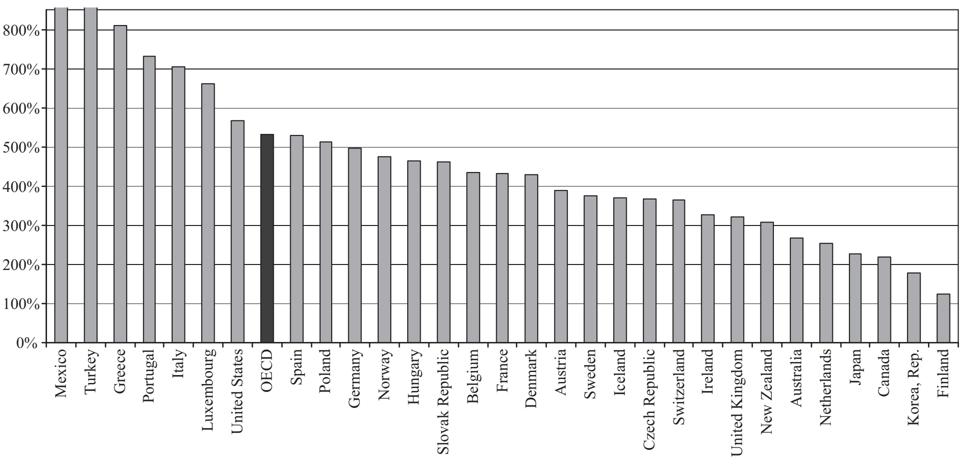

Figure 5. Present value of Scenario III (bring everyone up to minimum skills) in percentage of current GDP

Notes: Discounted value of future increases in GDP until 2090 due to a reform that brings every individual in the country up to a minimum skill level of 400 PISA points, expressed as a percentage of current GDP. Value is 1794% for Mexico and 1430% for Turkey.

Source: Authors' depiction based on the projection analysis reported in column 7 of Table 7.

leads to a larger stream of new ideas that produces technological progress at a higher rate. By contrast, in the augmented neoclassical growth model, changes in test scores lead to higher steady-state levels of income but do not affect the long-run growth path. Our empirical growth model captures the conditional convergence implied by the neoclassical model – but also by a set of endogenous growth models – through including the initial GDP level as a control variable. An alternative approach for the projections is thus to interpret the model in the neoclassical rather than endogenous-growth framework and have educational reforms affect the steady-state level of income but not its long-run growth.

To do so, we re-estimate our growth model with the logarithmic (rather than linear) per-capita GDP as control. The test-score coefficient hardly changes in this specification (1.718 rather than 1.864), and the coefficient on log initial income is –1.835. This matches standard parameter assumptions in the augmented neoclassical growth model (Mankiw *et al.*, 1992), implying an economy moves halfway to its steady state in about 38 years. With convergence, projections of growth rates with and without education reform will differ only during the transition to the new balanced growth path.

To characterize growth of the world technological frontier, we assume that a GDP-weighted average growth in the three countries with the largest shares of patents in the world – the United States, Japan and Germany – is 1.5% per year without reform. Together, the three countries currently account for over 70% of worldwide patents (measured in triadic patent families, OECD, 2008).

Table 8 shows the results of the projections based on the neoclassical model specification. In reform Scenario I, where each country increases by 25 PISA points, the value of the reform – the discounted value of the future increases in GDP – amounts to $90 trillion in present value terms compared to our previous projection of $123 trillion. In the neoclassical projections, the value of this 'uniform' reform differs across countries, because the projected growth rates vary with the level of GDP, but the overall OECD effect still equals 211% of current OECD GDP.

To illustrate the specific dynamics of the neoclassical projections, it is worth discussing a few details of the trajectories of the projection. Initial growth rates (with or without reform) in 2010–11 differ across countries, depending on how far away they are from their steady state. The United States is projected to grow at 1.1% initially, whereas the simple mean of OECD-country growth rates is higher at 1.9% due to many countries' room for catch-up.[13] Due to the convergence process, average OECD-country growth is down to 1.6% in 2090 without the reform, ranging from 1.0–1.8% across countries. With reform, the range is 1.2–2.0%, with an average of 1.7%. By 2104, the average is down to 1.5%, and by 2130, all countries' growth rates have converged to between 1.3% and 1.6% without the reform and to 1.4–1.7% with reform.

In the neoclassical projections, the difference in average growth rates between the scenarios with and without reform grows to a maximum of 0.28 percentage points in about 2060 and then declines back to 0.18 percentage points in 2090 and 0.06 percentage points in 2150. (Compare this to the endogenous growth model where the long-run growth rate stays 0.47 percentage points higher starting in 2070.) While the difference ultimately converges to zero (everywhere below 0.004 percentage points by 2300), the model underscores that this convergence process takes very long to take full effect.

In reform Scenario II, where each country improves to the test-score level of Finland, the present value of the reform amounts to $180 trillion in the neoclassical model, compared to the previous $275 trillion in the endogenous growth model. Note that this scenario implies that in the *very* long run, each country converges to the same steady-state level of per-capita GDP, as test scores are the only variable influencing the steady-state level in our model. However, in 2090 the per-capita GDP of the most advanced country would still be 70.6% higher than that of the least advanced country. By 2150, this difference would be down to 19.7%, and by 2300 to 1.2%.

[13] Based on our model which depicts only effects of test scores and evolving levels of income, Luxembourg and Norway – the two countries with the highest current levels of GDP per capita – are projected to converge to a lower balanced growth path (both without and with reform) and thus initially have the lowest growth rates. If these countries can keep their current advantage in per capita GDP relative to the other OECD countries in the future for reasons outside our model, this would increase the projected value of the educational reform in these countries.

Table 8. Projection results with 'neoclassical' model specification

	Scenario I: Increase avg. performance by 1/4 std. dev.		Scenario II: Bring each country to Finnish level of 546 points on PISA		Scenario III: Bring all to minimum of 400 points on PISA		Scenario II, Time horizon 2050		Scenario II, Time horizon 2150	
	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Australia	2,073	227	1,656	182	3,826	419	276	30	3,696	405
Austria	712	212	1,120	334	1,471	438	186	55	2,511	748
Belgium	926	221	1,207	288	2,334	557	200	48	2,711	647
Canada	3,282	234	2,227	159	5,209	371	368	26	4,984	355
Czech Republic	781	271	1,095	381	1,631	567	169	59	2,570	893
Denmark	435	206	859	407	869	412	143	68	1,932	916
Finland	560	272	0	0	651	316	0	0	0	0
France	5,026	221	8,552	376	11,090	488	1,397	61	19,389	853
Germany	6,521	213	12,466	408	15,347	502	2,049	67	28,191	922
Greece	671	185	2,548	702	1,459	402	417	115	5,837	1609
Hungary	618	296	1,322	632	1,302	623	193	93	3,213	1538
Iceland	25	204	46	367	46	371	8	61	102	819
Ireland	386	190	645	318	643	317	111	55	1,417	698
Italy	3,725	194	12,330	644	8,102	423	2,011	105	28,244	1474
Japan	12,584	273	2,772	60	24,595	534	442	10	6,329	137
Korea, Rep.	4,489	314	839	59	7,094	497	128	9	1,969	138
Luxembourg	46	105	144	330	93	214	29	66	291	665
Mexico	3,451	209	24,773	1504	7,160	435	3,583	217	62,461	3791
Netherlands	1,623	230	1,082	154	2,979	423	180	26	2,413	342
New Zealand	360	288	276	221	790	632	42	34	646	516
Norway	435	149	985	337	855	292	181	62	2,088	714
Poland	2,192	298	5,322	725	4,651	633	770	105	13,051	1777
Portugal	579	225	2,099	816	1,209	470	323	126	4,989	1940
Slovak Republic	337	283	749	630	682	573	111	93	1,809	1519

Table 8. (Continued)

	Scenario I: Increase avg. performance by ¼ std. dev.		Scenario II: Bring each country to Finnish level of 546 points on PISA		Scenario III: Bring all to minimum of 400 points on PISA		Scenario II, Time horizon 2050		Scenario II, Time horizon 2150	
	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Spain	3,142	202	8,281	531	6,275	403	1,359	87	18,825	1208
Sweden	784	209	1,259	336	1,542	412	210	56	2,818	753
Switzerland	722	208	831	239	1,598	460	141	41	1,837	528
Turkey	2,699	256	15,474	1467	6,363	603	2,162	205	39,523	3747
United Kingdom	5,504	231	6,308	265	10,918	459	1,032	43	14,243	599
United States	25,344	167	62,386	411	56,407	371	10,962	72	135,962	895
OECD	90,031	211	179,655	421	187,191	439	29,183	68	414,050	970

Notes: Discounted value of future increases in GDP until 2090, expressed in billion dollars (PPP) and as a percentage of current GDP. See text for reform parameters.

Source: Authors' calculations.

The present value of reform Scenario III, which brings all students to a minimum level of 400 PISA points, is \$187 trillion in the neoclassical projections, rather than the \$226 trillion of the endogenous-growth type projections.[14]

Longer time horizons imply larger differences between the two growth models. The final four columns of Table 8 report the present value calculated with the estimated neoclassical version for reform Scenario II with time horizons varying between 2050 and 2150, respectively. Compared to the endogenous-growth projections evaluated over the same time horizons (discussed in Table 9 below), the neoclassical projection is 81% of the endogenous-growth projection value for a time horizon through 2050, 65% through 2090, and 44% through 2150.

Several factors contribute to the closeness of the estimates over our time period for the two different growth models. First, our reform scenarios gradually introduce changes, reflecting the lags for the policy to become fully effective and for the new, better-educated workers to change the average skills of the labour force. Second, the biggest impacts of the differences across the alternative models occur in the distant future, and thus the impact is lessened by discounting to obtain present values and by disregarding any returns that might accrue after 2090. Third, even ignoring discounting, the estimated convergence parameters imply very long periods before any country returns to its balanced growth path following a perturbation because of policy.

6.5. Sensitivity to alternative parameter choices

The prior estimates come for specific parametric choices. It is useful to consider the sensitivity of the projections to key choices. We do this for our baseline endogenous-growth-type model specification under Scenario II that brings each country to the Finnish level of PISA scores. In the baseline specification, the total value of the reform in the OECD amounts to \$275 trillion, or 645% of the current GDP of OECD countries.

6.5.1. Growth parameter. The baseline model, based on column 2 of Table 2, projects 1.86% of additional average annual growth for a one standard deviation increase in test scores. The first four columns of Table 9 perform the projections for the lowest and highest estimated parameters for cognitive skills in the different specifications of Table 2 – that is, 1.398 and 1.968, respectively. This leads to an estimate of the total discounted value of the education reform for the OECD of \$196 trillion and \$295 trillion, respectively (or 459% and 690%, respectively, of current GDP).

[14] In the underlying growth regression that controls for log per-capita GDP, the coefficient on the proportion of the population with scores higher than 400 is 6.478 and the coefficient on log initial income is −2.107 (both significant at the 1% level).

Table 9. Projection results under alternative parameter assumptions on growth coefficients and time horizon

	Lowest coefficient		Highest coefficient		Lower bound confidence		Upper bound confidence		Time horizon 2050		Time horizon 2150	
	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Australia	1,548	170	2,217	243	1,317	144	2,899	318	308	34	5,805	636
Austria	1,128	336	1,641	489	954	284	2,182	650	219	65	4,555	1,357
Belgium	1,163	278	1,683	402	986	235	2,224	531	227	54	4,576	1,093
Canada	2,022	144	2,888	206	1,722	123	3,767	268	404	29	7,497	534
Czech Republic	862	299	1,249	434	730	254	1,655	575	168	58	3,423	1,189
Denmark	893	423	1,310	621	753	357	1,757	833	171	81	3,750	1,778
Finland	0	0	0	0	0	0	0	0	0	0	0	0
France	8,268	364	12,065	531	6,989	308	16,083	708	1,596	70	33,815	1,488
Germany	12,519	409	18,348	600	10,567	346	24,569	803	2,400	78	52,262	1,709
Greece	2,993	825	4,558	1256	2,493	687	6,354	1751	540	149	15,071	4,153
Hungary	957	458	1,409	674	807	386	1,896	907	182	87	4,084	1,954
Iceland	48	385	70	563	41	325	94	753	9	74	198	1,591
Ireland	725	357	1,057	521	613	302	1,409	694	140	69	2,957	1,457
Italy	13,726	717	20,702	1081	11,473	599	28,562	1491	2,516	131	65,827	3,436
Japan	2,145	47	3,034	66	1,833	40	3,920	85	436	9	7,618	165
Korea, Rep.	565	40	799	56	483	34	1,031	72	115	8	2,000	140
Luxembourg	299	685	450	1030	250	573	619	1417	55	126	1,415	3,238
Mexico	26,267	1594	42,714	2593	21,416	1300	63,912	3879	4,306	261	186,466	11,318
Netherlands	996	141	1,423	202	849	120	1,855	263	199	28	3,689	524
New Zealand	204	163	292	233	173	139	381	305	41	32	762	610
Norway	1,425	487	2,104	719	1,200	410	2,839	971	270	92	6,162	2,107
Poland	3,829	521	5,672	772	3,221	438	7,679	1045	722	98	16,813	2,289
Portugal	2,018	785	3,063	1191	1,684	655	4,254	1654	366	142	9,983	3,882
Slovak Republic	568	477	838	704	478	402	1,130	949	108	91	2,445	2,054

Table 9. (Continued)

	Lowest coefficient		Highest coefficient		Lower bound confidence		Upper bound confidence		Time horizon 2050		Time horizon 2150	
	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Spain	8,845	568	13,158	844	7,428	477	17,893	1148	1,656	106	39,653	2,544
Sweden	1,285	343	1,872	500	1,087	290	2,490	665	249	66	5,208	1,391
Switzerland	852	245	1,229	353	723	208	1,619	466	168	48	3,304	950
Turkey	13,273	1258	20,994	1990	10,920	1035	30,478	2889	2,265	215	81,227	7,700
United Kingdom	5,804	244	8,372	352	4,926	207	11,029	464	1,141	48	22,498	946
United States	80,503	530	119,336	785	67,686	445	161,693	1064	15,155	100	354,821	2,335
OECD	195,731	459	294,547	690	163,804	384	406,270	952	36,131	85	947,885	2,221

Notes: Scenario II: Bring each country to Finnish level of 546 points on PISA. Discounted value of future increases in GDP until 2090, expressed in billion dollars (PPP) and as a percentage of current GDP.

Source: Authors' calculations.

An alternative way to account for the imprecision of the growth coefficient estimate is to use the lower and upper bounds of the 95% confidence interval around the baseline growth coefficient. These parameter bounds imply the net present value of the education reform is between $164 trillion and $406 trillion (columns 5–8 of Table 9).

6.5.2. Time horizon. As suggested by Figure 3, the time horizon for benefit calculations is clearly important. Columns 9–12 of Table 9 report the net present value to which the reform results aggregate for time horizons to the years 2050 and 2150, respectively. As already indicated by Figure 3, this clearly makes a huge difference. While we have implicitly emphasized how important it is to adopt a long-term horizon when considering education reform, even by 2050 the present value of the reform already accumulates to $36 trillion, or 85% of the current GDP of the OECD countries. By contrast, when adopting a time horizon until 2150, the (appropriately discounted) value of the reform sums to a staggering $948 trillion, or more than 20 times the current GDP.

6.5.3. Speed of reform. The baseline scenarios assume that it takes 20 years for the education reform to be fully implemented. The first four columns of Table 10 alternatively assume reform durations of 10 and 30 years, respectively. The faster reform implementation leads to an increase of the reform value to $341 trillion, whereas the reform value is ‘only’ $223 trillion if the reform takes 30 years to implement. Thus, while faster reform efforts obviously lead to substantially higher returns, even a slow (but successful) reform, if begun today, would have an enormous impact.

6.5.4. Working life. Columns 5 and 6 of Table 10 report results under the assumption that the average working life is 35 years, rather than 40 years, which appears to be a more reasonable estimate for many OECD countries. A shorter working life means that the replacement of the workforce with better-educated individuals completes faster, so that the aggregate value of the education reform increases. Assuming a 35-year work life yields a projection estimate of $304 trillion for the total value of the education reform.

6.5.5. Discount rate. The rate at which future returns are discounted obviously makes a substantial difference for the net present value of reform. Thus, rather than using the common-practice 3% discount rate of the baseline model, columns 7–10 of Table 10 instead use discount rates of 2.5% and 3.5%, respectively. The resulting discounted present values of the projected returns are $369 trillion and $207 trillion, respectively. For a larger band of discount rate of 2% to 4%, the total discounted reform value would be $497 trillion and $157 trillion, respectively (not shown). However, other projections of long-run effects, in the area of climate

Table 10. Projection results under alternative parameter assumptions on reform duration, working life, and discount rate

	10-year reform		30-year reform		Working life 35 years		Discount rate 2.5%		Discount rate 3.5%		Discount rate Stern Report	
	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Australia	2,541	279	1,721	189	2,287	251	2,788	306	1,581	173	4,763	522
Austria	1,887	562	1,263	376	1,694	504	2,062	614	1,165	347	3,533	1,052
Belgium	1,933	462	1,299	310	1,736	415	2,115	505	1,197	286	3,621	865
Canada	3,310	236	2,246	160	2,980	212	3,633	259	2,062	147	6,204	442
Czech Republic	1,435	499	963	335	1,289	448	1,570	545	888	309	2,689	934
Denmark	1,508	715	1,004	476	1,352	641	1,645	780	928	440	2,823	1,338
Finland	0	0	0	0	0	0	0	0	0	0	0	0
France	13,875	610	9,272	408	12,449	548	15,152	667	8,559	377	25,981	1,143
Germany	21,116	691	14,067	460	18,934	619	23,034	753	13,000	425	39,530	1,293
Greece	5,279	1455	3,421	943	4,705	1296	5,703	1572	3,193	880	9,855	2,716
Hungary	1,623	777	1,078	516	1,454	696	1,768	846	997	477	3,037	1,453
Iceland	81	648	54	432	72	581	88	707	50	399	151	1,213
Ireland	1,216	599	813	401	1,091	538	1,328	654	750	370	2,277	1,122
Italy	23,939	1250	15,624	816	21,367	1115	25,926	1353	14,545	759	44,724	2,335
Japan	3,472	75	2,372	52	3,130	68	3,820	83	2,173	47	6,510	141
Korea, Rep.	914	64	625	44	824	58	1,005	70	572	40	1,713	120
Luxembourg	520	1191	340	779	465	1063	564	1290	317	724	972	2,225
Mexico	50,015	3036	30,929	1877	44,093	2676	53,101	3223	29,371	1783	92,736	5,629
Netherlands	1,630	231	1,106	157	1,468	208	1,790	254	1,016	144	3,056	434
New Zealand	334	267	227	181	301	241	367	293	208	166	627	501
Norway	2,425	829	1,607	549	2,172	742	2,640	903	1,488	509	4,537	1,551
Poland	6,539	890	4,323	589	5,854	797	7,114	969	4,006	545	12,233	1,665
Portugal	3,546	1379	2,304	896	3,161	1229	3,834	1491	2,148	835	6,621	2,575
Slovak Republic	965	811	640	538	865	726	1,051	883	593	498	1,807	1,518
Spain	15,180	974	10,005	642	13,579	871	16,498	1059	9,282	596	28,390	1,822

Table 10. (Continued)

	10-year reform		30-year reform		Working life 35 years		Discount rate 2.5%		Discount rate 3.5%		Discount rate Stern Report	
	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP	$ billion	% GDP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sweden	2,152	575	1,440	385	1,931	516	2,351	628	1,329	355	4,030	1,076
Switzerland	1,411	406	951	273	1,268	365	1,545	444	875	252	2,643	760
Turkey	24,470	2320	15,436	1463	21,672	2055	26,174	2481	14,550	1379	45,513	4,315
United Kingdom	9,609	404	6,475	272	8,638	363	10,523	442	5,960	251	18,005	757
United States	137,603	906	90,922	598	123,157	811	149,676	985	84,272	555	257,402	1,694
OECD	340,528	798	222,527	521	303,987	712	368,864	864	207,075	485	635,982	1,490

Notes: Scenario II: Bring each country to Finnish level of 546 points on PISA. Discounted value of future increases in GDP until 2090, expressed in billion dollars (PPP) and as a percentage of current GDP.

Source: Authors' calculations.

change, have used much lower rates at which to discount the future. In particular, although highly disputed, the influential Stern (2007) Review places a much higher value on future costs and benefits by employing a discount rate of only 1.4% (see, e.g., Nordhaus, 2007; Tol and Yohe, 2006). That report also assumes a slightly lower rate of potential growth of 1.3% (rather than 1.5% as in our other models), which we use in our next estimate. Note that what is ultimately relevant for the projections is the difference between discount rate and rate of potential growth, yielding an effective discount rate of 0.1% in this scenario. If we were to adopt the discounting practice of the Stern report, the present value of the education reform would sum to a staggering $636 trillion, or roughly 15 times the current GDP, by 2090 (final columns of Table 10).

7. POLICIES TO IMPROVE EDUCATIONAL ACHIEVEMENT

The previous discussion has stressed the immense long-term benefits of cognitive skills for economic growth. Table 11 summarizes the baseline projections normalized against the discounted value of the projected future OECD GDP over the same time span (until 2090). The value of the reform amounts to 4.3–13.8% of the present value of future GDP. Independent of whether the underlying economic model is specified in endogenous-growth or neoclassical terms, improved educational achievement is projected to have a large impact on future economic well-being of OECD countries.

Yet simply knowing that skill differences are important does not provide a guide to policies that might promote more skills. Indeed, a wide variety of policies have

Table 11. Summary of projection results

	Scenario I: Increase avg. performance by ¼ std. dev.	Scenario II: Bring each country to Finnish level of 546 points on PISA	Scenario III: Bring all to minimum of 400 points on PISA
	(1)	(2)	(3)
'Endogenous-growth' specification			
in $billion	123,108	275,429	226,333
in % of discounted future GDP	6.2	13.8	11.3
'Neoclassical' specification			
in $billion	90,031	179,655	187,191
in % of discounted future GDP	4.3	8.5	8.9

Notes: Discounted value of future increases in OECD GDP until 2090, expressed in billion dollars (PPP) and as a percentage of discounted value of all annual projected OECD GDPs until 2090.

Source: Authors' calculations.

been implemented within various countries without much evidence of success in either achievement or economic terms. We believe that the disappointing results of the past generally reflect pursuing policies for which there is little empirical support.

Substantial research has gone into understanding why achievement differs across students and across countries. While controversies have existed about the interpretation of various individual pieces of evidence, considerable agreement now exists about what kinds of approaches are unreliable avenues for change. There is perhaps less agreement on the precise approaches that might be followed, but even here there is a growing consensus about the general sets of policies that have shown promise based on more credible research approaches.

The work on achievement determinants generally falls under the heading 'education production functions'. The extensive work has taken a variety of perspectives and approaches. The general objective is to sort out the causal impacts of school factors (things that can potentially be manipulated through policy) from other influences on achievement including family background, peers, neighborhood influences, and the like (which are less readily amenable to policy change).

7.1. Evidence on school resources

The most extensive generally available evidence relates to the effects of resources. Many policies undertaken involve substantial flows of resources – direct spending, changes in teacher salaries, reductions in class size, and the like – made within the context of current school organization. The empirical evidence clearly documents the difficulties with such policies. Simply providing more resources gives, according to the available evidence, little assurance that student performance will improve significantly. The underlying analyses of resources include studies within individual countries and across different countries and have been extensively reviewed elsewhere (see Hanushek, 2003; Woessmann, 2007a; Hanushek and Woessmann, 2011).

The easiest way to see the situation is a simple cross-country picture. Figure 6 plots the descriptive association between educational expenditure per student and educational outcomes from the PISA 2006 study. Ignoring Mexico and Turkey where cumulative expenditure per student (over the age range of 6 to 15 years) is less than \$20,000, there is no association between educational spending and educational outcomes across OECD countries. Of course, many more factors such as students' family backgrounds enter the determination of educational outcomes.

The simplest way of addressing bias from unobserved time-invariant country factors is to ignore level differences and restrict the analysis to changes in expenditure and outcomes over time. From combining maths scores from PISA 2000, 2003, and 2006, it is possible to estimate a country-level regression based on either first differences or country fixed effects. Either approach yields an association between expenditure per student and educational outcomes that is not close to statistical

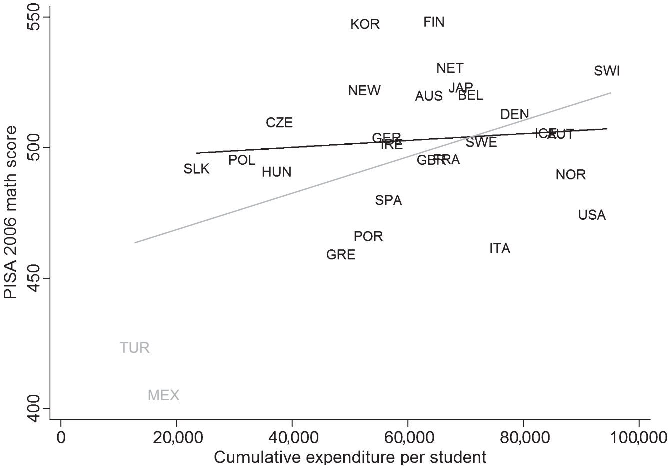

Figure 6. Educational spending and student achievement across countries

Notes: The figure plots the average maths achievement in PISA 2006 against cumulative expenditure on educational institutions per student between age 6 and 15, in US dollars, converted by purchasing power parities. Gray line: regression line for full sample. Black line: regression line for countries with expenditure above $20,000.

Source: Authors' depiction based on Organization for Economic Co-operation and Development (2007).

significance at conventional levels and actually negative (see Hanushek and Woessmann, 2010b).

For policy deliberations, information on the impact of resources from within individual countries has accumulated and is perhaps more appropriate than cross-country information. This work underscores the importance of including non-school factors by demonstrating that family differences are very important in explaining variations in achievement across students. It also yields consistent results on school resources.

The general picture about school resources in developed countries is now well known and has been reviewed in a variety of places (see Woessmann, 2005a for Europe and Hanushek, 2002, 2003 for the United States). The available studies concentrate on various common school inputs such as teacher experience or class size that are readily available in data sets and frequently identified as the focus of policy. The available econometric evidence now includes hundreds of separate estimates within the United States and other developed countries. Quite uniformly, there is little strong evidence that *any* of the following factors has a consistent impact on achievement: the level of teacher education, the pupil-teacher ratio, characteristics of administration, or facilities of the school. Specifically, aggregating results across studies, a minority of estimates are statistically different from zero (at the 5% level), and the studies do not even uniformly indicate improvements in performance with increased resources (see Hanushek, 2007; Woessmann, 2007a). A second line of studies focuses on financial inputs. A number of studies simply relate spending per student to achievement or capture teacher differences by teacher salaries. While these studies tend to be lower quality, they also fail to show a consistent relationship between financial resources and achievement.

These results have been controversial (see, e.g., Burtless, 1996; Greenwald *et al.*, 1996; Hanushek, 1996). But the current consensus is that policies beyond simple increases in resources are essential. A simplistic view of the results – convenient as a straw man in public debates – is that 'money never matters'. The research of course does not say that. Nor does it say that 'money cannot matter'. It simply underscores the fact that there has historically been a set of decisions and incentives in schools that have blunted any impacts of added funds, leading to inconsistent outcomes. More spending on schools has not led reliably to substantially better results.

7.2. Teacher quality

The most current research on school inputs and achievement has also led to another set of conclusions – that teacher quality is enormously important in determining student achievement. This work has concentrated on whether some teachers consistently produce more gains in student achievement than other teachers. (See, for example, Hanushek, 1971, 1992; Rockoff, 2004; Rivkin *et al.*, 2005; and a number of subsequent studies reviewed in Hanushek and Rivkin, 2010.) Working with extensive panel data on individual students from different US states, these studies have confirmed large differences among teachers in terms of outcomes in the classroom.

But, they have also shown that the observed differences are not closely related to commonly observed characteristics of teachers (such as amount of teacher education). Some attributes of teachers – such as having one or two years of experience – have explained part of the differences in teacher quality, but these factors are a small part of the overall variance in teacher results. This inability to identify specific teacher qualities makes it difficult to regulate or legislate having high-quality teachers in classrooms. It also contributes to our conclusion below that changes in the institutional structure and incentives of schools are fundamental to improving school outcomes.

7.3. Institutional structures and incentives in the school system

Existing evidence suggests some clear general policies related to institutional structure of schools that are important. Foremost among these, the performance of a system is affected by the incentives that actors face. That is, if the actors in the education process are rewarded (extrinsically or intrinsically) for producing better student achievement, and if they are penalized for not producing high achievement, achievement is likely to improve. The incentives to produce high-quality education, in turn, are created by the institutions of the education system – the rules and regulations that explicitly or implicitly set rewards and penalties for the people involved in the education process.

From existing work, three interrelated policies come to the forefront: promoting more competition, so that parental demand will create strong incentives to individual schools; autonomy in local decision making, so that individual schools and their leaders will take actions to promote student achievement; and an accountability system that identifies good school performance and leads to rewards based on this.

7.3.1. Choice and competition. Choice and competition through school vouchers were proposed a half century ago by Milton Friedman (1962). The simple idea is that parents, interested in the schooling outcomes of their children, will seek out productive schools, yielding demand-side pressure that creates incentives for each school to produce effective education and to ensure high-quality staff in addition to a good curriculum.

In many school systems (with the Netherlands being the most obvious example), privately managed schools (with public funding) provide alternatives for students. These schools, which also often have a religious affiliation, are part of the natural institutional framework. Unfortunately, little thorough evaluation has been done of these choice possibilities, in large part because there is no obvious comparison group (i.e., choice is instituted for an entire country and there is no example of the no-choice alternative). In a cross-country comparison, students in countries with a larger share of privately managed schools perform better on average (cf. Woessmann, 2007b, 2009; Woessmann *et al.*, 2009), and recent evidence corroborates the conclusion that this is due to a causal effect of private-sector competition (West and Woessmann, 2010).

In the United States there are limited examples of private school choice, ranging from the publicly funded school vouchers in Milwaukee, Cleveland and Washington DC, to privately financed voucher alternatives. The evaluations of these generally show that the choice schools do at least as well as the regular public schools, if not better (see Rouse, 1998; Howell and Peterson, 2002; Wolf *et al.*, 2010).

In Europe, Bradley and Taylor (2002) and Levačić (2004) find similar positive effects of school competition on the performance of English schools. Sandström and Bergström (2005) and Björklund *et al.*, (2004) provide evidence on significant positive effects of competition from privately operated schools on the performance of public schools in Sweden. Filer and Münich (2003) show that the introduction of a voucher-type system in the Czech Republic led to the creation of private schools in areas where public schools are doing badly and that the public schools facing private competition improved their performance.

7.3.2. Autonomy and decentralization. Several institutional features of a school system can be grouped under the heading of autonomy or decentralization, including fiscal decentralization, local decision making on different matters, and parental involvement. Almost any system of improved incentives for schools depends upon having school personnel in individual schools and districts heavily involved in decision making. It is difficult to compile evidence on the impact of autonomy, because the degree of local decision making is most generally a decision for a country (or state) as a whole, leaving no comparison group within countries. Across countries, students tend to perform better in schools that have autonomy in personnel and day-to-day decisions (Woessmann, 2003, 2007b; Woessmann *et al.*, 2009), in particular when there is accountability (see also the review in Hanushek and Woessmann, 2011).

The US states have varying amounts of local autonomy. One systematic form of school autonomy is 'charter schools', which are public schools that are allowed to perform quite autonomously. (Note that these are actually hybrids of choice schools and public-school autonomy, because they survive only if sufficient numbers of students attend them.) These schools are relatively new, a fact that complicates evaluation since many are still in the start-up phase. The evidence on them is mixed but indicates a variety of places where charter schools outperform the regular public schools after the initial start-up phase. But it also suggests in part that the regulations governing them and the particular competitive public schools they face have an influence.[15]

7.3.3. School accountability. Many countries around the world have been moving toward increased accountability of local schools for student performance. The United Kingdom has developed an elaborate system of 'league tables' designed to give parents full information about the performance of local schools. The United States has legislated a federal law ('No Child Left Behind') that all states must have an accountability system that meets certain general guidelines. It also sets into law a series of actions required when a school fails to bring all students up to proficiency in core subjects.

Evidence on the impacts of these systems has begun to accumulate. While there is some uncertainty given the newness of the overall federal accountability system (introduced in 2002), the best US evidence indicates that strong state accountability systems in fact lead to better student performance (Carnoy and Loeb, 2002; Hanushek and Raymond, 2005; Jacob, 2005; Dee and Jacob, 2009).

One institutional set-up that combines accountability with parental choice are systems that give students in schools that repeatedly do badly on the accountability

[15] A number of studies have based the analysis on student fixed effects, relying on students moving in and out of charter schools to identify the impact of charters (Booker *et al.*, 2007; Bifulco and Ladd, 2006; Hanushek *et al.*, 2007). These studies have generally pointed to a range of quality for charter schools but highlight start-up problems. Another set of studies considers charter schools that have more demand than open positions, requiring schools to choose their students by lottery (Hoxby and Murarka, 2009; Abdulkadiroglu *et al.*, 2009). Finally, matching methods have been used to compare public and private school performance, leading to the conclusion that there is wide variation in the quality of charter schools compared to the relevant public schools (CREDO, 2009, 2010).

test a voucher to attend private schools. In Florida, the threat of becoming subject to private-school choice if failing on the test has been shown to increase school performance particularly for disadvantaged students (West and Peterson, 2006; Figlio and Rouse, 2006).

Curriculum-based external exit exams are another means to introduce some form of accountability into the school system. They provide performance information which can hold both students and schools accountable. Students in countries with external exit exam systems tend to systematically outperform students in countries without such systems (Bishop, 1997, 2006; Woessmann, 2001, 2003, 2007b; Woessmann *et al.*, 2009). In Canada and Germany, the two national education systems where the existence of external exams varies across regions, students similarly perform better in regions with external exams (Bishop, 1997; Jürges *et al.*, 2005; Woessmann, 2010).

It is difficult to imagine choice or autonomy working well without a good system of student testing and accountability. Thus, the ideas about institutional structure are closely linked together. The international evidence clearly suggests that school autonomy, in particular local autonomy over teacher salaries and course content, is only effective in school systems that have external exams (Woessmann, 2005b, 2007b; Fuchs and Woessmann, 2007; Woessmann *et al.*, 2009). For example, school autonomy over teacher salaries is negatively associated with student achievement in systems without external exams, but positively in external-exam systems.

Finally, given the importance of high teacher quality, a promising candidate for improvement is the specific form of accountability that aims incentives directly at teachers. While convincing evidence on the effects of performance-related teacher pay is scarce, the more rigorous studies in terms of empirical identification tend to find a positive relationship between financial teacher incentives and student outcomes (cf. the surveys in Atkinson *et al.*, 2009 and Podgursky and Springer, 2007). Thus, Atkinson *et al.* (2009) find that the introduction of performance-related pay had a substantial positive impact on student achievement in England. Dolton and Marcenaro-Gutierrez (2011), using a panel of countries, provide evidence that aggregate changes in salaries over time lead to higher student performance. At the school level, monetary incentives for teachers based on their students' performance have been shown to improve student learning very significantly in Israel and in India (Lavy, 2002, 2009; Muralidharan and Sundararaman, 2009).

Most evaluations of performance pay systems nonetheless focus on whether existing teachers change their behavior – what is referred to as the 'effort' margin. There are many reasons to believe, however, that the attraction of new teachers and the retention of the more effective teachers – the 'selection' margin – are more important. (Indeed, in addressing what contributes to outcomes in the best performing countries, Barber and Mourshed (2007) emphasize both initial recruitment of teachers and the development of institutions to move ineffective teachers out of the classroom.) The importance of pay for selection is difficult to analyse because it generally involves considering longer-run incentives that are often at the aggregate level. For this, cross-country variation provides some indication that students perform better in countries that allow for teacher salaries to be adjusted based on performance in teaching (Woessmann, 2011).

7.4. The cost of reform

The projections provide the gross returns to improved schools, but these should be offset by the costs needed to obtain the achievement gains. Unfortunately, the costs are not easy to estimate. At one level, the costs of the institutional changes suggested previously appear mainly to be transition costs – expenses needed in training people (including parents) in new systems, developing testing and monitoring systems, covering short-run duplicate expenditures with the introduction of new schools and potential semi-fixed costs of old schools, and so forth.

Yet, it is also very possible that recruiting and retaining better teachers will involve higher salaries and added expenses. This would almost certainly be the case if more emphasis is placed on active retention policies that make teaching a more risky occupation. Further, the previous institutional changes may be insufficient without the addition of other programmes that entail added costs.

While it is difficult to estimate these costs directly, some reasonable bounds might be inferred from current spending. In 2007, spending on primary and secondary schools in the OECD ranged from 2.5% of GDP (Slovak Republic) to 5.1% of GDP (Iceland) with an average of 3.6% (OECD, 2010). Total spending including tertiary schooling averaged 5.7% in the OECD. Of the primary and secondary school expenditures, about 60% goes to teacher compensation, implying that a 50% increase in all teacher salaries – or equivalently doubling the salaries of half of the teachers – would amount to slightly over 1% of GDP.

It is useful to put this discussion into the context of the benefits indicated in Table 11. The benefits of a 25-point improvement in PISA scores average 4.3% of GDP in the conservative neoclassical estimates – implying that the very dramatic policies of paying all teachers 50% more would cost less than one-quarter of the projected benefits *if this policy led to a 25-point improvement*. It is important to emphasize the latter point, because many nations of the world have put in place large spending increases with significant teacher pay increases without getting the student achievement gains. In other words, this benefit-cost calculation only holds if the new policies reach the achievement goals.

That having been said, it is important to note that the most significant costs may be political costs. The projections show clearly that the economic gains come only after a lengthy time involving the reform of schools and the introduction of noticeable proportions of higher skilled workers into the labour force. This disjointed nature of costs and benefits is not unusual in many public programmes, but it is more extreme, because the benefits come only after a majority of current politicians have left office.

This mismatch may not be decisive. First, politicians in many countries already campaign on the possibility of improving schools. While current campaigning is often for failed programmes – increases in spending or reductions in class size that have not proved successful – it seems possible to redirect these efforts to more productive areas. Second, in matters such as climate change, politicians have actively engaged in long-run activities that are even more mismatched in terms of timing of costs and benefits.

8. CONCLUSIONS

It is generally the case that national attention to economic policies that deal with current aggregate demand conditions and with business cycles invariably take priority over longer-run policy considerations. Perhaps this has never been as true as today, when the most obvious focus of attention is the fiscal turmoil from the worldwide recession. The message of this paper is, however, that considering issues of longer-run economic growth may be much more important for the welfare of nations. Nobel Laureate Robert Lucas, in his presidential address to the American Economic Association, concluded that 'Taking US performance over the past 50 years as a benchmark, the potential for welfare gains from better long-run, supply-side policies exceeds by far the potential from further improvements in short-run demand management' (Lucas, 2003).

Our results show that education policy is closely associated with the long-run growth potentials of OECD countries. The regression analyses suggest that direct measures of educational outcomes, in terms of cognitive skills on international achievement tests, emerge as the one strong policy factor underlying growth differences across OECD countries. By contrast, a long battery of institutional and regulatory measures does not add to an explanation of the substantial differences in long-run growth rates that exist across OECD countries, mainly because all OECD countries share a common set of basic institutional structures that ensure a general functioning of market economies. Considering different skill dimensions, basic skills are robustly related to OECD-country growth, whereas the relation of the top-skill dimension with growth is at least substantially smaller than in non-OECD countries. When cognitive skills are accounted for, tertiary attainment *per se* is not significantly associated with long-run OECD growth.

Our projection analysis suggests that, under plausible parameter assumptions, the real present value of future improvements in GDP due to challenging but achievable educational reform scenarios amounts to $90–275 trillion. A plausible goal of having all OECD countries boost their average PISA scores by 25 points (one-quarter standard deviation) implies an aggregate gain of OECD GDP of $90–123 trillion. More aggressive goals, such as bringing all students to a level of minimal proficiency for the OECD or bringing all OECD countries to the level reached by Finland today, would imply aggregate GDP increases beyond $200 trillion according to historical growth relationships. The precise size of the reform value of such long-run projections is clearly up for debate. Nevertheless, our sensitivity analyses indicate that, while differences between an endogenous and neoclassical model framework and alternative parameter choices clearly make a difference, the estimates of the long-run effect of reasonable education reforms still yield enormous values no matter what.

Our projections do not by themselves indicate how schools should be changed. Nor do they solve the political economy issues of how any change should be achieved politically. They simply underscore the high cost of political inaction or misdirection.

In order to provide some guidance, we review the extensive relevant research on the determinants of educational achievement. Several conclusions appear. First, many of the traditional policies of simply providing more funds for schools or of adding specific resources such as smaller classes do not provide much hope for significant improvement in student achievement. Second, a growing body of research shows that teacher quality is a primary driver of student achievement but that differences in quality are not closely related to teacher education and experience. Because teacher quality is not easily measured and regulated, effective policies to improve quality appear to necessitate more careful attention to the incentives faced by schools and teachers. Here the research on educational institutions suggests productive policy approaches. In particular, evidence from both within and across countries points to the positive impact of competition among schools, of accountability and student testing, and of local school autonomy in decision making.[16] Research on these policies, separately and in combination, indicates some continuing uncertainty about the magnitude of any effects but does support more aggressive attention to these in setting school policies.

An important aspect highlighted by the projections is the dynamic nature of human capital and growth. Our basic characterization of growth indicates that higher cognitive skills offer a path of continued economic improvement, so that favourable policies today have growing impacts in the future. However, the full ramifications of schooling outcomes will not become apparent until reasonably far into the future. The economic gains from education reform are surely not reaped within matters of one or two political legislation periods. They rather require a long-run perspective that fully considers the time horizon of a child born today. In the discussion of climate policies, it has become custom to consider expected outcomes that materialize several generations from now. Education policy needs a

[16] In a variety of other work, the importance of different educational institutions is investigated. See the international study of Woessmann *et al.* (2009) and the United States analysis of Hanushek and Lindseth (2009). Mourshed *et al.* (2010) also provide a nuanced view of the impact of institutions on performance.

similar long-term perspective to fully capture the consequences of possible current reforms.

Discussion

Fabiano Schivardi
University of Caqliari and EIEF

This is an important paper. It contributes to the debate on the role of human capital in the determination of economic growth. It builds on an impressive body of previous work by the same authors. In concludes that improvements in the quality of human capital can greatly contribute to growth and proposes education as the growth-enhancing policy for the current century. I believe that this thesis deserves to be taken seriously. In the rest of the discussion I will elaborate more on the argument and I will offer some words of caution on the interpretation of the results and on the feasibility of policies aimed at improving the education system.

The paper has three parts. First, it shows that scores in the PISA tests – defined as cognitive skills (CS) – are very important in growth regressions. Then, using the results from the growth regressions, it constructs alternative GDP growth scenarios deriving from different assumptions on the evolution of CS. Finally, it discusses how to improve CS. The policy conclusion is clear cut: CS improvements can greatly enhance the growth performance of OECD countries. In addition to being large, the effects are also very robust with respect to possible confounding factors. In particular, in the growth regressions they outperform the average years of schooling (Barro–Lee), that are not significant once CS are included, as well as some of the other usual suspects, such as regulation and institutions. In general, it is hard to disagree with the qualitative conclusion that education is a fundamental driver of growth. More caution is required to interpret the quantitative statements. The quantitative results rest on a series of assumptions and require taking the estimates very seriously. Given that this is fundamentally a quantitative paper, I will propose some caveats in interpreting the conclusions and in translating them into policy recommendations.

Time series versus cross-sectional variation

The estimates are based on cross-sectional differences in PISA scores. That is, the authors use the average within-country values of the test scores as a regressor in a cross-sectional growth regression (GDP per capita) for the period 1960–2000. Therefore, the data variability that identifies the effects comes from cross-country differences. On the contrary, the evaluation of the growth effects through the simulations are based on changes in CS over time within each country. Clearly, there is a tension between these two aspects. In fact, cross-country regressions are subject to the usual criticisms – reverse causality and omitted variable bias. The authors show that CS beat some other horses in the regressions. But one can always propose a different horse. For example, 'culture' might be causing both CS and growth: there might be a cultural trait that induces both high school performance and high economic performance. Moreover, CS is computed as a 1960–2003 average, using all the available tests. Often these tests refer to the final period considered in the growth regressions. The implicit assumption is that they are a sort of fixed country attribute, which does not vary much over time.

The authors are well aware of this problem. Indeed, they dealt with it in the 2009 NBER working paper (Hanushek and Woessmann 2009), using IVs and also differences in achievements of immigrants trained abroad and in the United States. It would be useful to reference more directly that analysis. A different route to address these issues would be to follow Ciccone and Papaioannou (2009) and use the cross-sectoral differences in growth according to the importance of human capital for each sector. The idea is to determine if more human capital intensive sectors grow relatively more in countries with high CS. As shown by the seminal contribution of Rajan and Zingales (1998), this approach allows for the use of country fixed effects and it can account for all potential country fixed attributes. For example, the effects of specific 'cultural traits' should be uniform across sectors, while the relevance of human capital varies according to the technological content of each sector. In general, I believe much can be learned from the effects of CS on more than aggregate per capita GDP growth.

Can CS be improved substantially?

The relative stability of CS at the country level suggests that improving them is not an easy task. In fact there are various inputs in the human capital production function, such as schooling, innate ability, family inputs, etc. The most policy relevant input is schooling. The paper clearly states that we know little on what characteristics of the schooling system are more conducive to high student performance, despite the fact that a substantial amount of research has investigated the question. In fact, the most natural policy variables do not seem to affect substantially students' outcomes: amount of resources, class size, teachers' experience etc. More research will be needed to determine the extent to which and how CS can be improved.

CS versus tertiary education

CS are measured at the high school level. Based on this, the authors challenge, the Vandenbussche, Aghion, and Meghir (2006) idea that basic education is more important for developing countries, tertiary for developed ones. This is done by showing that: (a) in OECD countries what matters most is the share of students above basic CS, rather than above high CS; (b) the years of tertiary schooling are unimportant after controlling for CS. I find this analysis less convincing than the rest of the paper. It is unclear how CS relates to educational attainments (years of schooling). It might just be that in developed economies basic CS represents the relevant threshold to determine college attendance. Moreover, the US is an outlier in this analysis, as it has high growth, high tertiary achievements and a more modest CS performance. Indeed, this country is also the most technologically advanced. An alternative interpretation is that it represents the technological frontier, also thanks to the tertiary system, and all other OECD countries are followers. This would be in line with the Vandenbussche, Aghion, and Meghir (2006) interpretation, and would leave open the possibility that in the future the tertiary education might become as important also in other OECD countries.

A superficial (but at risk) reading of these results could be that only quality (measured by CS) matters, and quality early on. Even in this case, there are two possible interpretations: (a) Higher education is of second order importance; (b) Both higher and lower education matter, in a complementary way. It would be interesting to estimate something like:

$$human\ capital = \alpha\ quality + \beta\ years + \gamma(quality * years)$$

The results in the paper indicate that β is small, but γ might be large. The policy prescription would be very different: increasing years of schooling is also important. From a policy perspective, this in an easier goal than increasing CS. For example, it has been the crucial goal of the reform of the university system in Italy at the beginning of the 1990s. In a recent paper of mine with Roberto Torrini (2010), we find evidence that a higher supply of college graduates at the local level caused more productivity growth.

The reform scenarios and their effects

The main message is that the effects are so large that the net present value (NPV) of the reforms is always substantial, under different parameter values, experiments and growth framework. In terms of the distinction between endogenous growth and neoclassical framework, I would emphasize more the conditional convergence scenario, as scale effects require some specific assumptions on the input accumulation equation. Finally, the NPV calculations would have been perfect for computing the returns from investing in education and overlapping generations analysis: who pays? This could be the subject of future research.

Policies to improve CS

The results from a large body of literature, mostly for US schools, lead to the conclusion that simply increasing spending does not guarantee substantial improvements in the quality of education. Hanushek and Woessman propose incentives as the crucial element, promoted through competition, autonomy and accountability. This is very much an economist's perspective – and totally agreeable. Still, it would be useful to provide some evidence that these are important ingredients of the Korean and Finnish school systems, that are at the top of the CS distribution: is this the case? I suspect that this is not the whole story. It would be important to put some more attention to purely educational aspects: curriculum, teaching technology, class discipline and so on. Is there something we can learn from education experts? What is so special of the Korean and Finnish systems?.

Luigi Pistaferri[17]

Stanford University

1. Introduction

This paper deals with three issues.

First, it revisits the very large literature on the determinants of long term growth. It identifies three important factors for long-term growth: (a) human capital investments, (b) institutions, i.e., security of property rights etc., and (c) regulations, especially of product and labor markets (firing and hiring restrictions, extent of red tape, etc.). The authors' conclusion is that human capital variables appear to be the dominant force among the three.

The second issue addressed by the paper is quantitative. Having assessed that human capital is the dominant force for explaining differences in long run growth, how large is its effect?

Finally (and this is a theme that runs across the paper), are the answers to these questions different for developing and developed countries? Most of the empirical analysis is conducted on a sample of OECD countries, but the authors also show results for an extended sample, and argue that institutions don't matter among OECD countries but do matter for explaining growth differences between developed and developing countries; and that regulations don't affect long-term growth prospect but do have an effect in the short run.

The paper has three important take-away points:

(1) Using "average schooling" to measure human capital is misleading because, say, 12 years of schooling don't mean the same thing across countries. A given year of schooling delivers different increases in knowledge and skills in different countries, typically as a function of the strength of the education system in

[17] Stanford University.

which it is experienced. The authors try to avoid this pitfall and measure the human capital of a country with an index of cognitive ability. In particular, they propose using the average PISA score, which has the advantage of being comparable across countries because pupils in different countries all take the same test. Moreover, the PISA score is a sufficient statistics because it reflects quality and not just quantity, together with being also a reflection of all inputs in the process of human capital production, such as parental background, etc.

(2) Their new measure of human capital is a very powerful one – it improves the fit of the model and it knocks down the statistical significance of all other variables traditionally associated with long term growth, such as institution and regulation variables.

(3) Quoting almost verbatim from the paper, feasible reforms (that is, reforms that are being planned or have already been phased in in some countries) would result in *extremely large long term benefits* in terms of economic growth.

In short, this is a very good paper, with quite a lot in it, and a clear discussion of the possible policies that could be used to realize the growth increase predicted by their regressions, although – somewhat understandably – the authors shy away from actually making precise policy recommendations. The evidence is strong and the estimated effects of human capital on growth large. If these estimates are taken at face value, they ought to change the way we think about determinants of long-term growth and country-wide investment priorities.

2. Discussion

Let me now come to a discussion of some aspects of the paper that can be controversial. I will start by making two remarks on measurement.

(1) The paper argues that the PISA score offers a standardized measure of cognitive ability, i.e., something that is easily comparable across countries, unlike average years of schooling. I will offer some thoughts on whether this is effectively the case.

(2) The authors treat the concept of "variable quality" a bit asymmetrically. While it's true that an increase in "years of schooling" do not necessarily result in an equivalent increase in the stock of HK in two different countries, a similar argument can be made for the institutional and regulation measures used in their regressions.

Let me elaborate on the first point, the comparability of PISA score across countries. What is the PISA test score measuring? PISA tests just a few areas of the school curriculum: Reading, Math, and Science. It does not test knowledge of social sciences (such as economics!), history, arts, or foreign languages. But what if success in real life is more related to how much economics you know than how much chemistry formulae you're able to decipher? There is now a small literature in economics arguing that financial literacy (and not Mendeleev's periodic table) taught early on improves dramatically permanent income. Second, what kind of incentives do children have to answer correctly the PISA test questions? A small internet search reveals that in most countries children know (or are told) that their "school grade" (which is the one they really care about) is independent of how well (or bad) they perform on the PISA test, and may not even be told how well they did on the test, let alone have someone discussing the answers, etc. In other countries, in contrast, getting on the top of the PISA table seems to be a matter of national pride (for example, Korea, Taiwan and Singapore). Other countries appear to have changed their attitudes towards the importance of PISA test scores (for example, Norway and Germany have started to allocate resources to that specific goal after discovering that their kids were not doing as well as they thought they would). If these trends exist, they invalidate the use of fixed effects as a cure for all institutional differences that may exist across countries. Moreover, they may fire backs conceptually: Teachers may have incentives to "teach to the test" and the test score itself may lose the "economic" content the authors attach to it.

As for the second point, the use of indexes to measure the strictness of regulations in labor or product market may be misleading (this is a general point, it does not apply to this paper in particular). To give an example of the dangers involved, consider forming an index on the strictness of firing regulations. In many countries "small firms" are exempt from firing restrictions. Suppose countries A and B exempt firms with 15 employees or less. At first sight, it may seem that the two countries are similar in terms of the strictness of their firing restrictions. However, suppose that the firm size distribution in the two countries is such that in country B most of the firms are (relatively) large, while in country A most firms are small (below 15). In this case the regulation is really binding only for country B. Hence, policies that look similar may actually be quite different once one weights in enforcement, industrial structure (in this case), and so forth. It is therefore possible that the use of "correct" measures of institutions or regulatory powers might revert or weaken their negative results about the contributions of these variables to economic growth.

3. Results and conclusions

The paper tries to measure the effect on long-term growth of three types of policies: (1) Rising PISA score by 25 points; 2) Bringing every country to the level of highest-scorer Finland; and 3) Bringing all kids to a minimum proficiency level.

The projection exercise is interesting and useful. In principle, there are issues:

(1) Any reform (such as increasing the PISA score by 25 points) would probably require putting money somewhere (teachers' pay, etc.). Financing of these

reforms is not discussed, but higher taxes to pay for the reforms would hamper growth (and may even discourage human capital investments in turn). This most likely would be exaggerating the effects of the reform.

(2) The cost of the reform ought to be different for countries in different points of the distribution of PISA scores.

The authors tackle these issues is Section 7.4 and the findings remain robust, which is reassuring.

In conclusion, this a very interesting and ambitious paper which has some very strong findings. In particular, in OECD countries the role of institutions and regulation in explaining long-term growth differences disappears once a credible and comparable measure of human capital (the PISA test score) is used. It also finds that feasible reforms that enhance the human-capital of a country may have very large effects on long-term growth. What is somewhat absent from the paper is a clear mapping from possible policies onto "feasible" reforms, and this would I'm sure be the subject of future work.

Panel discussion

In opening the discussion, Marco Pagano asked if there was any country level evidence on the effect of major education reforms on long-term growth. Such analysis would provide better insight into the effects of changes in education on growth over time and would complement the cross-sectional analysis in this paper. He also contended that as the pay-off from policies which increase the quality of education is so far into the future the political will to make such investments would be very small and he wondered how such political incentives could be altered. He added that this issue was particularly relevant given that many countries are entering into a period of fiscal austerity.

Dalia Marin believed the schooling measure used in the paper underestimated the effect of schooling stock on long-run growth. She suggested an improved proxy for human capital stock may decrease the large effect of the quality of education on GDP growth found in the paper. In work she has done on the returns to education, where employee skill levels are weighted by their market wage, she found that the human capital stock adds 0.5% to GDP per capita per year in Germany. Volker Nocke pointed out that the cognitive skills measure used in the paper includes a measure of the cognitive skills of students still in education. He argued that it is more appropriate for the authors to use initial cognitive skills when estimating its effect on GDP growth.

Michalis Haliassos focused on the importance of institutions and the incentives of people to invest in education. For example, in countries where personal connections matter for getting a job this will undermine the importance of education and lead to the misallocation of resources and lower growth. In this case, increased spending on education would not necessarily lead to higher levels of education. Morten Ravn agreed that institutions and a cultural mentality to succeed were important factors that one must control for when estimating the returns to education. Monica Paiella focused on the importance of labour market institutions and in particular wage rigidities in teacher salaries. She noted that in many countries teachers' salaries are not linked to the performance of their students. Such rigidities are likely to impact on the overall cognitive development of students. Following Luigi Pistaferri's discussion on the scope of the PISA test for testing various aspects of cognitive skills, she noted that the purpose of the PISA test is to measure outcomes of learning beyond the school curriculum and therefore the exclusion of other social sciences from the test should not matter too much.

In reply to comments on the shortcomings of the cross-country regression approach made by Andrea Ichino and others, Eric Hanushek referred to other research he has completed which use a variety of approaches and continue to identify a positive effect of education on growth. In response to Marco Pagano's comments on the political economy of educational spending, Eric Hanushek conceded it was very difficult to answer how one could create the political incentives to invest in education given the pay-offs are in the very medium to long run. He did note that China has managed to develop an institutional structure which supports longer-run investments in education. Finally, Hanushek accepted the comments that more recent PISA test scores may be less representative of the cognitive development of a country's population as some countries now place greater pressure on students to perform well in this test.

REFERENCES

Abdulkadiroglu, A., J. Angrist, S. Dynarski, T.J. Kane and P. Pathak (2009). 'Accountability and flexibility in public schools: Evidence from Boston's charters and pilots', NBER Working Paper 15549, Cambridge, MA.

Acemoglu, D., S. Johnson and J.A. Robinson (2001). 'The colonial origins of comparative development: An empirical investigation', *American Economic Review*, 91(5), 1369–401.

—— (2005). 'Institutions as a fundamental cause of long-run growth', in P. Aghion and S.N. Durlauf (eds.), *Handbook of Economic Growth*, North Holland, Amsterdam, pp. 385–472.

Aghion, P. and P. Howitt (2006). 'Appropriate growth policy: A unifying framework', *Journal of the European Economic Association*, 4(2–3), 269–314.

Atkinson, A., S. Burgess, B. Croxson, P. Gregg, C. Propper, H. Slater and D. Wilson (2009). 'Evaluating the impact of performance-related pay for teachers in England', *Labour Economics*, 16(3), 251–61.

Barber, M. and M. Mourshed (2007). *How the World's Best-performing School Systems Come Out on Top*. McKinsey and Company.

Barro, R.J. and J.-W. Lee (2010). 'A new data set of educational attainment in the world, 1950–2010', NBER Working Paper 15902, Cambridge, MA (April).

Bifulco, R. and H.F. Ladd (2006). 'The impacts of charter schools on student achievement: Evidence from North Carolina', *Education Finance and Policy*, 1(1), 50–90.

Bils, M. and P.J. Klenow (2000). 'Does schooling cause growth?' *American Economic Review*, 90(5), 1160–83.

Bishop, J.H. (1997). 'The effect of national standards and curriculum-based examinations on achievement', *American Economic Review*, 87(2), 260–64.

—— (2006). 'Drinking from the fountain of knowledge: Student incentive to study and learn: Externalities, information problems, and peer pressure', in E.A. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, North Holland, Amsterdam, pp. 909–44.

Björklund, A., P.-A. Edin, P. Freriksson and A.B. Krueger (2004). 'Education, equality and efficiency: An analysis of Swedish school reforms during the 1990s', IFAU Report 2004:1, Institute for Labour Market Policy Evaluation, Uppsala.

Booker, K., S.M. Gilpatric, T. Gronberg and D. Jansen (2007). 'The impact of charter school attendance on student performance', *Journal of Public Economics*, 91(5–6), 849–76.

Börsch-Supan, A. (2000). 'A model under siege: A case study of the German retirement insurance system', *Economic Journal*, 110(461), F24–F45.

Bradley, S. and J. Taylor (2002). 'The effect of the quasi-market on the efficiency-equity trade-off in the secondary school sector', *Bulletin of Economic Research*, 54(3), 295–314.

Burtless, G. (ed.) (1996). *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Brookings, Washington, DC.

Carnoy, M. and S. Loeb (2002). 'Does external accountability affect student outcomes? A cross-state analysis', *Educational Evaluation and Policy Analysis*, 24(4), 305–31.

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

Cingano, F., M. Leonardi, J. Messina and G. Pica (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), 117–63.

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

CREDO (2009). *Multiple Choice: Charter School Performance in 16 States*, Center for Research on Education Outcomes, Stanford University, Stanford, CA.

—— (2010). *Charter School Performance in New York City*, Center for Research on Education Outcomes, Stanford University, Stanford, CA.

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

Dee, T. and B. Jacob (2009). 'The impact of No Child Left Behind on student achievement', NBER Working Paper 15531, Cambridge, MA.

Dolton, P. and O.D. Marcenaro-Gutierrez (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.

Figlio, D.N. and C.E. Rouse (2006). 'Do accountability and voucher threats improve low-performing schools?' *Journal of Public Economics*, 90(1–2), 239–55.

Filer, R.K. and D. Münich (2003). 'Responses of private and public schools to voucher funding', paper presented at the Annual Meeting of the American Economic Association, Washington, DC.

Friedman, M. (1962). *Capitalism and Freedom*, University of Chicago Press, Chicago.

Fuchs, T. and L. Woessmann (2007). 'What accounts for international differences in student performance? A re-examination using PISA data', *Empirical Economics*, 32(2–3), 433–62.

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

Goldin, C. and L.F. Katz (2008). *The Race between Education and Technology*, Harvard University Press, Cambridge, MA.

Greenwald, R., L.V. Hedges and R.D. Laine (1996). 'The effect of school resources on student achievement', *Review of Educational Research*, 66(3), 361–96.

Hagist, C., N. Klusen, A. Plate and B. Raffelhüschen (2005). *'Social health insurance: The major driver of unsustainable fiscal policy?'* CESifo Working Papers 1574, CESifo, Munich.

Hanushek, E.A. (1971). 'Teacher characteristics and gains in student achievement: Estimation using micro data', *American Economic Review*, 60(2), 280–88.

—— (1992). 'The trade-off between child quantity and quality', *Journal of Political Economy*, 100(1), 84–117.

—— (1996). 'A more complete picture of school resource policies', *Review of Educational Research*, 66(3), 397–409.
—— (2002). 'Publicly provided education', in A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, Elsevier, Amsterdam, 2045–141.
—— (2003). 'The failure of input-based schooling policies', *Economic Journal*, 113(485), F64–F98.
—— (2007). 'The alchemy of "costing out" an adequate education', in M.R. West and P.E. Peterson (eds.), *School Money Trials: The Legal Pursuit of Educational Adequacy*, Brookings, Washington DC, pp. 77–101.
Hanushek, E.A., J.F. Kain, S.G. Rivkin and G.F. Branch (2007). 'Charter school quality and parental decision making with school choice', *Journal of Public Economics*, 91(5–6), 823–48.
Hanushek, E.A. and D.D. Kimko (2000). 'Schooling, labor force quality, and the growth of nations', *American Economic Review*, 90(5), 1184–208.
Hanushek, E.A. and A.A. Lindseth (2009). *Schoolhouses, Courthouses, and Statehouses: Solving the Funding-achievement Puzzle in America's Public Schools*, Princeton University Press, Princeton, NJ.
Hanushek, E.A., P.E. Peterson and L. Woessmann (2010). *U.S. Math Performance in Global Perspective: How Well Does each State Do at Producing High-achieving Students?* Program on Education Policy and Governance, Harvard University, Cambridge, MA.
Hanushek, E.A. and M.E. Raymond (2005). 'Does school accountability lead to improved student performance?' *Journal of Policy Analysis and Management*, 24(2), 297–327.
Hanushek, E.A. and S.G. Rivkin (2010). 'Generalizations about using value-added measures of teacher quality', *American Economic Review*, 100(2), 267–71.
Hanushek, E.A. and L. Woessmann (2008). 'The role of cognitive skills in economic development', *Journal of Economic Literature*, 46(3), 607–68.
—— (2009). 'Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation', NBER Working Paper 14633, Cambridge, MA.
—— (2010a). *The High Cost of Low Educational Performance: The Long-run Economic Impact of Improving PISA Outcomes*, Organization for Economic Cooperation and Development, Paris.
—— (2010b). 'How much do educational outcomes matter in OECD countries?' NBER Working Paper 16515, Cambridge, MA (November).
—— (2011). 'The economics of international differences in educational achievement', in E.A. Hanushek, S. Machin and L. Woessmann (eds.), *Handbook of the Economics of Education*, North Holland, Amsterdam, Vol. 3, pp. 89–200.
Heston, A., R. Summers and B. Aten (2002). 'Penn World Table Version 6.1', Center for International Comparisons at the University of Pennsylvania (CICUP), Philadelphia.
Howell, W.G. and P.E. Peterson (2002). *The Education Gap: Vouchers and Urban Schools*, Brookings, Washington, DC.
Hoxby, C.M. and S. Murarka (2009). 'Charter schools in New York City: Who enrolls and how they affect their students' achievement', NBER Working Paper 14852, Cambridge, MA (April).
Jacob, B.A. (2005). 'Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools', *Journal of Public Economics*, 89(5–6), 761–96.
Jürges, H., K. Schneider and F. Büchel (2005). 'The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany', *Journal of the European Economic Association*, 3(5), 1134–55.
Lavy, V. (2002). 'Evaluating the effect of teachers' group performance incentives on pupil achievement', *Journal of Political Economy*, 110(6), 1286–317.
—— (2009). 'Performance pay and teachers' effort, productivity, and grading ethics', *American Economic Review*, 99(5), 1979–2011.
Levačić, R. (2004). 'Competition and the performance of English secondary schools: Further evidence', *Education Economics*, 12(2), 177–93.
Levine, R. and D. Renelt (1992). 'A sensitivity analysis of cross-country growth regressions', *American Economic Review*, 82(4), 942–63.
Lucas Jr, R.E. (2003). 'Macroeconomic priorities', *American Economic Review*, 93(1), 1–14.
Mankiw, N.G., D. Romer and D. Weil (1992). 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics*, 107(2), 407–37.

Moore, M.A., A.E. Boardman, A.R. Vining, D.L. Weimer and D. Greenberg (2004). '"Just give me a number!" Practical values for the social discount rate', *Journal of Policy Analysis and Management*, 23(4), 689–812.

Mourshed, M., C. Chijioke and M. Barber (2010). *How the world's Most Improved School Systems Keep Getting Better*. McKinsey and Company.

Muralidharan, K. and V. Sundararaman (2009). 'Teacher performance pay: Experimental evidence from India', NBER Working Paper 15323, Cambridge, MA.

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

Nicoletti, G., S. Scarpetta and O. Boylaud (2000). 'Summary indicators of product market regulation with an extension to employment protection legislation', Economic Department Working Paper 226, OECD, Paris (February).

Nordhaus, W.D. (2007). 'A review of the Stern review on the economics of climate change', *Journal of Economic Literature*, 45(3), 686–702.

Organization for Economic Co-operation and Development (1999). *Employment Outlook*, OECD, Paris.

—— (2007). *PISA 2006: Science Competencies for Tomorrow's World*. Vol. 1 – Analysis, OECD, Paris.

—— (2008). *Compendium of Patent Statistics*, OECD, Paris.

—— (2009a). *OECD Economic Outlook, Volume 2009/1, No. 85*, OECD, Paris.

—— (2009b). *Society at a Glance 2009: OECD Social Indicators*, OECD, Paris.

—— (2010). *Education at a Glance 2010: OECD Indicators*, OECD, Paris.

PISA-Konsortium Deutschland (2008). *PISA 2006 in Deutschland: Die Kompetenzen der Jugendlichen im dritten Ländervergleich*, Waxmann, Münster.

Podgursky, M.J. and M.G. Springer (2007). 'Teacher performance pay: A review', *Journal of Policy Analysis and Management*, 26(4), 909–49.

Rajan, R. and L. Zingales (1998). 'Financial dependence and growth', *American Economic Review*, 88, 559–86.

Rivkin, S.G., E.A. Hanushek and J.F. Kain (2005). 'Teachers, schools, and academic achievement', *Econometrica*, 73(2), 417–58.

Rockoff, J.E. (2004). 'The impact of individual teachers on student achievement: Evidence from panel data', *American Economic Review*, 94(2), 247–52.

Rouse, C.E. (1998). 'Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program', *Quarterly Journal of Economics*, 113(2), 553–602.

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

—— (1997). 'Fundamental sources of long-run growth', *American Economic Review*, 87(2), 184–88.

Sala-i-Martin, X., G. Doppelhofer and R.I. Miller (2004). 'Determinants of long-term growth: A Bayesian Averaging of Classical Estimates (BACE) approach', *American Economic Review*, 94(4), 813–35.

Sandström, F.M. and F. Bergström (2005). 'School vouchers in practice: Competition will not hurt you', *Journal of Public Economics*, 89(2–3), 351–80.

Schivardi, F. and R. Torrini (2010). *Structural change and human capital in the Italian productive system, mimeo*, University of Cagliari.

Stern, N. (2007). *The Economics of Climate Change: The Stern Review*, Cambridge University Press, Cambridge.

Tol, R.S.J. and G.W. Yohe (2006). 'A review of the Stern review', *World Economics*, 7(4), 233–50.

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

Venn, D. (2009). 'Legislation, collective bargaining and enforcement: Updating the OECD employment protection indicators', OECD Social, Employment and Migration Working Paper 89, OECD, Paris.

West, M.R. and P.E. Peterson (2006). 'The efficacy of choice threats within school accountability systems: Results from legislatively-induced experiments', *Economic Journal*, 116(510), C46–C62.

West, M.R. and L. Woessmann (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. (2001). 'Why students in some countries do better: International evidence on the importance of education policy', *Education Matters*, 1(2), 67–74.
—— (2003). 'Schooling resources, educational institutions, and student performance: The international evidence', *Oxford Bulletin of Economics and Statistics*, 65(2), 117–70.
—— (2005a). 'Educational production in Europe', *Economic Policy*, 20(43), 446–504.
—— (2005b). 'The effect heterogeneity of central exams: Evidence from TIMSS, TIMSS-Repeat and PISA', *Education Economics*, 13(2), 143–69.
—— (2007a). 'International evidence on expenditure and class size: A review', in *Brookings Papers on Education Policy 2006/2007*, Brookings, Washington DC, pp. 245–72.
—— (2007b). 'International evidence on school competition, autonomy and accountability: A review', *Peabody Journal of Education*, 82(2–3), 473–97.
—— (2009). 'Public-private partnerships and student achievement: A cross-country analysis', in R. Chakrabarti and P.E. Peterson (eds.), *School choice international: Exploring public-private partnerships*, MIT Press, Cambridge, MA, pp. 13–45.
—— (2010). 'Institutional determinants of school efficiency and equity: German states as a microcosm for OECD countries', *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics*, 230(2), 234–70.
—— (2011). 'Cross-country evidence on teacher performance pay", *Economics of Education Review*, 30(3), 404–18.
Woessmann, L., E. Luedemann, G. Schuetz and M.R. West (2009). *School Accountability, Autonomy, and Choice around the World*, Edward Elgar, Cheltenham.
Wolf, P., B. Gutmann, M. Puma, B. Kisida, L. Rizzo and N. Eissa (2010). *Evaluation of the DC Opportunity Scholarship Program: Impacts after Three Years*, Institute for Education Sciences, Washington, DC (March).
Wölfl, A., I. Wanner, T. Kozluk and G. Nicoletti (2009). *'Ten years of product market reform in OECD countries: Insights from a revised PMR indicator'*, OECD Economics Department Working Paper 695, OECD, Paris.