German Economic Review 6(3): 269–286

The Economics of School Quality

Eric A. Hanushek
Stanford University, National Bureau of Economic Research and University of Texas at Dallas

Abstract. *Recent national concerns about student performance on PISA tests and similar assessments appear warranted because of the direct relationship between student cognitive skills and both individual earnings and national growth. The evidence on the impacts of school quality indicates very large economic effects. Available research shows, however, that improving school quality is difficult and not closely related to spending levels. One approach supported by research is improving teacher quality, which can yield sizable gains in student performance. Instituting appropriate policies to alter the teacher force will necessarily take significant time.*

JEL classification: *I*2, *J*2, *O*4.

Keywords: Economic growth; school quality; individual productivity; teacher quality; school reform; student achievement.

1. INTRODUCTION

No country in the world has been as attuned to issues of educational quality as Germany has over the past few years. The national reaction to the performance of German 15-year-olds on the PISA tests of 2000 went far beyond that seen in other countries. Economic analysis suggests that the attention given to school performance in Germany is warranted. School quality has powerful effects on individual productivity and on national growth. But further analysis suggests that achieving improvements in the quality dimension will be difficult and will take long-term commitments to change.

This article brings together economic analyses of school quality to provide a picture both of the benefits to quality improvements and of the potential costs of them. Much of the early and continuing development of empirical work on human capital concentrates on the role of school attainment, that is, the quantity of schooling. Quantity of schooling is easily measured, and data on years attained, both over time and across individuals, are readily available.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

Nonetheless, today's policy concerns in most corners of the world revolve much more around issues of quality than issues of quantity.

It is somewhat common in policy debates simply to note that economic analysis shows that schooling is very important (based on the returns to more years of schooling) and then turn to policies that are designed to improve quality of schooling. This approach, however, gives no guidance to how to compare the costs and benefits of any program of school quality improvement. Indeed, it might be very misleading if the returns to quality are very different than the returns to quantity of schooling.

One of the challenges in understanding the impact of school quality differences in human capital has been simply knowing how to measure quality. Much of the discussion of quality – in part related to new efforts to provide better accountability – has identified cognitive skills as the important dimension. And, clearly, much of the concern over PISA results and other international test scores implicitly assumes this is an appropriate measure. The question here is whether this proxy for school quality – students' performance on standardized tests – is correlated with economic outcomes including individuals' performance in the labor market and the economy's ability to grow. Until recently, little comprehensive data have been available to show any relationship between differences in cognitive skills and any related economic outcomes. Such data are now becoming available.

It is important to understand what the evidence on quality says and to use this in any policy discussions. Two very important conclusions emerge. First, improving school quality can lead to dramatic economic gains – ones that could in the aggregate pay for all of national expenditure on schooling after a generation. But, second, realizing these gains will require more patience and commitment than usually present in the political system.

2. IMPACTS OF QUALITY ON ECONOMIC GROWTH

Economic growth determines how much improvement will occur in the overall standard of living of society. Moreover, the education of each individual has the possibility of making others better off (in addition to the individual benefits discussed below). Specifically, a more educated society may lead to higher rates of invention; may make everybody more productive through the ability of firms to introduce new and better production methods; and may lead to more rapid introduction of new technologies. These externalities provide extra reason for being concerned about schooling.

Economists have developed a variety of models and ideas to explain differences in growth rates across countries – invariably featuring the importance of human capital (see Barro and Sala-i-Martin, 2003). The empirical work supporting growth analyses has emphasized school attainment differences across countries. Again, this is natural because, while compiling comparable data on many things for different countries is difficult,

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

assessing quantity of schooling is more straightforward. The typical study finds that quantity of schooling is highly related to economic growth rates. But, quantity of schooling is a very crude measure of the knowledge and cognitive skills of people – particularly in an international context.[1]

Hanushek and Kimko (2000) go beyond simple quantity of schooling and delve into quality of schooling.[2] We develop a consistent measure of labor force quality based on information about international differences in mathematics and science knowledge. The basic statistical models relate annual growth rates of GDP per capita (g_c) to our measure of labor force quality (T_c), the initial level of income (Y^0), the quantity of schooling (S_c), and a vector of other control variables (Z_c, which includes in different specifications the population growth rates, political measures, openness of the economies, and the like):

$$g_c = \alpha_0 + \eta T_c + \alpha_1 Y_c^0 + \alpha_2 S_c + Z_c \phi + v_c \quad (1)$$

We combine all of the international test scores available through 1991 into a single composite measure of quality and consider statistical models that explain differences in growth rates across nations during the period 1960 to 1990.[3] In 1963 and 1964, the International Association for the Evaluation of Educational Achievement (IEA) administered the first of a series of mathematics tests to a voluntary group of countries. Five subsequent testings, sponsored by the IEA and others including the OECD, assessed both math and science and expanded on the group of countries that have been tested. There is some movement in student performance across time of country

1. Critics of the existing research built on school attainment or school enrollment rates have raised several questions. First, even though there might be a correlation between growth and school attainment, it may not be a causal relationship. Growing countries may use a portion of their wealth to buy more schooling, leading to the observed correlation (Bils and Klenow, 2000). Second, the estimated impacts are sensitive to the precise specification of the underlying statistical analyses, and it is very difficult with the available data to distinguish among alternative estimates (Levine and Renelt, 1992; Levine and Zervos, 1993; Benhabib and Spiegel, 1994). Finally, the estimates of schooling on growth are significantly different than what would be expected from the micro relationship between individual earnings and schooling – possibly reflecting the misuse of education in socially unproductive ways (Pritchett, 2001; Easterly, 2002). Other authors have shown, however, that a number of the research anomalies disappear when measurement issues are dealt with (Topel, 1999; Krueger and Lindahl, 2001; Soto, 2002). Importantly, these authors do not even directly address what is perhaps the most important measurement issue – and the subject of this article. Variations in cognitive skills and measured quality show that the knowledge at a given level of schooling completion in some countries has virtually nothing in common with that in other countries.
2. Barro and Lee (2001) provide an analysis of qualitative differences that also includes literacy.
3. We exclude the three TIMSS tests from 1995–2003 and the PISA tests from 2000–03 because they were taken outside of the analytical period on economic growth. We combine all test measures available for each country over the 1965–91 period into a single measure for each country. (Note that the United States and the United Kingdom are the only countries to participate in all of the testing.)

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

performance on the tests, but for the one country that can be checked – the United States – the pattern is consistent with other US data.[4]

Estimation across countries with test information explains a substantial portion of the variation in economic growth across countries. Most important, the quality of the labor force (T_c) as measured by math and science scores is extremely important, both in statistical terms and in magnitude. One standard deviation difference on test performance is related to 1 per cent difference in annual growth rates of GDP per capita.[5] Because the added growth compounds, it leads to powerful effects on national income and on societal well-being.

The subsequent considerations of economic benefits to quality improvement employ these growth estimates. Before discussing the magnitude of these effects, it is useful to consider their interpretation and credibility.

One common concern in analyses such as this is that schooling might not be the actual cause of growth but, in fact, may just reflect other attributes of the economy that are beneficial to growth. For example, the East Asian countries consistently score very highly on the international tests, and they also had extraordinarily high growth over the 1960–90 period. It may be that other aspects of these East Asian economies have driven their growth and that the statistical analysis of labor force quality simply is picking out these countries. But in fact, even if the East Asian countries are excluded from the analysis, a strong – albeit slightly smaller – relationship is still observed with test performance. This test of sensitivity of the results seems to reflect a basic importance of school quality, a factor that contributes also to the observed growth of East Asian countries.

Another concern might be that other factors that affect growth, such as efficient market organizations, are also associated with efficient and productive schools – so that, again, the test measures are really a proxy for other attributes of the country. In order to investigate this, we concentrate on immigrants to the United States who received their education in their home countries. We find that immigrants who were schooled in countries that have higher scores on the international math and science examinations earn more in the United States. This analysis makes allowance for any differences in school attainment, labor market experience or being native English-language speakers. In other words, skill differences as measured by the international tests are clearly rewarded in the United States labor market, reinforcing the validity of the tests as a measure of individual skills and productivity.

4. The National Assessment of Educational Progress (NAEP) in the United States is designed to follow performance of US students for different subjects and ages. NAEP performance over this period exhibits a sizable dip in the 1970s, a period of growth in the 1980s, and a leveling off in the 1990s – exactly the pattern of US scores on the various international tests given over the same time period.
5. The details of this work can be found in Hanushek and Kimko (2000) and Hanushek (2003b). Importantly, the effect school quantity is greatly reduced by including quality, but adding other factors potentially related to growth (including aspects of international trade, private and public investment, and political instability) leaves the effects of labor force quality unchanged.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

Finally, the observed relationships could simply reflect reverse causality, that is, that countries that are growing rapidly have the resources necessary to improve their schools and that better student performance is the result of growth, not the cause of growth. This issue is potentially important when policy alternatives are considered below, but, as a simple test of this, we investigated whether the international math and science test scores were systematically related to the resources devoted to the schools in the years prior to the tests. They were not. If anything, we found relatively better performance in those countries spending less on their schools.

One final issue needs consideration: the United States has never done well on these international assessments, yet its growth rate has been very high for a long period of time. The reconciliation is that quality of the labor force is just one aspect of the economy that enters into the determination of growth. A variety of factors clearly contribute, and these factors work to overcome any deficits in quality. These other factors may also be necessary for growth. In other words, simply providing more or higher-quality schooling may yield little in the way of economic growth in the absence of other elements, such as the appropriate market, legal and governmental institutions to support a functioning modern economy. Past experiences investing in less developed countries that lack these institutional features demonstrates that schooling is not itself a sufficient engine of growth.[6]

Three other factors immediately come to mind as being important in US growth and as potentially masking to detrimental effects of low school quality. First, almost certainly the most important factor sustaining the growth of the US economy is the openness and fluidity of its markets. The United States maintains generally freer labor and product markets than most countries in the world. The government generally has less regulation on firms (both in terms of labor regulations and in terms of overall production), and trade unions are less extensive than those in many other countries. Even broader, the United States has less intrusion of government in the operation of the economy – not only less regulation but also lower tax rates and minimal government production through nationalized industries. These factors encourage investment, permit the rapid development of new products and activities by firms, and allow US workers to adjust to new opportunities. While identifying the precise importance of these factors is difficult, a variety of analyses suggest that such market differences could be very important explanations for differences in growth rates.[7]

6. Indeed, some have questioned the precise role of schooling in growth. Easterly (2002), for example, notes that education without other facilitating factors such as functioning institutions for markets and legal systems may not have much impact. He argues that World Bank investments in schooling for less developed countries that do not ensure that the other attributes of modern economies are in place have been quite unproductive. Similarly, Pritchett (2001) worries about socially unproductive uses of schooling.
7. See, for example, Krueger (1974), World Bank (1993) and Parente and Prescott (1994, 1999).

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

Second, over the twentieth century, the expansion of the education system in the United States outpaced that around the world. The United States pushed to open secondary schools to all citizens. With this came also a move to expand higher education with the development of land grant universities, the GI bill, and direct grants and loans to students. In comparison with other nations of the world, the US labor force has been better educated, even after allowing for the lesser achievement of its graduates. In other words, more schooling with less learning each year has yielded more human capital than found in other nations that have less schooling but learn more in each of those years. (This advantage has clearly ended as many OECD countries have expanded schools to exceed that found in the United States; Organisation for Economic Co-operation and Development, 2003.)

Finally, the analysis of growth rates across countries emphasizes quality of the elementary and secondary schools of the United States. It did not include any measures of the quality of US colleges. By most evaluations, US colleges and universities rank at the very top in the world. A number of the economic models of economic growth in fact emphasize the importance of scientists and engineers as a key ingredient to growth. By these views, the technically trained college students who contribute to invention and to development of new products provide a special element to the growth equation. Here, again, the United States appears to have the best programs. If this view is correct, US higher education may continue to provide a noticeable advantage over other countries.

This detail on the modeling is important, because the results figure into the analysis of school quality. The summary statement at this point is that the relationship between math and science skills on the one hand and productivity and growth on the other comes through clearly when investigated in a systematic manner across countries. The magnitude of the effect underscores the importance of high-quality schooling.

3. THE BENEFITS OF QUALITY IMPROVEMENT

The impact of improved quality can be directly calculated from the considerations of how quality affects growth rates for economies. Two dimensions of quality improvement are important: how much improvement in quality is considered and how fast is any change to be realized.

Most policy discussions do not pay much attention to the speed of change. But, if student achievement is to be raised by improving the schools, it will take time to introduce the school programs, it will take time for students to gain the new knowledge and skills, and it will take time for them to enter the labor market and have their impact felt. To start, we can simply look at the effects under different assumptions about the speed of reform – 10 years, 20 years or 30 years to achieve the desired improvements in student performance. (The discussion below provides ways of assessing the realism of differing speeds of reform.)

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

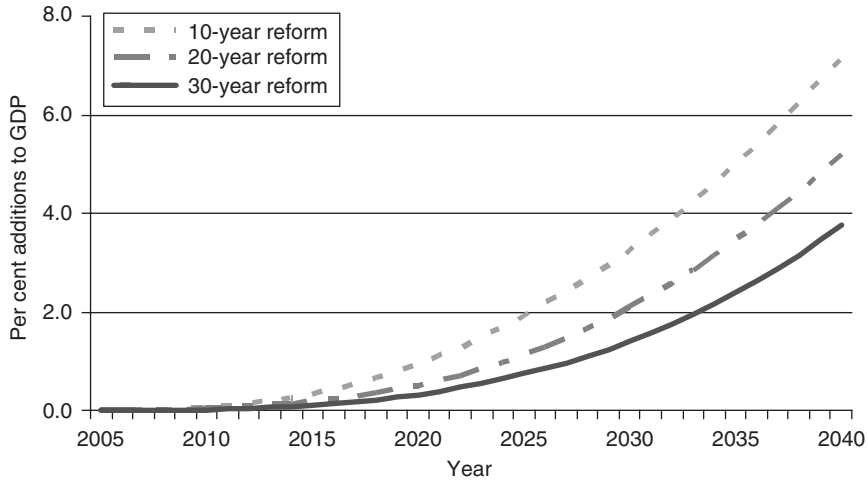

Figure 1 Improved GDP with moderately strong knowledge improvement (0.5 S.D.)

For our purposes, it is also important to be clear about the magnitude of improvement that is being considered. As a benchmark, we set a 'moderately strong' reform at a 0.5 standard deviation improvement on the international math and science tests that are used to index quality in the growth models. In terms of the PISA scores in 2003, this moderately strong reform would move Germany from the middle of the OECD distribution of math scores to around the level of Belgium or Canada. This movement would not be the top of the distribution but would be substantial.

Figure 1 illustrates the impact that reform could be expected to have over time if it is successful at achieving such moderately strong knowledge improvement. The curves sketch out the path of GDP improvement that would occur with a reform plan that reaches its improvement goal within 10, 20 or 30 years.[8]

Consider just the slow improvement of schools over a 30-year period. In 2040, the GDP would be almost 4 per cent higher than projected without the schooling reforms. This magnitude would cover total school spending in many countries of the world. Of course, faster reforms would yield even greater gains in GDP.

8. These calculations simulate a linear improvement across 12 grades in school that is sufficient to obtain a 0.5 S.D. improvement in the given reform period. After the reform period, school quality remains constant at the new higher level. Students enter the labor force each year with the appropriate 'reform quality' of schooling, and they age over time so that the average labor force quality gradually changes.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

4. IMPACTS OF QUALITY ON INDIVIDUAL PRODUCTIVITY AND INCOMES

School quality also has direct implications for the productivity and earnings of individuals.[9] Recent analyses underscore this impact of quality.

Most attention has been, and still is, directed at the returns to years of school or school attainment. Beginning with Mincer (1970, 1974), economists have employed readily available census data to do estimate what is now simply referred to as a 'Mincer equation':

$$\ln(y_i) = a_0 + \rho S_i + a_1 \exp + a_2 \exp^2 + X_i\beta + \varepsilon_i \tag{2}$$

where y_i is earnings, S_i is years of schooling, $\exp_i$ is labor market (or potential) experience, X_i is a vector of other individual attributes, and ε_i is an error term. The object of attention, ρ, is interpreted as the rate of return to a year of schooling, and this has been estimated for a very large number of countries around the world (see Psacharopoulos, 1994).[10]

A variety of researchers, however, have investigated how quality enters, and they document that the earnings advantages to higher achievement on standardized tests are quite substantial.[11] These results are derived from different specific approaches, but the basic underlying analysis involves estimating a standard 'Mincer' earnings function and adding a measure of individual cognitive skills:

$$\ln(y_i) = a_0 + \rho S_i + \gamma T_i + a_1 \exp + a_2 \exp^2 + X_i\beta + \varepsilon_i \tag{3}$$

where T_i is the individual's measured cognitive skill and γ is the return to quality.

While these various analyses emphasize different aspects of individual earnings, they typically find that measured achievement has a clear impact on earnings after allowing for differences in the quantity of schooling, the

9. By implication, the distribution of schooling outcomes also enters into the distribution of income in society, although those effects are not considered here.
10. There has been some controversy over exactly how to estimate the rate of return to school attainment. The main issue has revolved around whether or not a causal interpretation can be given to ρ. The argument has been that higher-ability students are more likely to continue in schooling. Therefore, part of the higher earnings observed for those with additional schooling really reflects pay for added ability and not for the additional schooling. Early discussion of ability bias can be found in Griliches (1974). Economists have pursued a variety of analytical approaches for dealing with this. The approaches have included looking for circumstances where the amount of schooling is affected by things other than the student's valuation of continuing and considering the income differences among twins (see Card, 1999). The various adjustments for ability differences typically make small differences on the estimates of the value of schooling, and Heckman and Vytlacil (2001) argue that it is not possible to separate the effects of ability and schooling.
11. A separate line of research investigates expenditure and resource differences across schools, but these measures are known to be poor measures of school quality differences (Hanushek, 2002).

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

experiences of workers, and other factors that might also influence earnings.[12] In other words, higher quality as measured by tests similar to those currently being used in accountability systems around the country is closely related to individual productivity and earnings.

The magnitude of the effects is easiest to see from three recent US studies that provide direct and quite consistent estimates of the impact of test performance on earnings (Mulligan, 1999; Murnane *et al.*, 2000; Lazear, 2003). These studies employ different nationally representative datasets that follow students after they leave schooling and enter the labor force. When scores are standardized, they suggest that one standard deviation increase in mathematics performance at the end of high schools translates into 12 per cent higher annual earnings.

Murnane *et al.* (2000) provide evidence from the High School and Beyond and the National Longitudinal Survey of the High School Class of 1972. Their estimates suggest some variation with males obtaining a 15 per cent increase and females a 10 per cent increase per standard deviation of test performance. Lazear (2003), relying on a somewhat younger sample from the National Education Longitudinal Study of 1988 (NELS88), provides a single estimate of 12 per cent. These estimates are also very close to those in Mulligan (1999), who finds 11 per cent for the normalized AFQT score in the National Longitudinal Survey of Youth (NLSY) data. By way of comparison, estimates of the value of an additional year of school attainment are typically 7–10 per cent.

These estimates are very significant: a person one-half standard deviation above the current mean performance on these cognitive tests can expect to earn 6 per cent per year more than the average person each and every year of the working life.

There are reasons to believe that these estimates provide a lower bound on the impact of higher achievement. First, these estimates are obtained fairly early in the work career (mid-20s to early 30s), and other analysis suggests that the impact of test performance becomes larger with experience.[13] Second, the labor market experiences that are observed begin in the mid-1980s and extend into the mid-1990s, but other evidence suggests that the value of skills and of schooling has grown throughout and past that period. Third, future general improvements in productivity are likely to lead to larger returns to skill.[14]

12. The clearest analyses are found in the following references (which are analyzed in Hanushek, 2002). See Bishop (1989, 1991), O'Neill (1990), Blackburn and Neumark (1993, 1995), Grogger and Eide (1993), Murnane *et al.* (1995, 2000, 2001), Neal and Johnson (1996), Mulligan (1999), Altonji and Pierret (2001) and Lazear (2003).
13. Altonji and Pierret (2001) find that the impact of achievement grows with experience, because the employer has a chance to observe the performance of workers.
14. These analyses typically compare workers of different ages at one point in time to obtain an estimate of how earnings will change for any individual. If, however, productivity improvements occur in the economy, these will tend to raise the earnings of individuals

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

A limited number of additional studies are available for developed countries outside of the United States. McIntosh and Vignoles (2001) study wages in the United Kingdom and find strong returns to both numeracy and literacy.[15] Finnie and Meng (2002) and Green and Riddell (2003) investigate returns to cognitive skills in Canada. Both suggest that literacy has a significant return, but Finnie and Meng (2002) find an insignificant return to numeracy. This latter finding stands at odds with most other analyses that have emphasized numeracy or math skills. Even without a broader empirical base, it seems plausible to believe that the kinds of returns to quality found in the US are similar to those existing in other countries.

Without pursuing it in any detail, it is important to note that these returns to higher skills also appear to hold across developing countries. The literature on returns to cognitive skills in developing countries is restricted to a relatively limited number of countries: Ghana, Kenya, Morocco, Pakistan, South Africa and Tanzania.[16] There are reasons for caution in interpreting the precise magnitude of estimates. First, the estimates appear to be quite sensitive to the estimation methodology itself. Both within individual studies and across studies using the same basic data, the results are quite sensitive to the techniques employed in uncovering the fundamental parameter for cognitive skills.[17] Second, the evidence on variations within developing countries is not entirely clear. For example, Jolliffe (1998) finds little impact of skills on farm income, while Behrman *et al*. (forthcoming 2005) suggest an equivalence across sectors at least on theoretical grounds. Nonetheless, the overall summary is that the available estimates of the impact of cognitive skills on outcomes suggest strong economic returns within developing countries.

Quantity of schooling also has a direct effect in the estimates of individual earnings, implying that part of the full return to school quality can come through continuation in school.[18] There is substantial US evidence that students who do better in school, either through grades or scores on standardized achievement tests, tend to go farther in school. Murnane *et al*. (2000) is perhaps most interesting because of its linkage to the earning effects

over time. Thus, the impact of improvements in student skills are likely to rise over the work life instead of being constant as portrayed here.

15. Because they look at discrete levels of skills, it is difficult to compare the quantitative magnitudes directly to the US work.
16. Moreover, a number of studies actually employ the same basic data, albeit with different analytical approaches, but come up with somewhat different results. The available studies are: Boissiere *et al*. (1985), Knight and Sabot (1990), Glewwe (1996), Angrist and Lavy (1997), Jolliffe (1998), Moll (1998), Vijverberg (1999), Behrman *et al*. (forthcoming 2005).
17. The sensitivity to estimation approach is not always the case; see, for example, Jolliffe (1998). A critique and interpretation of the alternative approaches within a number of these studies can be found in Glewwe (2002).
18. This work has not, however, investigated how achievement affects the ultimate outcomes of additional schooling. For example, if over time higher-achieving students tend increasingly to attend further schooling, these schools may not be forced to offer as many remedial courses.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

of quality. They separate the direct returns to measured skill from the indirect returns of more schooling (through alternative specifications of equation (3)) and suggest that perhaps one-third to one-half of the full return to higher achievement comes from further schooling.[19]

As with the estimation of the returns to the quantity of schooling, there are potential concerns about the causal interpretation of the quality estimates. The largest issue, which will be addressed below, is the impact of schools versus other inputs such as families. The measured achievement employed in the earnings models clearly represents a variety of influences.

5. REFORM PROGRAMS

Policy-makers around the world have taken up the pursuit of improved school quality. Sometimes it is based on concerns about the observed performance on assessments – ones like PISA that provide direct information on relative performance. Sometimes it is based simply on their instincts or on the political popularity of discussing school quality issues.

One important feature, however, pervades much of the existing reform discussion. Reform policies have generally been expensive, but they have not led to widespread improvements in student performance. The United States experience in seeking school quality improvement probably is not representative. Nonetheless, after truly substantial increases in the resources to school, measured performance went up only slightly between 1970 and 2000.[20]

The existing evidence suggests that common improvement strategies center on such things as increasing teacher qualifications or reducing class size do not have a powerful effect on student outcomes (see discussion in Hanushek, 2003a). Although clearly controversial, the past analyses of resource policies do not indicate that continuation of these as a very hopeful way to achieve student performance increases.

19. Other relevant work, including Rivkin (1995), finds that variations in test scores capture a considerable proportion of the systematic variation in high school completion and in college continuation, so that test score differences can fully explain black–white differences in schooling. Bishop (1991) and Hanushek *et al*. (1996), in considering the factors that influence school attainment, find that individual achievement scores are highly correlated with continued school attendance. Neal and Johnson (1996) in part use the impact of achievement differences of blacks and whites on school attainment to explain racial differences in incomes. Their point estimates of the impact of cognitive skills (AFQT) on earnings and school attendance appear to be roughly comparable to that found in Murnane *et al*. (2000). Behrman *et al*. (1998) find strong achievement effects on both continuation into college and quality of college; moreover, the effects are larger when proper account is taken of the various determinants of achievement. Hanushek and Pace (1995) find that college completion is significantly related to higher test scores at the end of high school.
20. Between 1960 and 2000, real spending per pupil in the United States more than tripled. These increases are compared to scores of 17-year-olds on the National Assessment of Educational Progress.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

One possible explanation for past failure, supported by research into the determination of achievement, is that insufficient attention has been given to teacher quality. By many accounts, the quality of teachers is the key to student performance. But the research evidence suggests that many of the policies that have been pursued have not been very productive. Specifically, while the policies may have led to changes in measured aspects of teachers, they have not improved the quality of teachers when identified by student performance.[21]

Output-based measures of teacher quality are rather new, and they are non-existent for research outside of the United States. Thus, this part of the analysis is based entirely on the US teacher market and the quality distribution found there.

Rivkin *et al*. (2005) describe estimates of differences in teacher quality on an output basis. Specifically, the concern is identifying good and bad teachers on the basis of their performance in obtaining gains in student achievement. An important element of that work is distinguishing the effects of teachers from the selection of schools by teachers and students and the matching of teachers and students in the classroom. In particular, highly motivated parents search out schools that they think are good, and they attempt to place their children in classrooms where they think the teacher is particularly able. Teachers follow a similar selection process (Hanushek *et al*., 2004). Thus, from an analytical viewpoint, it is difficult to sort out the quality of the teacher from the quality of the students that she has in her classroom. The analysis of teacher performance goes to great lengths to avoid contamination from any such selection and matching of kids and teachers.[22] In the end, we estimate the distribution of teacher quality: one standard deviation in teacher quality is at least 0.11 standard deviations of annual growth in student achievement.

It is useful to put this estimate of the variation in quality into perspective. This estimate implies that having a 'good' teacher (one standard deviation of quality above the mean) as compared to the mean teacher quality would lead the average student to move up over four percentile points in the year. If a student had a good teacher as opposed to an average teacher for five years in a row, the increased learning would be sufficient to close entirely the average gap between a typical low-income student and a higher-income student (i.e. one not on free or reduced lunch).

21. For a review of existing literature, see Hanushek and Rivkin (2004). This paper describes various attempts to estimate the impact of teacher quality on student achievement.
22. The analysis uses annual test score data on all public school students in the State of Texas. Several entire cohorts of students (over 200,000 per cohort) are tracked across time. To deal with selection, it concentrates on differences among teachers within a given school in order to avoid the potential impact of parental choices of schools. Moreover, it employs a strategy that compares grade-level performance across different cohorts of students, so that the matching of students to specific teachers in a grade can be circumvented. As such, it is very much a lower-bound estimate on differences in teacher quality.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

In another attempt to estimate the variation in teacher quality, we analyze variations across classrooms within a large school district in Texas (Hanushek *et al.*, 2005). In this, we match individual teachers and students and look at achievement-based quality measures of each teacher compared to all of the teachers in the district or, alternatively, all of the other teachers in each school. On a basis comparable to the prior estimates, we obtain an estimate of teacher quality between 0.15 and 0.18 standard deviations of student achievement.

These estimates of the importance of teacher quality permit calculations of what would be required to yield the reform results discussed earlier. Specifically, consider what kinds of teacher policies might yield a 0.5 standard deviation improvement in student performance. Obviously an infinite number of alternative hiring plans could be used to arrive at any given end point. A particularly simple plan is employed here to illustrate what is required.

Consider a steady improvement plan where the average new hire is chosen at a constant amount better than the average existing teacher in each year.[23] For example, the average teacher in the current distribution is found at the 50th percentile. Consider a policy where the average of the new teachers hired is set at the 56th percentile and where future hires continue to be at this percentile each year of the reform period. (Note that, because better teachers are being hired each year, the mean rises. Thus, always hiring at the 56th percentile means that the average new hire in each year is better than the average in the previous year.)

What kinds of teacher replacement policies would be feasible? To illustrate an overall reform policy, the calculations simply assume that replacement rates are the same as found in the United States. In 1994–95 school year, 6.6 per cent of all teachers left teaching; 13.8 per cent of all teachers either left teaching or moved to a new school.

By maintaining the lower replacement rate of all teachers exiting teaching (6.6 per cent annually) but retaining all other teachers each year, a replacement policy set at the 56th percentile would yield a 0.5 standard deviation improvement in student performance after a 20-year period. If instead we thought of applying these new standards to all teacher turnover (exits plus the 7.2 per cent who change schools), a 0.5 S.D. improvement in student performance could be achieved in 10 years.

Figure 2 displays the annual hiring improvement that is necessary to achieve a 0.5 standard deviation improvement under a 10-, 20- and 30-year reform plan and based on applying it to either just those exiting or the higher turnover rates that include transfers. As is obvious, the stringency of the new hiring is greater when there is a shorter reform period and when fewer new (higher quality) teachers are brought in each year. Achieving a 0.5 S.D. boost in achievement in

23. Because the previous estimates ignore teacher quality differences between schools in the central city and suburban areas, these calculations use an estimate of the standard deviation of teacher quality of 0.22 S.D. of student performance.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

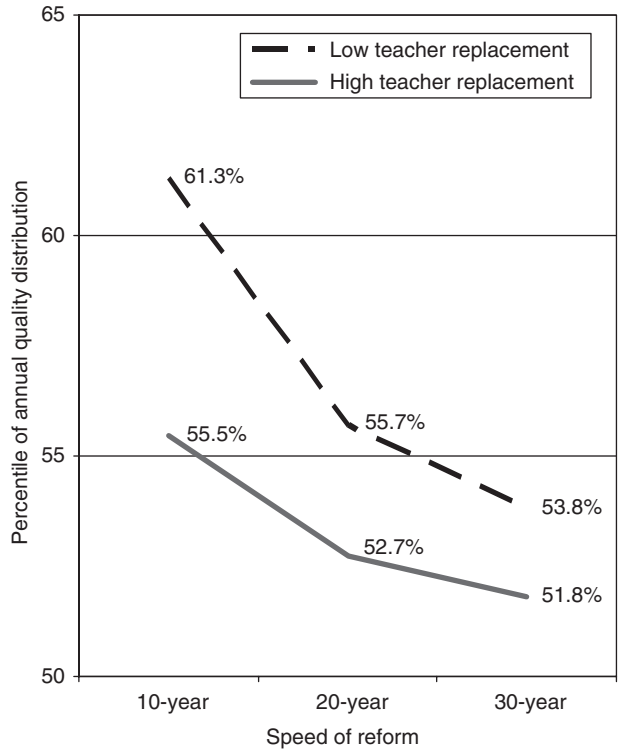

Figure 2 Annual required hiring percentile for moderately strong improvement in student achievement

10 years by upgrading just those who exit each year implies hiring at the 61st percentile, but this declines to the 52nd percentile for a 30-year plan where the higher turnover population is subject to these new hiring standards.

These calculations are meant to illustrate two points. First, existing research into student achievement and teacher quality shows that teachers have significant leverage on performance. By implication, if better teachers can be hired and retained, significant changes in student achievement can be obtained. Second, without dramatic changes in policies about teacher retention, feasible reform will take a quite long period of time. Specifically, unless larger numbers of current teachers are fired and replaced, changing the character of the teaching force takes time.

6. CONCLUSIONS

The previous analysis shows that a country can get substantial gains in student performance if it can adopt policies that will improve the teaching force. But, a clear message is that patience and persistence are required.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

If policies are aimed at the quality of teachers, one cannot simply eliminate all current teachers and start over. Instead, if the changes come through selective hiring and retention policies whenever there are replacements, it takes a considerable length of time before the teaching force is likely to be different. And then it takes time for students to experience these differences.

This contrasts with policies that simply aim at lifting the quantity of schooling completed. Through just expanding the system (adding more schooling opportunities on top of the old ones) and inducing current students to stay in school longer, the average attainment can be increased fairly rapidly. The changes in completion of upper secondary schools that have been accomplished over the past three decades in OECD countries illustrate the rates of increase possible when talking about quantity of schooling. On average, contrasting the completion rates for upper secondary schooling of 25–34-year-olds with 55–64-year-olds, countries increased the completion rate by 0.8 percentage points per year for 30 years (Organisation for Economic Co-operation and Development, 2003).

Policies aimed at student achievement and school quality cannot feasibly obtain such rapid changes. This means that policy-makers will have to stay with policies for a very long period of time – a time during which economic effects will be difficult to discern.

Nonetheless, it is worth it. Individuals will see higher productivity and earnings. And national income will grow substantially. As illustrated in Figure 1, as an increasing number of new, higher-quality graduates are found in the labor market, GDP growth can be expected to lead to sizable gains. Under reasonable scenarios, most countries would be able to pay for *all the primary and secondary school spending* when the policies have had time to play out.

The underlying empirical analysis into student achievement allows for direct inputs of families into education along with indirect impacts through the selection of a school and possibly a teacher within a school. Thus, it reinforces the basic point that the observed scores of students – either within a country or across countries – only partly reflects the value-added of schools. Nonetheless, from a policy viewpoint, the analysis assumes that effective policies to change schooling through the families are difficult and that the only policy option is to deal through the schools.

This paper has also not dealt with the question of how the quality of the teaching force can be improved. A variety of alternative incentives schemes have been suggested, although we have little experience with any of them (Hanushek and Rivkin, 2004).[24] Nonetheless, the magnitude of gains suggests that a wide range of substantial incentive schemes could be introduced with the gains from an improved economy.

24. Additionally, teacher-training institutions will undoubtedly have to change if the overall quality of the teaching force is to increase by the substantial amounts discussed.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

ADDRESS FOR CORRESPONDENCE

Hoover Institution, Stanford University, Stanford, CA 94305-6010, USA.
E-mail: hanushek@stanford.edu

REFERENCES

Altonji, J. G. and C. R. Pierret (2001), 'Employer Learning and Statistical Discrimination', *Quarterly Journal of Economics* 116(1), 313–350.

Angrist, J. D. and V. Lavy (1997), 'The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco', *Journal of Labor Economics* 15, S48–S76.

Barro, R. J. and J.-W. Lee (2001), 'International Data on Educational Attainment: Updates and Implications', *Oxford Economic Papers* 53(3), 541–563.

Barro, R. J. and X. Sala-i-Martin (2003), *Economic Growth*, 2nd edn, McGraw-Hill, New York.

Behrman, J. R., L. G. Kletzer, M. S. McPherson and M. O. Schapiro (1998), 'The Microeconomics of College Choice, Careers, and Wages: Measuring the Impact of Higher Education', *Annals of the American Academy of Political and Social Science* 559 (September), 12–23.

Behrman, J. R., D. Ross and R. Sabot (2005), 'Improving the Quality versus Increasing the Quantity of Schooling: Evidence for Rural Pakistan', forthcoming in *Topics in the Economics and Growth of Developing Areas of the B.E. Journals of the Economics and Growth of Developing Areas*.

Benhabib, J. and M. M. Spiegel (1994), 'The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data', *Journal of Monetary Economics* 34(3), 143–174.

Bils, M. and P. J. Klenow (2000), 'Does Schooling Cause Growth?', *American Economic Review* 90(5), 1160–1183.

Bishop, J. (1989), 'Is the Test Score Decline Responsible for the Productivity Growth Decline?', *American Economic Review* 79(1), 178–197.

Bishop, J. (1991), 'Achievement, Test Scores, and Relative Wages', in: M. H. Kosters (ed.), *Workers and their Wages*, AEI Press, Washington, DC, pp. 146–186.

Blackburn, M. L. and D. Neumark (1993), 'Omitted-Ability Bias and the Increase in the Return to Schooling', *Journal of Labor Economics* 11(3), 521–544.

Blackburn, M. L. and D. Neumark (1995), 'Are OLS Estimates of the Return to Schooling Biased Downward? Another Look', *Review of Economics and Statistics* 77(2), 217–230.

Boissiere, M. X., J. B. Knight and R. H. Sabot (1985), 'Earnings, Schooling, Ability, and Cognitive Skills', *American Economic Review* 75(5), 1016–1030.

Card, D. (1999), 'Causal Effect of Education on Earnings', in: O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, North-Holland, Amsterdam, pp. 1801–1863.

Easterly, W. (2002), *The Elusive Quest for Growth: An Economist's Adventures and Misadventures in the Tropics*, MIT Press, Cambridge, MA.

Finnie, R. and R. Meng (2002), 'Minorities, Cognitive Skills, and Incomes of Canadians', *Canadian Public Policy* 28, 257–273.

Glewwe, P. (1996), 'The Relevance of Standard Estimates of Rates of Return to Schooling for Educational Policy: A Critical Assessment', *Journal of Development Economics* 51, 267–290.

 © Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

Glewwe, P. (2002), 'Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes', *Journal of Economic Literature* 40(2), 436–482.
Green, D. A. and W. C. Riddell (2003), 'Literacy and Earnings: An Investigation of the Interaction of Cognitve and Unobserved Skills in Earnings Generation', *Labour Economics* 10, 165–184.
Griliches, Z. (1974), 'Errors in Variables and Other Unobservables', *Econometrica* 42(6), 971–998.
Grogger, J. T. and E. Eide (1993), 'Changes in College Skills and the Rise in the College Wage Premium', *Journal of Human Resources* 30(2), 280–310.
Hanushek, E. A. (2002), 'Publicly Provided Education', in: A. J. Auerbach and M. Feldstein (eds), *Handbook of Public Economics*, Elsevier, Amsterdam, pp. 2045–2141.
Hanushek, E. A. (2003a), 'The Failure of Input-Based Schooling Policies', *Economic Journal* 113(485), F64–F98.
Hanushek, E. A. (2003b), 'The Importance of School Quality', in: P. E. Peterson (ed.), *Our Schools and Our Future: Are We Still at Risk?*, Hoover Institution Press, Stanford, CA, pp. 141–173.
Hanushek, E. A. and D. D. Kimko (2000), 'Schooling, Labor Force Quality, and the Growth of Nations', *American Economic Review* 90(5), 1184–1208.
Hanushek, E. A. and R. R. Pace (1995), 'Who Chooses to Teach (and Why)?', *Economics of Education Review* 14(2), 101–117.
Hanushek, E. A. and S. G. Rivkin (2004), 'How to Improve the Supply of High Quality Teachers', in: D. Ravitch (ed.), *Brookings Papers on Education Policy 2004*, Brookings Institution Press, Washington, DC, pp. 7–25.
Hanushek, E. A., J. F. Kain, D. M. O'Brien and S. G. Rivkin (2005), 'The Market for Teacher Quality', Working Paper No. 11154, National Bureau of Economic Research (February).
Hanushek, E. A., J. F. Kain and S. G. Rivkin (2004), 'Why Public Schools Lose Teachers', *Journal of Human Resources* 39(2), 326–354.
Hanushek, E. A., S. G. Rivkin and L. L. Taylor (1996), 'Aggregation and the Estimated Effects of School Resources', *Review of Economics and Statistics* 78(4), 611–627.
Heckman, J. J. and E. Vytlacil (2001), 'Identifying the Role of Cognitive Ability in Explaining the Level of and Change in the Return to Schooling', *Review of Economics and Statistics* 83(1), 1–12.
Jolliffe, D. (1998), 'Skills, Schooling, and Household Income in Ghana', *World Bank Economic Review* 12, 81–104.
Knight, J. B. and R. H. Sabot (1990), *Education, Productivity, and Inequality*, Oxford University Press, New York.
Krueger, A. B. and M. Lindahl (2001), 'Education for Growth: Why and For Whom?', *Journal of Economic Literature* 39(4), 1101–1136.
Krueger, A. O. (1974), 'The Political Economy of the Rent Seeking Society', *American Economic Review* 64(3), 291–303.
Lazear, E. P. (2003), 'Teacher Incentives', *Swedish Economic Policy Review* 10(3), 179–214.
Levine, R. and D. Renelt (1992), 'A Sensitivity Analysis of Cross-Country Growth Regressions', *American Economic Review* 82(4), 942–963.
Levine, R. and S. J. Zervos (1993), 'What have we Learned about Policy and Growth from Cross-Country Regressions?', *American Economic Review* 83(2), 426–430.
McIntosh, S. and A. Vignoles (2001), 'Measuring and Assessing the Impact of Basic Skills on Labor Market Outcomes', *Oxford Economic Papers* 53, 453–481.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005

Mincer, J. (1970), 'The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach', *Journal of Economic Literature* 8(1), 1–26.
Mincer, J. (1974), *Schooling Experience and Earnings*, NBER, New York.
Moll, P. G. (1998), 'Primary Schooling, Cognitive Skills, and Wage in South Africa', *Economica* 65, 263–284.
Mulligan, C. B. (1999), 'Galton versus the Human Capital Approach to Inheritance', *Journal of Political Economy* 107(6), S184–S224.
Murnane, R. J., J. B. Willett, M. J. Braatz and Y. Duhaldeborde (2001), 'Do Different Dimensions of Male High School Students' Skills Predict Labor Market Success a Decade Later? Evidence from the NLSY', *Economics of Education Review* 20(4), 311–320.
Murnane, R. J., J. B. Willett, Y. Duhaldeborde and J. H. Tyler (2000), 'How Important are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?', *Journal of Policy Analysis and Management* 19(4), 547–568.
Murnane, R. J., J. B. Willett and F. Levy (1995), 'The Growing Importance of Cognitive Skills in Wage Determination', *Review of Economics and Statistics* 77(2), 251–266.
Neal, D. A. and W. R. Johnson (1996), 'The Role of Pre-Market Factors in Black–White Differences', *Journal of Political Economy* 104(5), 869–895.
O'Neill, J. (1990), 'The Role of Human Capital in Earnings Differences between Black and White Men', *Journal of Economic Perspectives* 4(4), 25–46.
Organisation for Economic Co-operation and Development (2003), *Education at a Glance: OECD Indicators 2003*, Organisation for Economic Co-operation and Development, Paris.
Parente, S. L. and E. C. Prescott (1994), 'Barriers to Technology Adoption and Development', *Journal of Political Economy* 102(2), 298–321.
Parente, S. L. and E. C. Prescott (1999), 'Monopoly Rights: A Barrier to Riches', *American Economic Review* 89(5), 1216–1233.
Pritchett, L. (2001), 'Where has All the Education Gone?', *World Bank Economic Review* 15(3), 367–391.
Psacharopoulos, G. (1994), 'Returns to Investment in Education: A Global Update', *World Development* 22, 1325–1344.
Rivkin, S. G. (1995), 'Black/White Differences in Schooling and Employment', *Journal of Human Resources* 30(4), 826–852.
Rivkin, S. G., E. A. Hanushek and J. F. Kain (2005), 'Teachers, Schools, and Academic Achievement', *Econometrica* 73(2), 417–458.
Soto, M. (2002), 'Rediscovering Education in Growth Regressions', Working Paper No. 202, OECD Development Centre (November).
Topel, R. (1999), 'Labor Markets and Economic Growth', in: O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Elsevier, Amsterdam, pp. 2943–2984.
Vijverberg, W. P. M. (1999), 'The Impact of Schooling and Cognitive Skills on Income from Non-Farm Self-Employment', in: P. Glewwe (ed.), *The Economics of School Quality Investments in Developing Countries: An Empirical Study of Ghana*, St Martin's Press (with University of Oxford), New York.
World Bank (1993), *The East Asian Miracle: Economic Growth and Public Policy*, Oxford University Press, New York.

© Verein für Socialpolitik and Blackwell Publishing Ltd. 2005