CREDIT Research Paper

No. 08/04

International School Test Scores and Economic Growth

by

Simon Appleton, Paul Atherton and Michael Bleaney

Abstract

We expand Hanushek and Kimko's (2000) analysis of the relationship between schooling quality, as measured by scores in international tests, and growth. We take account of another fifteen years of growth and approximately twice as many test score results. We treat the data first as a panel, relating growth only to test scores at earlier dates, and then as a cross-section. In both cases we find the effect of schooling quality on growth to be statistically significant but substantially smaller than that reported by Hanushek and Kimko (2000) and Hanushek and Woessmann (2007).

JEL Classification: O4

Keywords: Growth, human capital, education

Centre for Research in Economic Development and International Trade, University of Nottingham

CREDIT Research Paper

No. 08/04

International School Test Scores and Economic Growth

by

Simon Appleton, Paul Atherton and Michael Bleaney

Outline

The Authors

Paul Atherton is a Research Student, Simon Appleton is an Assistant Professor in Development Economics, and Michael Bleaney is Professor in Economics in the School of Economics, University of Nottingham.

Corresponding author: Paul Atherton, lexpa1@nottingham.ac.uk.

Research Papers at www.nottingham.ac.uk/economics/credit/

1. INTRODUCTION

Human capital is central to theoretical and empirical models of economic growth. Typically empirical analysis proxies human capital by measures of time spent in education. However, since pupils may acquire differing amounts of skill and knowledge from a given period in education, the concept of human capital is perhaps better captured by direct measures of knowledge acquired. This has been attempted through internationally standardized tests of student performance, which reveal significant cross-country differences in achievement. This paper focuses on whether these test scores appear to affect economic growth.

Hanushek and Kimko (2000) provide evidence that average country scores in these tests have strong effects on growth. They use data from six international tests of student performance in mathematics and science. In their basic cross-section regression for average per capita GDP growth 1960-90 for 31 countries, Hanushek and Kimko (2000, Table 2, p.1190) [hereafter HK] find these test scores to be extremely significant, and to perform better than measures of the quantity of schooling. Moreover the estimated effects of educational quality are strikingly large: a one-standard deviation increase in educational quality is estimated to increase per capita growth by 1.4 per cent per annum. The exercise is extended to fifty countries and forty years of growth (1960-2000) by Hanushek and Woessmann (2007) [hereafter HW], who obtain an even larger estimate of 2.0 percentage points of additional per capita growth from an improvement in educational quality by one standard deviation.

A key data limitation is that international tests of school performance are infrequent, with few countries involved in the earlier part of the 1960-90 period over which growth was measured. In response to this, both HK and HW aggregate the performance of each country in all tests entered, at whatever date, to generate a single measure of "labor force quality" over the growth period. However, because many countries only participated in later years, in many cases this procedure implies using test performance at a later date to explain economic growth over an earlier time interval. For example, HK use Mozambique's tests scores from 1991 to explain the country's growth in the period 1960-90. HK justify this procedure by assuming that schooling quality changes slowly over time. However, one may wonder whether test scores in 1991 might better be regarded as at least in part a consequence of economic growth in the period 1960-90 rather than a cause. This suspicion is heightened by the consideration that the test scores relate to students still in school and not the whole workforce. In a direct causal sense, student test scores can only have a lagged effect on economic growth.

The HK approach to measuring test scores contrasts with the now-standard procedure for the *number of years* of schooling, where demographic information is combined with past data on years of schooling to produce an average estimate for the whole labor force, rather than just new entrants (Barro and Lee, 1993). Lack of a sufficient run of data prevents such an approach for test scores, but a good first step is to relate test scores to growth only at a strictly *later* date, when the people whose scores are being measured are likely to have entered the labour force. This is the innovation of this paper. Using this procedure, we find that the estimated effect of test scores on five-year average growth rates at various lags is statistically significant but substantially weaker than that reported by HK and HW. The estimated effects are of similar magnitude to theirs only when we relate test scores to growth at an *earlier* date, a result which has no obvious explanation. Cross-section results using the same procedure as HK and HW, but with the latest version of Penn World Tables, confirm our findings.

The paper proceeds as follows: Section 2 reviews the existing literature on empirical growth studies that incorporate human capital. Section 3 presents results that relate test scores to growth at a strictly later date, treating the data as a panel. Section 4 presents some new cross-section results, and Section 5 concludes.

2. MEASURES OF THE QUALITY OF EDUCATION IN GROWTH STUDIES: A REVIEW OF THE LITERATURE

2.1. Test scores as measures of human capital

Direct international comparisons of schooling systems have been made possible through a series of international tests which have been conducted since 1963 under the auspices of multiple bodies, most notably the International Association for the Evaluation of Educational Achievement (IEA) and the International Assessment of Educational Progress (IAEP)[1]. These tests provide the basis for qualitative education variables to proxy for human capital.

Hanushek and Kimko (2000) use these to construct a measure of labour force quality. They draw from six studies, conducted in 1963/64, 1970, 1982, 1985, 1988 and 1991, which consist of twenty-six performance series, all with differing ages (of participants), component tests and from different years. Each cohort of test scores is multiplicatively transformed to an international mean of 50 and then averaged by country to create a country-specific measure of

[1] Other tests exist, such as the OECD programme for international student assessment (PISA), the International Adult Literacy Survey (IALS) and the Adult Literacy and Life skills (ALLS).

labour force quality at all dates, denoted QL1. Additional variables in the baseline regression, which uses a sample of 31 countries, are initial income, population growth and the Barro-Lee measure of quantity of schooling.

HK find that the quality measures have positive and significant effects on growth, although the magnitude is rather high (a ten-point increase in test scores is associated with 1.34 percentage points of additional growth). These findings are robust to the inclusion of additional variables, following Levine and Renelt (1992). Hanushek and Woessmann (2007) use a similar procedure, but they incorporate test scores from more recent years, including the Third International Maths and Science study (TIMSS) conducted in 1995 and 1999 and the Programme for International Student Assessment (PISA) studies conducted in 2000 and 2003. This enlarges the sample to 50 countries, and they estimate that a ten-point increase in average test scores is associated with a 1.55 percentage point improvement in growth rates, which is even larger than HK's original estimate of 1.34. Barro (2001), who treats each subject area (maths, science and reading) as separate, also finds that quality of schooling is more important for growth than quantity.

3. THE MODEL

Hanushek and Kimko (2000) estimate the following model:

$$1)\quad G_i = \alpha + \beta_1 n_i + \beta_2 \log(GDP_i) + \beta_3 S_i + \beta_4 T_i + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

where G_i denotes annual percentage growth in GDP per capita, measured at purchasing power parity, for country *i* over the period 1960-90, n_i denotes the average population growth rate over the sample interval, GDP_i denotes the level of GDP per capita in 1960, S_i denotes the average years of schooling of the labour force, T_i is the measure of labour force quality derived from the test scores and ε_i is an error term. Data on years of schooling (the average years of schooling of the total population aged 15 or above) are taken from the Barro and Lee dataset (1993). The growth data that we use in replicating this regression are drawn from the latest Penn World Tables (6.1 and 6.2), enabling the period to be expanded to incorporate data up to 2004. Population growth is an average of all possible data points in the period, rather than an average of the various subsets of years. HW estimate a similar model using growth data from 1960 to 2000.

Few countries participated in the international tests before 1980, so cross-sectional growth regressions using data back to 1960, as in HK and HW, tend to be dominated by growth information that chronologically precedes test score information. To address this issue we treat the data as a panel and estimate

$$\text{2)} \quad \begin{aligned} G_{it} &= \alpha_i + \beta_1 n_{it} + \beta_2 \log\left(GDP_{it}\right) + \beta_3 S_{it} + \beta_4 T_{it-x} + \varepsilon_{it} \\ \varepsilon_{it} &\sim \mathrm{N}(0, \sigma^2) \end{aligned}$$

where variables are defined as in equation (1) but allowed to vary over time, t, which we measure in five year periods. We consider a variety of lags of test scores, with x varying from 1 to 3 (corresponding to lags of 5 to 15 years). To gain insight into the differences between our results and HK's, we also consider non-positive values of *x*, ranging from -2 (where the growth period commenced 10 years before the testing year) to zero (where the growth period commences in the testing year).

The longer the lag from test scores to growth, the smaller the number of test scores that can be used in the regression. The 2003 rounds are excluded from all regressions, the 1999 rounds are lost when the lag is expanded to 10 years and the 1995 rounds are lost when the lag is expanded to 15 years. Similarly, the 1963 rounds are lost when the lag is expanded to -10 years. Since the model compares growth directly from different periods, unobserved time-specific factors may be influencing both the inputs and outputs of the growth process. To control for this, time dummies are in all regression specifications.

A further econometric issue arises over possible unobserved country-specific factors - the α_i in equation (2). These can be purged from the estimates by using fixed effects panel data estimates, removing some potential omitted variables bias[2]. However, this estimation comes at a serious cost - it means that we cannot use the information on long-run cross-country differences in the quality of schooling to gauge possible effects on growth. Instead, we will be reliant on temporal variations in school quality between tests. These will aggravate potential measurement error - such as that which arises from problems in comparability between the different waves of tests. Consequently, we report results from both pooled cross-sections (which constrain $\alpha_i = \alpha$) and from fixed-effects estimates.

[2] A third possibility would be modelling country specific terms as random effects. However, Hausman tests rejected the consistency of random effects for all lag-specifications. As such, random effects estimates are not reported.

Pooled cross sectional estimates are presented in Table 1, with the impact of test scores on growth evaluated after various lag lengths. Columns (1) shows the impact when it is assumed to be contemporaneous. Columns (2)-(4) impose a lagged effect. Columns (5) and (6) relate test scores to *past* growth rates (included only for illustration of the effect of such misspecification).

In the pooled results, test scores are significant at the 0.05 level for all lags except +15. The estimated effect of a ten-point improvement in average test scores is to increase per capita growth by 0.86 percentage points five years later and by a very similar amount (0.76 percentage points) ten years later. These numbers are considerably smaller than those estimated by HK (1.34 percentage points) and HW (1.55 percentage points). According to Table 1, the estimated effect of test scores on growth is much larger for negative values of *x*.

Table 1: The impact of lagged test scores on growth - pooled results

Dependent Variable: Average GDP per capita growth rate for five year period						
Lag between testing year and commencement of growth period						
	Contemporaneous	*Lag*			*Lead*	
	(1) 0 Years	(2) 5 years	(3) 10 years	(4) 15 years	(5) -10 years	(6) -5 years
Log(Initial GDP level)	-0.530 (0.216)**	-0.442 (0.308)	-0.391 (0.252)	-0.093 (0.209)	-0.903 (0.276)***	-0.513 (0.283)*
Population Growth	-0.298 (0.180)	0.703 (0.316)**	0.769 (0.306)**	-0.518 (0.244)**	-0.065 (0.233)	0.019 (0.205)
Average Years of Schooling	0.041 (0.0975)	-0.304 (0.143)**	-0.226 (0.134)*	0.101 (0.096)	-0.021 (0.132)	-0.149 (0.132)
Test Scores	0.056 (0.0246)**	0.086 (0.0360)**	0.076 (0.0333)**	0.025 (0.0246)	0.184 (0.0329)***	0.138 (0.033)***
RMSE	1.613	2.071	1.738	1.064	2.065	2.079
Number of Observations	126	95	71	56	117	116
Number of Countries	41	38	29	24	41	40
Standard errors in parentheses (significant at 0.01***, 0.05**, 0.10*)						

There is no obvious explanation for this result, but it is likely to account for the higher coefficients obtained by HK and HW in cross-section regressions.

Allowing for country-specific fixed effects, as in Table 2, produces contrasting results: test scores are not significantly related to growth in any specification except that with a fifteen-year lag. However, the magnitude of the impact of test scores with a fifteen-year lag is comparable to the numbers obtained from the pooled cross-sections. The estimated effect of a ten-point improvement in average test scores is to increase per capita growth by 0.72 percentage points fifteen years later.

Table 2: The impact of lagged test scores on growth

Dependent Variable: Average GDP per capita growth rate for five year period

Lag between testing year and commencement of growth period				
	Contemporaneous	*Lag*		
	(1) 0 Years	(2) 5 years	(3) 10 years	(4) 15 years
Log(Initial GDP level)	-6.064 (1.233)***	-14.22 (2.307)***	2.823 (4.556)	-4.32 (2.926)
Population Growth	0.229 (0.340)	0.416 (0.723)	2.212 (1.249)*	3.437 (1.239)
Average Years of Schooling	0.313 (0.286)	-0.301 (0.408)	-0.469 (0.530)	0.329 (0.316)
Test Scores	-2.242 (3.286)	-0.0111 (0.04438)	0.0695 (0.0524)	0.07167 (0.0299)**
RMSE	1.123	1.367	1.424	0.736
Number of Observations	126	95	71	56
Number of Countries	41	38	29	24

As previously discussed, there are pros and cons to using fixed effects estimates over pooled cross sectional ones. However, a case can be made for regarding the fixed effects estimates with a fifteen year lag as the preferred specification. As the tests are taken at age 13, after the lag the participants would be aged 28, a more plausible age than 18 or 23 when evaluating the impact on growth rates. Furthermore, there are strong arguments to suggest country specific effects will be non-zero and correlated with the variables in the pooled specification, biasing the results.

4. SOME NEW CROSS-SECTION ESTIMATES

In this section we re-estimate the cross-section regression of HK over a longer data period, both with and without updated test score data. The results are presented in Table 3. The top half of Table 3 uses data up to 2000, as in Hanushek and Woessmann (2007). In the first column only HK's original 31 countries are included; in the second column the sample is expanded to 42 countries using the expanded TIMSS data; and in the third column the data set is expanded further to 49 countries, using all possible data. The results for test scores are similar to those reported by HK, with an additional average test score of 1 point adding between 0.12 and 0.14 percentage points to growth.

In the bottom half of Table 3, we repeat the exercise using PPP-adjusted growth data up to 2004. The test score coefficient is much smaller, and close to that obtained in Table 1, with an additional average test score of one point adding only about 0.05 percentage points to growth. It is clear that it is the addition of more recent growth data, rather than more recent test score results, that is responsible for the difference in results.

Table 3: Aggregate measure of labour force quality

Dependent Variable:	**PPP-adjusted average growth rate per capita 1960-00**		
Labour force measure	QL1 with TIMSS	Expanded TIMSS	All Tests, Including Reading
Initial income level	-1.863 (0.604)***	-1.781 (0.421)***	-1.249 (0.389)***
Population growth	-0.632 (0.380)	-0.474 (0.269)*	-0.657 (0.274)**
Average years of schooling	0.3572 (0.270)	0.299 (0.187)	0.151 (0.172)
test scores	0.136 (0.049)**	0.119 (0.031)***	0.125 (0.030)***
N	31	42	49
R^2	0.471	0.554	0.315
	PPP-adjusted average growth rate per capita 1960-04		
Labour force measure	QL1 with TIMSS	Expanded TIMSS	All Tests, Including Reading
Initial income level	-1.923 (0.406)***	-1.957 (0.331)***	-1.264 (0.279)***
Population growth	-0.869 (0.256)***	-0.760 (0.224)***	-0.765 (0.234)***
Average years of schooling	0.355 (0.181)*	0.322 (0.148)*	0.094 (0.072)*
test scores	0.057 (0.033)*	0.044 (0.024)*	0.048 (0.025)*
N	31	42	49
R^2	0.574	0.519	0.356
Standard errors in parentheses (significant at 0.01***, 0.05**, 0.10*)			

5. CONCLUSIONS

Existing literature relating test scores to growth may over-estimate any positive impacts because of the strong correlation between test scores and growth rates at *earlier* dates, which is probably spurious. When growth is measured strictly subsequent to test scores, the estimated effect diminishes to about half that reported by Hanushek and Kimko (2000). The results vary depending on whether country-specific effects are included. Without such effects, there is a significant positive impact on growth if the relationship is contemporaneous or has a lag of five or ten years. However, a fifteen-year lag is arguably more plausible and gives a significant effect only when country fixed effects are allowed for.

Although it is possible that the smaller coefficients could be caused by measurement error, which is larger when information from all tests for each country are not amalgamated into one figure and when fixed-effects estimates are used, they are consistent with cross-section results from the latest data, using the same methodology as HK.

6. APPENDIX

6.1. Summary of Tests

Country	IEA Maths	IEA Science	IEA Reading	2nd IEA Maths	First IAEP	2nd IEA Science	2nd IEA Reading	Second IAEP	TIMSS 1995	TIMSS 1999	PIRLS	TIMSS 2003	Pisa 2003
Argentina											42		
Armenia												46.95	
Australia	27	30.75				59.33			53.7	53.25		51.6	52.4
Austria									54.85				50.6
Bahrain												41.95	
Belgium	43.4	22.88	74	57.5			49.4		52.8	54.65		52.65	52.9
Belize											32.7		
Botswana							33					36.5	
Brazil								42.13					35.6
Bulgaria									55.25	51.45	55	47.75	
Canada				56.25	51.44	62	51.1	65.4	52.9	53.2	54.4	53.2	53.2
Chile			61							40.6		40	
Colombia									39.8		42.2		
Cyprus							48.9		46.8	46.8	49.4	45	
Czech Republic									56.9	52.95	53.7		51.6
Denmark							50		49				51.4
Egypt												41.35	
Estonia												54.15	
Finland	37.71	25.63	74	45.5		61.67	56.45			52.75			54.4
France	30			57.7			54	66.4		51.8	52.5		51.1
Ghana												26.55	

Greece							50.65		49.04		52.4		44.5
Hong Kong				55.1			52.6		55.5	55.6	52.8	57.1	55
Hungary		36.38	70	56.8		72.33	51.75	70.9	54.55	54.2	54.3	53.6	49
Iceland							52.7		49.05		51.2		51.5
Indonesia							39.4			41.9		41.55	36
Iran			39						44.9	43.5	41.4	43.2	
Ireland					48.68		51	61.9	53.25				50.3
Israel	46.14			49.9				66.4	52.3	46.7	50.9	49.2	
Italy			65			55.67	52.2	66.95	48.6	48.75	54.1	48.75	46.6
Japan		39		60.3		67.33			58.8	56.45		56.1	53.4
Jordan								48.5		43.9		44.9	
Korea					55.89	60.33		75.45	58.6	56.8		57.35	54.2
Kuwait									41.11		39.6		
Latvia									48.9	50.4	54.5	51	48.3
Lebanon												41.3	
Lithuania									47.65	48.5	54.3	51.05	
Luxemburg				45.4									49.3
Macau													52.7
Macedonia										45.25	44.2	44.2	
Malawi			34										
Malaysia										50.5		50.9	
Mexico													38.5
Moldova										46.4	49.2	46.6	
Morocco										33	35	39.15	
Netherlands	30.57	22.25	69	59.3		66	49.95		55.05	54.25	55.4	53.6	53.8
New Zealand		30.25		45.6			53.65		51.65	50.05	52.9	50.7	52.3
Nigeria				40.8			40.1						
Norway						59.67	52		51.5		49.9	47.74	49.5

Palestine												41.25	
Philippines						38.33	43			34.5		37.75	
Poland						60.33							49
Portugal							50.05	55.45	46.7				46.6
Romania									48.4	47.2	51.2	47.25	
Russia									53.65	52.75	52.8	51.1	46.8
Saudi Arabia												36.5	
Serbia												47.2	43.7
Singapore						55	52.45		62.5	58.6	52.8	59.1	
Slovakia									54.55	53.45	51.8	51.25	49.8
Slovenia									55.05	53.15	50.2	50.65	
South Africa									34	25.9		25.4	
Spain					50.78		49.7	61.45	50.2			48.8	48.5
Swaziland				32.3									
Sweden	21.86	27.13	72	40.6		61.33	54.25		52.7		56.1	51.15	50.9
Switzerland							52.35	72.25	53.35				52.7
Taiwan										57.7		57.8	
Thailand						55	47.7		52.35	47.45		41.7	
Trinidad and Tobago							46.5						
Tunisia										43.9		40.7	35.9
Turkey										43.1	44.9		42.3
UK	32.93	26.6925	71	49.2	51.47	27.835		64.45	51.85	51.7	54.05	51.3	
Uruguay					47.62			61.15					42.2
US	25.43	27	67	51.4		55	54.1		51.7	50.85	54.2	51.55	48.3
Venezuela							40						
Zimbabwe							37.2						

REFERENCES

Barro, R. J., and J.-W. Lee (1993): "International Comparisons of Educational Attainment.," *Journal of Monetary Economics*, 32, 363-394.

— (2001): "International Data on Educational Attainment: Updates and Implications.," *Oxford Economic Papers*, 53.

Hanushek, E. A. (1995). "Interpreting recent research on schooling in developing countries" *The World Bank Research Observer* **10**(2): 227.

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

Hanushek, E., and L. Woessman (2007): "The Role of Education Quality for Economic Growth," *World Bank Working Paper* 4122.

Summers, R., and A. Heston (1991): "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988," *The Quarterly Journal of Economics*, 106, 327-368.