Available online at www.sciencedirect.com
SCIENCE DIRECT®
Economics of Education Review 25 (2006) 447–462
Economics of Education Review
www.elsevier.com/locate/econedurev

Alternative school policies and the benefits of general cognitive skills

Eric A. Hanushek*

Hoover Institution, Stanford University, Stanford, CA 94305, USA

Received 29 July 2004; accepted 31 August 2005

Abstract

School policy debates vacillate between policies emphasizing improvement in general skills and those aimed at strengthening the linkages between schools and the workplace. While these policies do not necessarily conflict, each is frequently motivated by shortcomings in the other. This paper presents basic evidence about the very substantial impacts of general cognitive skills on individual earnings and on economic growth. The calculations are then put in a school policy framework that emphasizes the importance of considering both the magnitude and the speed of quality improvements. It then considers alternative school reform policies focused on improvements in teacher quality, identifying how much change is required. Finally, teacher bonus policies are put into the context of potential benefits.

JEL classification: I2; J3; O4

Keywords: Economic impact; Rate of return; Efficiency; Human capital; Economic development

1. Introduction

Although not entirely appropriate, many school policy discussions contrast programs aimed at improving general cognitive skills with those aimed at improving the transition from school to work. The underlying theme, particularly from those arguing for improved school-workplace linkages, is that general skills are not obviously relevant for the workplace and that schools could improve the careers of students by emphasizing more relevant skills. The arguments are made most strongly for students who will not proceed to colleges or universities but will instead directly enter the job market. This paper does not directly consider the contrast between work-related and job-related programs but instead provides a benchmark of the economic value general cognitive skills.[1] Analysis of the benefits and costs of school reform indicates investments that improve the quality of schools offer exceptional rewards to society. Thus, any consideration of focused job-to-work programs

*Tel.: +1 650 736 0942; fax: +1 650 723 1687.
E-mail address: hanushek@stanford.edu.

[1]One perspective in the vocational education literature is that alternative ways of teaching basic skills that emphasize practical problems and that are motivated by workplace ideas are more effective for some students than programs that use less direct motivation of abstract concepts. This paper does not directly consider this evidence, but it does underscore the economic value of such instruction if it truly enhances student cognitive skills.

0272-7757/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2004.08.011

must take into account potential costs that might arise from lowered general skills.

The findings about the importance of standard measures of school quality are particularly relevant in the context of US accountability policies that emphasize performance on standardized tests in core areas. Some people have suggested that the achievement emphasized by current state accountability systems is not very important and that other aspects of student performance—creativity, the ability to work in teams, specific vocational knowledge, or personality traits—should be the focus of attention. While these other aspects are undoubtedly valuable, the analysis here strongly affirms an emphasis on basic cognitive skills by demonstrating their substantial economic returns.

This analysis highlights what is known about the economic benefits of investments in school quality as identified by general cognitive skills. Two decades ago, the federal government released a report, *A Nation at Risk* (National Commission on Excellence in Education, 1983), that identified some serious problems with school quality as found in measured student performance.[2] While it precipitated an unbroken period of concern about US schools, it did not lead to any substantial improvements in school quality (Peterson, 2003). This paper provides evidence that the economic motivations for improving schools are real and continuing.

From the discussion of the importance of student outcomes, it then attempts to provide some bounds on the economics of school reform. In providing a general economic analysis of school policies, it is clear that the benefits of reform are generally easier to estimate than the costs, although some information on costs is provided at the end.

The central messages are: first, the economic impact of reforms that enhance student achievement will be very large; second, reform must be thought of in terms of both the *magnitude* of changes and the *speed* with which any changes occur. Third, based on current knowledge, the most productive reforms are almost certainly ones that improve the quality of the teacher force. Fourth, such policies are likely to be ones that improve the hiring, retention, and pay of high quality teachers, i.e. selective policies aimed at the desired outcome.

[2] *A Nation at Risk* relied on flawed analysis of the economic impacts of school quality, leading some critics to attack the logic of the case rather than to consider improving the schools.

Finally, it is important to emphasize that this is a narrow discussion because it only considers a series of direct market outcomes. Other outcomes would enhance the benefits and will be pointed out along the way—even valuing these is not currently possible. Underestimating the benefits is nonetheless not overly important given the central message that investments that truly improve student cognitive outcomes are likely to be so valuable.

2. Benefits of enhanced school quality

Economists have devoted considerable attention to understanding how human capital affects a variety of economic outcomes, but much of the discussion has been quite general. The underlying notion is that individuals make investment decisions in themselves through schooling and other routes. The accumulated labor market skills from these investments over time represent an important component of the human capital of an individual. But, here, economic analyses are frequently imprecise about the exact skills or the nature of their measurement.

It is commonly presumed that formal schooling is an important contributor to the skills of an individual and to human capital, but not the only factor. Parents, individual abilities, and friends undoubtedly contribute. Schools nonetheless have a special place because they are most directly affected by public policies. Thus, we frequently emphasize the role of schools.

The human capital perspective immediately makes it evident that the real issues are ones of long-run outcomes. Future incomes of individuals are related to their past investments. It is not their income while in school or their income in their first job. Instead, it is their income over the course of their working life.

The distribution of income in the economy similarly involves both the mixture of people in the economy and the pattern of their incomes over their lifetime. Specifically, most measures of how income and well-being vary in the population do not take into account the fact that some of the low-income people have low incomes only because they are just beginning a career. Their lifetime income is likely to be much larger as they age, gain experience, and move up in their firms and career. What is important is that any noticeable effects of the current quality of schooling on the distribution of skills and income will only be realized years in the

future, when those currently in school become a significant part of the labor force. In other words, most workers in the economy were educated years and even decades in the past—and they are the ones that have the most impact on current levels of productivity and growth, if for no reason other than that they represent the larger share of active workers.

Much of the early and continuing development of empirical work on human capital concentrates on the overall role of school attainment, that is, the quantity of schooling. The revolution in the United States during the twentieth century was universal schooling. Moreover, quantity of schooling is easily measured, and data on years attained, both over time and across individuals, are readily available.

Today, however, policy concerns revolve much more around issues of quality than issues of quantity. The US completion rates for high school and college have been roughly constant for a quarter of a century. Meanwhile, the standards movement in schools has focused on what students know as they progress through schools. This trend is substantially reinforced by federal accountability legislation (the No Child Left Behind Act of 2001), which emphasizes student proficiency in basic subjects as measured by standardized achievement tests.[3]

2.1. Impacts of quality on individual incomes

One of the challenges in understanding the impact of 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, while there is ongoing debate about the testing and measurement of these skills, a debate that interacts with discussions of school-to-work programs, most parents and policy makers alike accept the notion that cognitive skills are a key dimension of schooling outcomes. The question is whether this proxy for school quality—students' performance on standardized tests—is correlated with 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.

Much of the work by economists on differences in worker skills has actually been directed at the issue of determining the average labor market returns to additional schooling and the possible influence of differences in ability. 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. Economists have pursued a variety of analytical approaches to deal with this, including adjusting for cognitive test scores, but this work generally ignores issues of variation in school quality.[4]

There is mounting evidence that quality measured by test scores is directly related to individual earnings, productivity, and economic growth. A variety of researchers documents that the earnings advantages to higher achievement on standardized tests are quite substantial.[5] While these 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 experiences of workers, and other factors that might also influence earnings. In other words, quality measured by tests

[3]For a discussion and analysis of accountability systems, see Hanushek and Raymond (2005).

[4]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. The only explicit consideration of school quality typically investigates expenditure and resource differences across schools, but these are known to be poor measures of school quality differences (Hanushek, 2002). Early discussion of ability bias can be found in Griliches (1974).

[5]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. This approach relates the logarithm of earnings to years of schooling, experience, and other factors that might yield individual earnings differences. The clearest analyses are found in the following references (which are analyzed in Hanushek (2002)). See Bishop (1989, 1991), O'Neill (1990), Grogger and Eide (1993), Blackburn and Neumark (1993, 1995), Murnane, Willett, and Levy (1995), Neal and Johnson (1996), Mulligan (1999), Murnane et al.(2000) Murnane, Willett, Duhaldeborde, and Tyler (2000), Altonji and Pierret (2001), Murnane, Willett, Braatz, and Duhaldeborde (2001), Lazear (2003), and Rose (2006).

similar to those currently used in accountability systems is closely related to individual productivity and earnings.

Three recent studies provide direct and quite consistent estimates of the impact of test performance on earnings (Mulligan, 1999; Murnane, Willett, Duhaldeborde, & Tyler, 2000; Lazear, 2003). These studies employ different nationally representative data sets 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 percent higher annual earnings.[6] By way of summary, median earnings in 2001, while differing some by age, were about $30,000, implying that a one standard deviation increase in performance would boost these by $3600 for each year of work life. The full value to individual earnings and productivity is simply the annual premium for skills integrated over 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.[7] Second, these analyses concentrate on labor market experiences from the mid 1980s and into the mid 1990s, but these might not be entirely representative of the current situation because other evidence suggests that the value of skills and of schooling has grown throughout and past that period. Third, future general improvements in productivity might lead to larger returns to skill if the recent trends of higher rewards to more skilled workers continue.[8]

Another part of the return to higher skills comes through continuation in school. 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.[9] Murnane et al. (2000) separate the direct returns to measured skill from the indirect returns of more schooling and suggest that perhaps one-third to one-half of the full return to higher achievement comes from further schooling. Note also that the effect of quality improvements on school attainment incorporates concerns about drop out rates. Specifically, higher student achievement keeps students in school longer, which will lead among other things to higher graduation rates at all levels of schooling.

This work has not, however, investigated how achievement affects the ultimate outcomes of higher education. For example, if over time lower-achieving students tend increasingly to attend college, colleges may be forced to offer more remedial courses, and the variation of what students know and can do at the end of college may expand commensurately. This possibility, suggested in *A Nation at Risk*, has not been fully investigated, but may fit into considerations of the widening of the

[6]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 percent increase and females a 10 percent increase per standard deviation of test performance. Lazear (2003), relying on a somewhat younger sample from NELS88, provides a single estimate of 12 percent. These estimates are also very close to those in Mulligan (1999), who finds 11 percent for the normalized AFQT score in the NLSY data. By way of comparison, estimates of the value of an additional year of school attainment are typically 7–10 percent, suggesting that the economic value of an additional year of schooling is equivalent to 0.6–0.8 standard deviations of test scores. For policy purposes, the right comparison would reflect the costs of the alternatives. While these comparisons are not currently feasible, we return to this discussion below.

[7]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.

[8]These estimates, 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 over time. In the past few decades, these increases have favored the more educated and skilled, thus increasing the return to skill. If these trends continue, the impact of improvements in student skills are likely to rise over the work life instead of being constant as portrayed here. On the other hand, such skill-biased change has not always been the case, and technology could push returns in the opposite direction.

[9]See, for example, Dugan (1976), Manski and Wise (1983). 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, Rivkin, and Taylor (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. Behrman, Kletzer, McPherson, and Schapiro (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.

distribution of income within schooling categories.[10]

The impact of test performance on individual earnings provides a simple summary of the primary economic rewards to an individual. This estimate combines the impacts on hourly wages and on employment/hours worked. It does not include any differences in fringe benefits or nonmonetary aspects of jobs. Nor does it make any allowance for aggregate changes in the labor market that might occur over time.

These estimates also do not directly provide information about the source of any skill differences. As the education production function literature suggests, a variety of factors influence achievement, including family background, peers, school factors, and individual ability (Hanushek, 1979, 1986). Rose (2006), for example, estimates models of earnings that attempt to distinguish between skills deriving from high school quality differences and those from ability and earlier schooling factors. Interestingly, her estimates indicate that the source of any skill differences has little impact on the economic returns, i.e. skills developed early and possible related to ability have the same impact on earnings as skills developed late through schooling.[11] This suggests that skill improvements, regardless of their source, have strong economic effects.

2.2. Impacts of quality on economic growth

The relationship between measured labor force quality and economic growth is perhaps even more important than the impact of human capital and school quality on individual productivity and incomes. 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 just discussed). 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 the quality of schooling.

The current economic position of the United States is largely the result of its strong and steady growth over the twentieth century. Economic models that explain differences in growth rates across countries invariably feature the importance of human capital.[12]

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, 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.

Hanushek and Kimko (2000) go beyond simple quantity of schooling and delve into quality of schooling. We incorporate the information about international differences in mathematics and science knowledge that has been developed through testing over the past four decades. And we find a remarkable impact of differences in school quality on economic growth.

The international comparisons of quality come from piecing together results of a series of tests administered over four decades. 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. These initial tests suffered from

[10]This logic is most clear for the college graduates. For high school graduates, the movement into the college category could leave the high school group more homogeneous and could work in the opposite direction. Empirical evidence on income inequality within schooling groups suggests that inequality has increased over time for both college and high school groups, but the increase for college is larger (Murnane et al., 1995). This analysis also suggests that increased demand for skills is one of the elements in this growing inequality.

[11]She does, however, find a larger difference in the effects of skills on female earnings as compared to male earnings than found in the prior work.

[12]Barro and Sala-i-Martin (2003) review recent analyses and the range of factors that are included. 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. As discussed below, schooling clearly interacts with other factors, and these other factors have been important in supporting US growth.

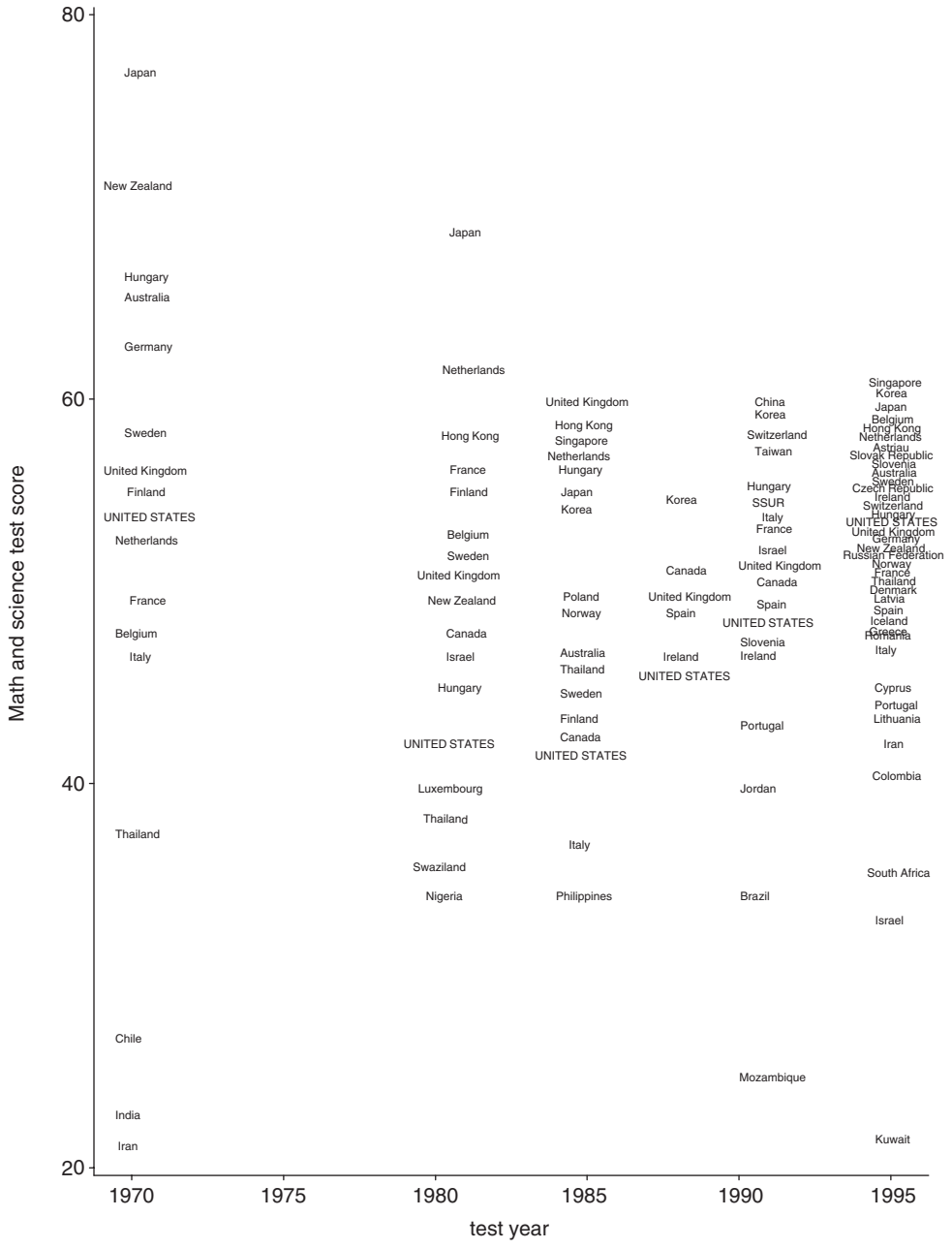

Fig. 1. Normalized test scores on mathematics and science examinations, 1970–1995.

a number of problems, but they did prove the feasibility of such testing.[13]

Subsequent testing, sponsored by the IEA and others, has included both math and science and has expanded on the group of countries that have been tested. In each, the general model has been to develop a common assessment instrument for different age groups of students and to work at obtaining a representative group of students taking the tests. An easy summary of the participating countries and their test performance is found

[13]The problems included issues of developing an equivalent test across countries with different school structure, curricula, and language; issues of selectivity of the tested populations; and issues of selectivity of the nations that participated. The first tests did not document or even address these issues in any depth. Subsequent assessments paid much more attention to selectivity issues.

in Fig. 1. This figure tracks performance aggregated across the age groups and subject area of the various tests and scaled to a common test mean of 50.[14] The United States and the United Kingdom are the only countries to participate in all of the testing.

There is some movement across time of country performance on the tests, but for the one country that can be checked—the United States—the pattern is consistent with other data. 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, also exhibits a sizable dip in the seventies, a period of growth in the eighties, and a leveling off in the nineties (see Campbell, Hombo, & Mazzeo, 2000 and http://nces.ed.gov/nationsreportcard/).

This figure also highlights a central issue here. In terms of measured math and science skills, US students have not been competitive on an international level. They have scored below the median of countries taking the various tests. Moreover, this figure—which combines scores across different age groups—disguises the fact that US performance is much stronger at young ages but falls off dramatically at the end of high school (Hanushek, 2003b).

Kimko's and my analysis of economic growth is very straightforward. We combine all of the available earlier test scores into a single composite measure of quality and consider statistical models that explain differences in growth rates across nations during the period 1960–1990.[15] The basic statistical models, which include the initial level of income, the quantity of schooling, and population growth rates, explain much of the variation in economic growth across countries.

Most important, the quality of the labor force as measured by math and science scores is extremely important. One standard deviation difference on test performance is related to 1 percent difference in annual growth rates of gross domestic product (GDP) per capita.[16] Before discussing the economic implications of these estimates, however, it is important to know whether or not these should be taken as causal impacts.

A series of separate analyses addresses the issue of whether the effect of quality is causal, a question frequently asked about international growth comparisons. Questions about causality arise because it is natural to believe that countries that grow and become richer decide to spend some added income on more schooling. Additionally, countries doing well on the tests could have other, concomitant attributes that influence their overall economic performance. The tests in Hanushek and Kimko (2000) involve: (1) investigation of international spending differences and test performance; (2) consideration of performance of immigrants in the US using the test score measures; and (3) exclusion of the high scoring East Asian countries.

As a starting point in the causal interpretation, a series of simple, cross-country educational production functions were estimated. Neither expenditure per pupil nor expenditure as a proportion of GDP influence the international achievement scores, thus ruling out the simplest version of the causality problem. The richer countries putting money back in the schools do not get higher achievement.

Perhaps the most persuasive test comes from looking at immigrants into the United States. For immigrants educated in their home country, differences in the international test scores used in our growth models contribute significantly to differences in earnings within the US labor market.[17] For those immigrants educated in the US, home country test scores have no influence on US earnings. Thus,

[14]The details of the tests and aggregation can be found in Hanushek and Kimko (2000) and Hanushek and Kim (1995). This figure excludes the earliest administration and runs through the Third International Mathematics and Science Study (TIMSS) (1995). Other international tests have been given and are not included in the figure. First, reading and literacy tests have been given in 1991 and very recently. The difficulty of unbiased testing of reading across languages plus the much greater attention attached to math and science both in the literature on individual earnings and in the theoretical growth literature led to the decision not to include these test results in the empirical analysis. Second, the more recent follow-ups to the 1995 TIMSS in math and science (given in 1999 and 2003) along with PISA tests administered in 2000, 2002, and 2003 by the OECD are excluded from the figure simply for presentational reasons.

[15]We exclude the two TIMSS tests from 1995 and later and the PISA tests because they were taken outside of the analytical period on economic growth. We combine the test measures over the 1965–1991 period into a single measure for each country. The underlying objective is to obtain a measure of quality for the labor force in the period during which growth is measured.

[16]The details of this work can be found in Hanushek and Kimko (2000) and Hanushek (2003b). Importantly, 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.

[17]These estimates come from standard Mincer models of earnings with schooling, potential experience, and potential experience squared.

it does not appear that our test-based measure of skills is simply a proxy for omitted factors that influence growth but instead represents true productivity differences.

Finally, because East Asian countries have tended to score at the top on the international math and science tests and simultaneously had high growth in the 1960–1990 period, one might be concerned that the test measures are simply identifying the East Asian countries as opposed to measuring the importance of labor force quality. To assess this, we omit various subsets of the East Asian countries and look at the influence of achievement on the growth of the remaining countries. The effect of one standard deviation difference on test performance falls to three-fourths of a percent difference in annual growth rates, indicating that skills were an important element of the East Asian growth but also are very important in other areas of the world.

Each of the tests lends support to a causal interpretation of the effects of test-based measured quality on economic growth rates. Nonetheless, similar to the modeling of individual earnings, the tests do not identify the source of test differences—parents, schools, or what-have-you. We return to those issues below.

This quality effect, while possibly sounding small, is actually very large and significant. Because the added growth compounds, it leads to powerful effects on US national income and on societal well-being. To underscore the importance of quality, it is possible to simulate the effects of improved achievement. At this point, the discussion begins to be phrased in terms of improvements in schools, because that is the site of most public policies. But, achievement gains that come from other sources—say, better childhood nutrition—can also be evaluated within this framework if they lead to student achievement gains.

As a benchmark, consider a policy introduced in 2005 that leads to an improvement of scores of graduates of one-half standard deviation by the end of a decade. This would be a significant change that would put US student performance in line with that of students in a variety of middle performing European countries, e.g. Sweden, the Czech Republic, or Ireland. But US students still would not be at the top of the world rankings. Such benchmark performance is noticeably below the lofty goal of the 1989 governor's summit that envisioned being first in the world in math and science by 2000—a goal that we did not attend during the 1990s.

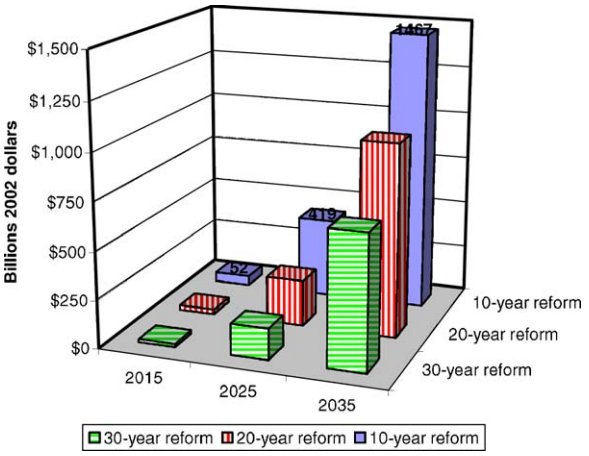

Fig. 2. Growth dividend from 0.5 s.d. reforms begun in 2005.

Reaching this America 2000 goal would imply more than doubling this benchmark.

Such a path of improvement would not have an immediately discernible effect on the economy, because new graduates are always a small portion of the labor force, but the impact would mount over time. If past relationships between quality and growth hold, GDP in the United States would end up two percent higher by 2025 and five percent higher by 2035. Remember that our current GDP is roughly $12.5 trillion, meaning that one percent is $125 billion.

This kind of performance change may or may not be feasible, but the impact on GDP illustrates the real importance of effective school reform. To give some idea of the range of possible outcomes, Fig. 2 continues the goal of a one-half standard deviation improvement in student performance but aims to achieve this over different time periods ranging from 10 to 30 years. A 30-year reform plan would still yield a gain to the economy in 2035 of over $700 billion dollars, or almost four percent of GDP then, itself an attractive outcome.[18]

The summary of this analysis is that improvements in schooling outcomes are likely to have very powerful impacts on individuals (the previously identified effect on earnings) and on the economy as a whole. The impact on the aggregate economy will raise the whole economy over and above the individual differences estimated above.

[18]All calculations are stated in constant 2002 dollars. GDP follows from the Congressional Budget Office projections of potential GDP. Potential GDP in trillions is projected to be: $16.6 T in 2015; $22.0 T in 2025; and $29.3 T in 2035.

3. Feasible teacher quality policies

The prior analysis has simply projected the benefits of achieving various goals for student achievement. A first question is whether or not achieving such gains could be feasible with realistic reform strategies.

Past reform efforts clearly do not support feasibility. During the two decades since publication of *A Nation at Risk*, a variety of approaches have been pursued (Peterson, 2003). These have involved expanding resources in many directions, including increasing real per pupil spending by more than 50 percent. Yet performance has remained unchanged since 1970 when we started obtaining evidence from NAEP (see Campbell et al., 2000).

The aggregate picture is consistent with a variety of other studies indicating that resources alone have not yielded any systematic returns in terms of student performance (Hanushek, 1997, 2003b). The character of reform efforts can largely be described as "same operations with greater intensity." Thus, pupil–teacher ratios and class size have fallen dramatically, teacher experience has increased, and teacher graduate degrees have grown steadily—but these have not translated into higher student achievement. On top of these resources, a wide variety of programs have been introduced with limited aggregate success. The experience of the past several decades vividly illustrates the importance of true reform, i.e. reform that actually improves student achievement.

One explanation for past failure is simply that we have not directed sufficient attention to teacher quality. By many accounts, the quality of teachers is the key element to improving 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.[19]

Rivkin, Hanushek, and Kain (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, Kain, & Rivkin, 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. To do this, it concentrates entirely 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.

The estimates of this analysis indicate that the differences in annual achievement growth between an average and a good teacher (i.e. the average versus a teacher one standard deviation higher in the achievement distribution) are at least 0.11 standard deviations of student achievement.[20]

An alternative approach to estimating the variation in teacher quality considers achievement gains across individual classrooms for a large district in Texas (Hanushek, Kain, O'Brien, & Rivkin, 2005). Looking at just variations among teachers within schools, the comparable estimate of a standard deviation in teacher quality is 0.15 standard deviations in annual student achievement gains. This latter analysis uses a different strategy to estimate the quality of teachers, but the bounds for quality differences within schools remain quite tight.

Before going on, it is useful to put these estimates of the variation in quality into perspective. Specifically, since the Coleman Report (Coleman et al.(1966)), many people have presumed that family influences entirely dominate student achievement and that schools can do little to overcome the deficits of poor family preparation. The average gap

[19]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.

[20]For this calculation, a teacher at the mean of the quality distribution is compared to a teacher 1.0 s.d. higher in the quality distribution (84th percentile), labeled a "good teacher".

between a typical low income student and a student not on free or reduced lunch on the math tests in Texas that were used in the prior work is approximately one-half standard deviation (depending on grade level). While a bit speculative, if these annual gains accumulate in a linear fashion, disadvantaged students assigned to a good teacher (as opposed to an average teacher) for 4–5 years in a row could, by these estimates, entirely close the income gap.[21]

For our policy calculations, a reasonable estimate is that differences in quality are twice the lower bound (0.22 s.d.). This larger estimate reflects likely differences in teacher quality among schools (plus the series of other factors that bias the previously discussed estimate downwards). The higher second estimates of teacher quality also leave out between school variations and for other reasons are likely themselves to be biased downward (Hanushek et al., 2005). The teacher quality estimate for policy calculations of 0.22 s.d. of student achievement is 50 percent higher than this alternative estimate.

These estimates of the importance of teacher quality permit some calculations of what would be required to yield the reforms discussed earlier. To begin with, 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 kind of policy is required.

To fix the idea of changing teacher quality, the analysis pursues a simple quality upgrade policy. Specifically, consider that schools only hire new teachers when an existing teacher decides to leave. A simple plan would be a steady improvement plan where the average new hire is maintained at a constant amount better than the average teacher in any given year. 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 (compared to the existing stock) each year of the reform period. By maintaining this standard for replacement of all teachers exiting

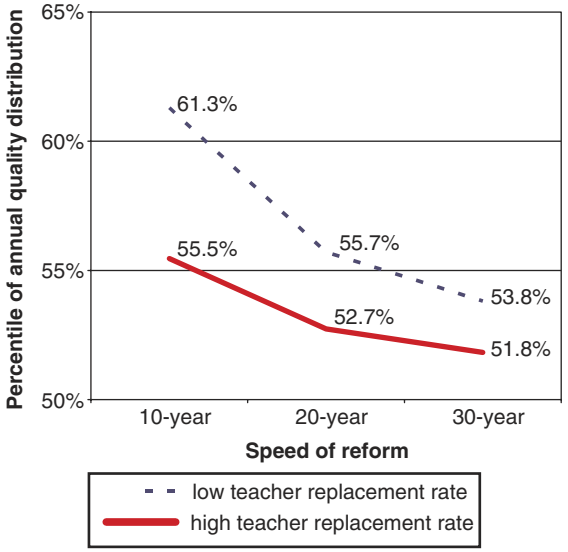

Fig. 3. Annual required hiring percentiles for 0.5 s.d. reforms.

teaching (6.6 percent annually in 1994–1995) but retaining all other teachers, this policy would yield a 0.5 standard deviation improvement in student performance after a 20 year period.[22] If instead we thought of applying these new standards to all teacher turnover (exits plus the 7.2 percent who change schools), a 0.5 s.d. improvement in student performance could be achieved in 10 years.

Fig. 3 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 10 years by upgrading just those who exit teaching 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.

[21]It has not been possible to analyze the dynamics of different patterns of teachers, so estimates of the cumulative effects of good and bad teachers involve extrapolating from the estimated teacher effects.

[22]Exiting teachers are assumed on average to be the average quality. This is reasonably close to the findings in Hanushek et al. (2005). The chosen replacement levels are, of course, somewhat arbitrary, because any policy changes may directly affect both replacement rates and the quality of exiting teachers. For example, the next estimates based on all school changers would imply that only high quality teachers seeking to relocate are hired—and this might slow overall turnover.

These calculations demonstrate the challenge of achieving substantial improvements in achievement. It requires significantly upgrading the quality of the current teacher force. It also requires a lengthy commitment to policies that lead to continuing quality improvements.

Several aspects of these scenarios deserve note. First, the improvements that are required apply to the teacher distribution that exists *each* year. In other words this standard requires continual improvement in terms of the current teachers. The continual improvement comes from the fact that the distribution of teachers improves each year because of the higher quality teachers hired in prior years. At the same time, it does not imply that all new teachers reach these levels, only that the average teacher does. There will still be a distribution of teachers in terms of quality.

In fact, it is easy to summarize what the distribution of teachers must look like in terms of the current distribution of teachers. In order to achieve a 0.5 standard deviation improvement in student achievement, the average teacher (after full implementation of reform) must be at the 58th percentile of the current distribution. (In order to achieve a 1.0 s.d. improvement, the average teacher must be at the 65th percentile of the current distribution). The annual adjustments given previously simply translate these quality calculations into the path required for reaching them under different reform periods.

It is important to emphasize, however, that the calculations also freeze many aspects of teaching. They assume no change in teacher turnover rates. Of course, teacher turnover will be affected by a variety of policies aimed at quality improvement such as salary policy, tenure, etc. The calculations also assume that turnover is unrelated to quality—as it largely is with today's passive teacher management approach. An active selection and teacher retention policy could, however, lead to improvements in overall teacher quality that would offer relief from the stringency of hiring standards that are required. For example, a policy that retained the best teachers two years longer and dropped the least effective teachers two years sooner would by itself lead to substantial improvements in the average quality of the teacher force.

The required improvements in the teaching force could also be achieved in other ways, at least conceptually. For example, a new professional development program that boosts the quality of current teachers would accomplish the same purpose. However, any such program must be in addition to the current amount of professional development, including obtaining master's degrees and completing in-service training, because the existing professional development activities are already reflected in the current quality distributions.

4. Cost considerations

Analyzing reform policies directly in terms of their costs is not feasible, because we know very little about the supply function for teacher quality. While there has been some work on the cost of hiring teachers with different characteristics (such as experience or advanced degrees), these characteristics do not readily translate into teacher quality (Hanushek, 2003a; Hanushek & Rivkin, 2004).

There are two alternative ways to consider the costs related to any policies aimed at improving the teaching force. First, the prior calculations of the benefits provide an estimate of the upper bounds on the costs of feasible policies (i.e. costs must be less than benefits in order for the policy to be efficient). Second, while limited by current experience, actual programs similar to those being contemplated can be evaluated in terms of costs to achieve any outcome.

Much of the current discussion of teacher quality is centered on statements about the overall level of salaries. It seems clear that teacher salaries have slipped relative to alternative earnings of college workers, particularly for women (Hanushek & Rivkin, 1997, 2004).[23] This does not, however, give much policy guidance, because we do not know how teacher quality responds to different levels of salaries (Hanushek & Rivkin, 2004).

Overall increases in salaries would undoubtedly increase the pool of applicants and presumably, by climbing higher in the pool of college graduates, could be expected to have some increase in quality as seen from the alternative jobs from which the new applicants come. However, the impact on overall teacher quality would depend, first, on how correlated skills demanded in other jobs are with teaching skills and, second, on the ability of schools to select better teachers from this enlarged pool. Existing research does not provide complete evidence on these, although there are strong suspicions

[23]There is a current debate about how salaries of teachers compare to those in different professions; see Podgursky (2003).

that hiring is not particularly precise (Ballou, 1996; Ballou & Podgursky, 1997). Moreover, policies that simply raise salaries across-the-board (even if advanced as a way to increase the attractiveness of the profession) would almost certainly slow any reform adjustments, because they would lower teacher turnover and make it more difficult to improve quality through new hiring.

The previously cited analyses on teacher quality suggest that most of the variance in quality appears to lie within schools and not between schools (Rivkin et al., 2005; Hanushek et al., 2005). Moreover, because quality is not related to experience (except in the very earliest years) or to having graduate degrees, changing overall teacher salaries is not a very refined instrument for changing the quality of the teacher force.

Changing the quality of teaching force almost certainly must rely upon either salaries and other employment terms that are directed at quality or on differential retention policies that emphasize classroom outcomes. On this score, little current evidence exists. As discussed below, while many merit pay plans have not succeeded in the past, these have tended to involve small amounts of pay and have seldom considered the long run effects that operate through selection of teachers. Lacking specific evidence on the quality-salary locus, we instead begin by calculationg the magnitude of economically feasible programs.

It is clear that very large programs to improve teacher quality are economically feasible (if they truly increase quality). The aggregate growth numbers suggest that the annual growth dividend from an effective reform plan would cover most conceivable program costs over a relatively short period of time. For example, a 10-year reform plan that yielded a one standard deviation improvement in student performance would produce an annual reform dividend that more than covered the *entire* expenditure on K-12 education by 2025.[24] Of course, as suggested previously, a reform program of this magnitude and speed would require dramatic changes in hiring of new teachers. But a 20-year reform program with a 0.5 s.d. improvement would produce a sufficient dividend to cover all K-12 expenditure by 2035.

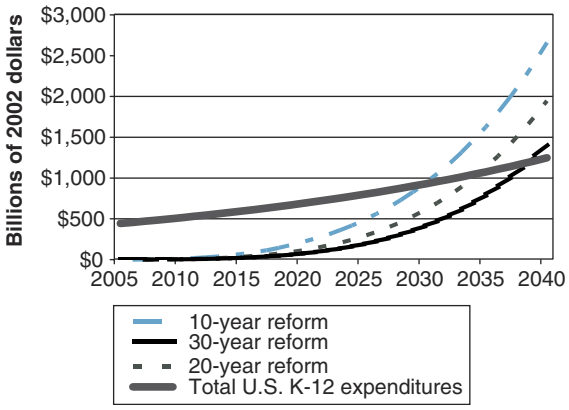

Fig. 4. Annual growth dividends from 0.5 s.d. reforms begun in 2005 compared to K-12 expenditures in US.

Fig. 4 traces out the growth dividend relative to the total education budget for the United States. Educational expenditure for K-12 is calculated to grow at a real 3 percent annually, and the growth dividend of a 0.5 standard deviation reform plan (of varying speed) is plotted against this. This figure shows vividly how true reform (i.e. reform that actually yields improvement in student performance) has a cumulative effect on the economy.

Alternatively, consider a set of teacher bonuses. If half of the teachers received bonuses averaging 50 percent of salary, the average bonus today would be approximately $12,500 per year.[25] There are different ways to judge the magnitude of this. First, in aggregate terms the total annual expenditure for teacher bonuses in 2025 would be approximately $81 billion, or slightly over 10 percent of the total K-12 expenditure in that year.[26] This magnitude is identical to the annual reform dividend from growth in 2025 from a 30-year reform yielding just a 0.25 standard deviation improvement.

But, teacher bonuses can be considered from another perspective. A one-half standard deviation improvement in performance raises the future salaries of the students (not counting any growth effects) by around $1800 per year. This annual addition to earnings translates into a present value of $30,000 for each student.[27] A bonus to a teacher

[24]These calculations assume that K-12 expenditures growth at 3 percent (real), implying that the current half trillion dollar expenditure would grow the $777 billion in 2025.

[25]These calculations assume that the current average teacher salary is $50,000—a figure close to the National Education Association survey data.

[26]These calculations assume a constant teacher force of 3 million (compared to 2.8 million in 2000) and a 3 percent real growth in teacher salaries.

[27]These calculations assume 35 years of working life and a 5 percent net discount rate. The net discount rate represents the

of $12,500 per year could then be recouped in increased student earnings with a pupil-teacher ratio of six or more, as long as the bonuses elicited at least a 0.5 standard deviation improvement in student skills.[28] In other words, the minimum average class size that justifies such bonuses is very small.

The alternative of extrapolating from existing incentive programs is not feasible. Estimating the costs of achieving improvements in the teacher force is generally impossible based directly on current data. We simply have limited experience with any policies that alter the incentives for hiring and retaining high quality teachers (and which also evaluate the outcomes).

Evidence from existing merit pay plans, for example, is not relevant for consideration of hiring new, higher quality teachers. Specifically, these plans are designed largely to increase teacher "effort" as opposed to attracting and retaining a new set of teachers.[29]

A few incentive schemes have been evaluated, and they provide suggestive but not very generalizable results. For example, one promising program is the Teacher Advancement Program (TAP) of the Milken Family Foundation. This is a broad program with several elements, but a unique component is a teacher evaluation and bonus system based on performance in the classroom. The separate components have been not been costed out or evaluated fully, but the initial results suggest that the overall program appears to cost about $400 per student and to have achieved performance gains of about 0.4 standard deviations (compared to a set of control schools).[30] If generalizable, this program at even half the performance result would be economically justified by either gains in individual earnings or aggregate effects.

Another evaluation is found in an experiment in Israel (Lavy, 2002). Schools were placed in competition with each other, and teachers in the highest performing schools received salary bonuses. These salary bonuses, given to the entire school faculty, were rather modest (approaching three percent at the top). Nonetheless, schools competing for bonuses did better than another set of schools that just received resources.[31] This program shows that schools react to incentives, but it is unclear how to translate that into costs and benefits for a set of US schools.

The conclusion of the cost considerations is simple. The benefits from quality improvements are very large. Thus, they can support incentive programs that are quite large and expansive *if the programs work*. US schools have in fact expanded in a variety of ways over the past four decades—real expenditures per pupil in 2000 are more than three times those in 1960. It is just that these past programs have not led to significant improvements in student performance. Put another way, the benefits do not justify all types of expenditure. They do justify many conceivable programs if they can be shown to be effective.

5. What is not considered

These calculations simplify many facets of the problem and ignore many others. It is useful to list some of the major factors that have been ignored.

On the benefit side, the discussion ignores all nonmonetary gains. For example, none of the potential improvements in society—from improved functioning of our democracy to lowered crime—are considered. Moreover, other possible gains such as improved health outcomes or better child development are not included (even though they could conceptually be estimated).[32] While there is evidence that a variety of these nonmonetary factors are related to quantity of schooling, there is simply limited evidence about the relationship with quality.

On the cost side, improved school performance is likely to lower other schooling costs. For example,

(*footnote continued*)
interest rate above any annual growth in real income as would occur with general productivity improvements. Thus, this is a high discount rate, since a 3 percent growth in earnings per year would imply that the gross discount rate is 8 percent.

[28]The calculations assume that the teacher bonuses apply to teachers in grades K through 12 for each student and that they would be spread across all of the students in the "average" class for the teachers.

[29]The standard citation on merit pay and its ineffectiveness is Cohen and Murnane (1986). A discussion of alternative perspectives on merit pay is found in Hanushek et al. (1994).

[30]This program is currently in its initial phases and the evaluation is on-going. Some preliminary results can be found in Schacter, Thum, Reifsneider, and Schiff (2004). Cost figures come from private correspondence.

[31]This statement reflects the cost-effectiveness of the two programs. The additions to resources were much larger than the bonuses, and schools with added resources obtained larger absolute test score gains than schools with just teacher incentives.

[32]See, for example, Black, Devereux, and Salvanes (2003) for a suggestive analysis of Norwegian experiences.

improvements in early reading could well lessen the costs of special education (Lyon & Fletcher, 2001). Current remedial costs, both in K-12 and in higher education, would almost certainly decline with better classroom instruction (Greene, 2000).

Both of these elements reinforce the previous economic analyses and further swing the case toward investing in improved quality. Yet, since the previous calculations are so clear, no effort is made to include these, potentially important, elements.

6. Conclusions

Again, it is necessary to put this discussion within the context raised originally. Many people are dissatisfied with the performance of US schools. The dissatisfaction comes from a variety of viewpoints, and the suggested remedies also go in many directions.

One of the on-going debates about school policy has been a very fundamental one of what kinds of skills and programs to emphasize. While perhaps over-simplifying, it involves how closely schools should be aligned with the world of work. This debate typically centers on the types of programs for students who are likely not to continue with further schooling, often suggesting an increased emphasis on vocational skills. While often implicit, a related argument is that emphasis on general cognitive skills (as in school accountability systems) is not productive.

This paper does not directly contrast the alternative approaches. Instead, it investigates the economic benefits to improvements in cognitive skills. The prior analysis demonstrates that better student outcomes generate considerable benefits. While these benefits have not been previously quantified, the presumption that they exist has surely propelled much of the interest in our schools that has existed at least since *A Nation at Risk*.

These findings are particularly relevant to current attention to school outcomes. The federal law, *No Child Left Behind*, requires states to institute accountability systems that insure all students are proficient in core subjects. These accountability systems emphasize measured cognitive skills of just the kind shown to have high payoffs in the labor market and for society.

Looked at from the different perspective of enhancing school-to-work programs, the evidence suggests that alternative programs must be very highly effective if they involve a loss in cognitive skills. In other words, if there is a trade-off between more vocationally oriented programs and academic results, the former programs must be sufficient to compensate for any losses through reduced academic skills.

Of course, there does not have to be a trade-off. If in fact more vocationally rich curriculum provides better motivation to a segment of students and this motivation pays off in better academic skills, the potential concerns go away.

A part of the picture, however, that has not received as much attention is what is required to achieve the student outcome gains. This analysis uses available information about the current distribution of teacher quality to sketch out the kinds of changes that would be required for reform programs of differing magnitude and speed. This analysis highlights the fact that reform, at least as seen from past experience, will require a significant upgrading of the teaching force.

The benefit picture indicates that improvements in student performance have truly substantial impacts on individual productivity and earnings and on the growth and performance of the aggregate economy. The economic gains could in fact cover some substantial changes in expenditure on schools.

Past history, however, provides a key caution. The US has devoted substantial attention to its schools. In just the two decades since *A Nation at Risk*, the nation has increased real spending on schools by over 50 percent. But it has gotten little in terms of student outcomes. We have accumulated considerable experience on things that do not work, but much less on policies that will succeed.

The available evidence does indicate that improvement in the quality of the teacher force is central to any overall improvements. And improving the quality of teachers will almost certainly require a new set of incentives, including selective hiring, retention, and pay.

Acknowledgments

Helpful comments were received from David Neumark and an anonymous referee. This research was supported by the Packard Humanities Institute and The Teaching Commission.

References

Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics*, *116*(1), 313–350.

Ballou, D. (1996). Do public schools hire the best applicants? *Quarterly Journal of Economics*, *111*(1), 97–133.

Ballou, D., & Podgursky, M. (1997). *Teacher pay and teacher quality*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Barro, R. J., & Sala-i-Martin, X. (2003). *Economic growth* (2nd ed.). Cambridge, MA: MIT Press.

Behrman, J. R., Kletzer, L. G., McPherson, M. S., & Schapiro, M. O. (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*, 12–23.

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* (pp. 146–186). Washington, DC: The AEI Press.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2003). Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. NBER Working Paper No. w10066 (November).

Blackburn, M. L., & Neumark, D. (1993). Omitted-ability bias and the increase in the return to schooling. *Journal of Labor Economics*, *11*(3), 521–544.

Blackburn, M. L., & Neumark, D. (1995). Are OLS estimates of the return to schooling biased downward? Another look. *Review of Economics and Statistics*, *77*(2), 217–230.

Campbell, J. R., Hombo, C. M., & Mazzeo, J. (2000). *NAEP 1999 trends in academic progress: Three decades of student performance*. Washington: National Center for Education Statistics.

Card, D. (1999). Causal effect of education on earnings. In O. Ashenfelter, & D. Card (Eds.), *Handbook of labor economics* (pp. 1801–1863). Amsterdam: North-Holland.

Cohen, D. K., & Murnane, R. J. (1986). Merit pay and the evaluation problem: Understanding why most merit pay plans fail and a few survive. *Harvard Educational Review*, *56*(1), 1–17.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., et al. (1966). *Equality of educational opportunity*. Washington, DC: US Government Printing Office.

Dugan, D. J. (1976). Scholastic achievement: Its determinants and effects in the education industry. In J. T. Froomkin, D. T. Jamison, & R. Radner (Eds.), *Education as an industry* (pp. 53–83). Cambridge, MA: Ballinger.

Easterly, W. (2002). *The elusive quest for growth: An economists' adventures and misadventures in the tropics*. Cambridge, MA: The MIT Press.

Greene, J. P. (2000). *The cost of remedial education: How much Michigan pays when students fail to learn basic skills*. Midland, MI: Mackinac Center for Public Policy.

Griliches, Z. (1974). Errors in variables and other unobservables. *Econometrica*, *42*(6), 971–998.

Grogger, J. T., & Eide, E. (1993). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources*, *30*(2), 280–310.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, *14*(3), 351–388.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, *24*(3), 1141–1177.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, *19*(2), 141–164.

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

Hanushek, E. A. (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?* (pp. 141–173). Stanford, CA: Hoover Institution Press.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). The market for teacher quality. Working Paper No. 11154 (February).

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, *39*(2), 326–354.

Hanushek, E. A., & Kim, D. (1995). Schooling, labor force quality, and economic growth. Working Paper 5399, National Bureau of Economic Research (December).

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor force quality, and the growth of nations. *American Economic Review*, *90*(5), 1184–1208.

Hanushek, E. A., & Pace, R. R. (1995). Who chooses to teach (and why)? *Economics of Education Review*, *14*(2), 101–117.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, *24*(2), 297–327.

Hanushek, E. A., & Rivkin, S. G. (1997). Understanding the twentieth-century growth in US school spending. *Journal of Human Resources*, *32*(1), 35–68.

Hanushek, E. A., & Rivkin, S. G. (2004). How to improve the supply of high quality teachers. In D. Ravitch (Ed.), *Brookings papers on education policy 2004* (pp. 7–25). Washington, DC: Brookings Institution Press.

Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics*, *78*(4), 611–627.

Hanushek, E. A., et al. (1994). *Making schools work: Improving performance and controlling costs*. Washington, DC: Brookings Institution.

Heckman, J. J., & Vytlacil, E. (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.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review*, *10*(3), 179–214.

Lavy, V. (2002). 'Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy*, *110*(6), 1286–1317.

Lyon, G. R., & Fletcher, J. M. (2001). Early warning system: How to prevent reading disabilities. *Education Matters*, *1*(2), 23–29.

Manski, C. F., & Wise, D. A. (1983). *College choice in America*. Cambridge: Harvard University Press.

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

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

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

National Commission on Excellence in Education (1983). *A nation at risk:The imperative for educational reform.* Washington, DC: US Government Printing Office.

Neal, D. A., & Johnson, W. R. (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.

Peterson, P. E. (Ed.). (2003). *Our schools and our future: Are we still at risk?*. Stanford, CA: Hoover Press.

Podgursky, M. (2003). Fringe benefits. *Education Next*, *3*(3), 71–76.

Rivkin, S. G. (1995). Black/white differences in schooling and employment. *Journal of Human Resources*, *30*(4), 826–852.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Rose, H. (2006). Do Gains in Test Scores Explain Labor Market Outcomes? *Economics of Education Review*, this volume, doi:10.1016/j.econedurev.2005.07.005.

Schacter, J., Thum, Y. M., Reifsneider, D., & Schiff, T. (2004). *TAP preliminary results report: Year three results from Arizona and year one results from South Carolina*. Santa Monica, CA: Miken Family Foundation.