



Stanford
Center for International
Development

Working Paper No. 198

Some Simple Analytics of School Quality

by
Eric A. Hanushek*

December 2003



Stanford University
John A. and Cynthia Fry Gunn Building
366 Galvez Street | Stanford, CA | 94305-6015

* Hoover Institution, Stanford University; University of Texas at Dallas; and National Bureau of Economic Research. This research was supported by the Packard Humanities Institute and The Teaching Commission.

Some Simple Analytics of School Quality

By Eric A. Hanushek

Public investment in schooling is at its heart an exercise in economics. All governments of the world assume a substantial role in providing education for their citizens. A variety of motivations lead societies to provide such strong support for schooling – some of which come from pure economics and others of which come from ideas of improved political participation, of social justice, and of general development of society. No matter what the motivation, however, the fundamental question of ‘how much should society invest?’ remains. Public investment in education comes at the expense of other public and private uses of the funds – although being an investment there is the prospect that any expenditure will be partially or fully offset by increased productivity and output that is engendered. Analysis of the benefits and costs of school reform indicates investments that improve the quality of schools offer exceptional rewards to society.

Most consideration of economic aspects of education has concentrated on school attainment, or the quantity of education. This is natural. First, it is easy to calculate the economic return on such an investment – both the costs and benefits are fairly clear. Second, until recently, relatively limited data have been available on the quality of schools. Third, there are great uncertainties about how to change quality and what it costs. Nonetheless, the policy issues today are ones of quality. 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. While it precipitated an unbroken period of concern about U.S. schools, it did not lead to any substantial improvements in school quality (Peterson (2003)).

This overview highlights what is known about investments in school quality. It then attempts to provide some bounds on the economics of school reform. It remains 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 – but providing valuations of these is not currently possible.

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.

Benefits of Enhanced School Quality

Economists have devoted considerable attention to understanding how human capital affects a variety of economic outcomes. The underlying notion is that individuals make investment decisions in themselves through schooling and other routes. The accumulated skills that are relevant for the labor market from these investments over time represent an important component of the human capital of an individual. The investments made to improve skills then return future economic benefits in much the same way that a firm's investment in a set of machines (physical capital) returns future production and income. In the case of public education, parents and public officials act as trustees for their children in setting many aspects of the investment paths.

In looking at human capital and its implications for future outcomes, economists are frequently agnostic about where these skills come from or how they are produced. Although we return to that below, it is commonly presumed that formal schooling is one of several important contributors to the skills of an individual and to human capital. It is 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. For this reason, 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 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 U.S. 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 and the knowledge and skills of graduates.

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, 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 for dealing with this, including adjusting for measured cognitive test scores, but this work generally ignores issues of variation in school quality.¹

¹ 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 Vytlačil (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

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.² 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, 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.

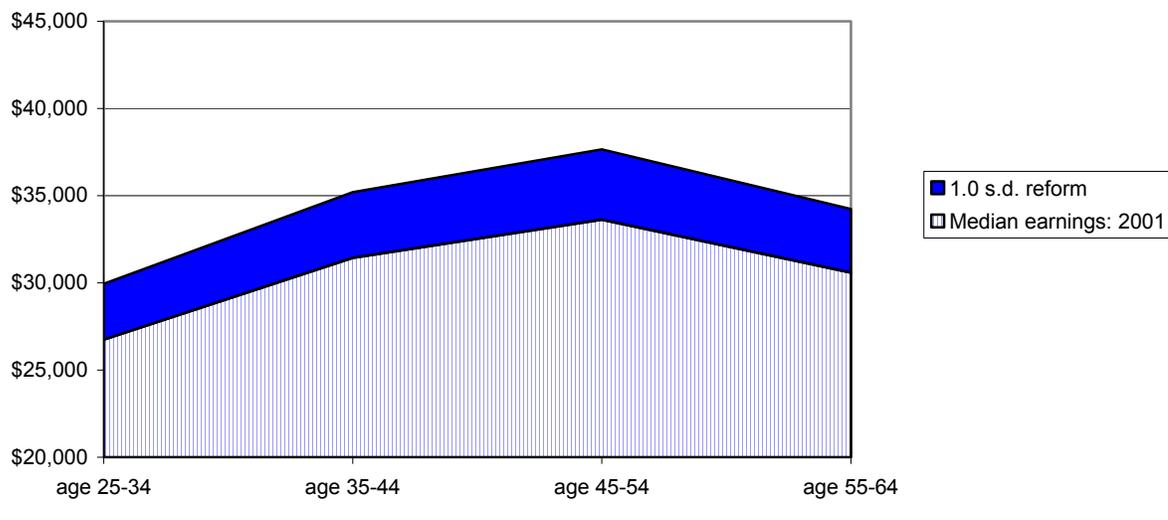
Three recent studies provide direct and quite consistent estimates of the impact of test performance on earnings (Mulligan (1999), Murnane et al. (2000); Lazear (2003)). These studies employ different nationally representative data sets that follow students after they leave schooling and enter the labor 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.³ The impact of one standard deviation in test performance is illustrated in Figure 1 which builds on the level of median annual earnings for workers in 2001. By way of summary, median earnings, while differing some by age, were about \$30,000, implying that a one

of school quality differences (Hanushek (2002)). Early discussion of ability bias can be found in Griliches (1974).

² 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); Altonji and Pierret (2001); Murnane et al. (2001); and Lazear (2003).

³ 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.

**Figure 1. Median U.S. Individual Earnings with 1.0 s.d.
Reform ($\gamma=0.12$)**



standard deviation increase in performance would boost these by \$3,600 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 (mid20's to early 30s), and other analysis suggests that the impact of test performance becomes larger with experience.⁴ Second, the labor market experiences that are observed begin the mid1980's and extend into the mid1990s, 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.⁵

Another part of the return to school quality comes through continuation in school.⁶ There is substantial U.S. evidence that students who do better in school, either through grades or scores on standardized achievement tests, tend to go farther in school.⁷ Murnane et al. (2000) separate

⁴ 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.

⁵ These estimates, as highlighted in Figure 1, 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. Thus, the impact of improvements in student skills are likely to rise over the work life instead of being constant as portrayed here.

⁶ 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. 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 for dealing with this, including adjusting for measured cognitive test scores, but this work generally ignores issues of variation in school quality. 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)).

⁷ 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.

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. (Figure 1 is just the direct effects of skills, not including the indirect effects coming through added 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 investigated, but may fit into considerations of the widening of the distribution of income.

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.

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

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 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.

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. Economists have developed a variety of models and ideas to explain differences in growth rates across countries – invariably featuring the importance of human capital.⁸

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.

⁸ Barro and Sala-I-Martin (1995) 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 U.S. growth.

The international comparisons of quality come from piecing together results of a series of tests administered over the past 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 a number of problems, but they did prove the feasibility of such testing and set in motion a process to expand and improve on the undertaking.⁹

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 in figure 2. This figure tracks performance aggregated across the age groups and subject area of the various tests and scaled to a common test mean of 50.¹⁰ 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 U.S. students for different subjects and ages. NAEP performance over this period,

⁹ 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.

¹⁰ 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-up to the 1995 TIMSS in math and science (given in 1999) is excluded from the figure simply for presentational reasons.

shown in Appendix Figure A1, also exhibits a sizable dip in the seventies, a period of growth in the eighties, and a leveling off in the nineties.

This figure also highlights a central issue here. The U.S. has not been competitive on an international level. It has scored below the median of countries taking the various tests. Moreover, this figure – which combines scores across different age groups – disguises the fact that U.S. performance is much stronger at young ages but falls off dramatically at the end of high school (Hanushek (2003)).

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 to 1990.¹¹ The basic statistical models, which include the initial level of income, the quantity of schooling, and population growth rates, explain a substantial portion 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.¹² A series of separate tests addresses the issue of whether the effect of quality is causal, a question frequently asked about international growth comparisons. Each test is consistent with a causal interpretation.¹³

¹¹ We exclude the two TIMSS tests from 1995 and 1999 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.

¹² The details of this work can be found in Hanushek and Kimko (2000) and Hanushek (2003). 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.

¹³ Questions about causality arise in studies of the quantity of schooling because countries that grow and become richer may decide to spend some of their added income on more schooling. 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 U.S. using the test score measures; and 3) exclusion of

This quality effect, while possibly sounding small, is actually very large and significant. Because the added growth compounds, it leads to powerful effects on U.S. national income and on societal well-being.

To underscore the importance of quality, it is possible to simulate the effects of alternative reforms of U.S. schools. As a benchmark, consider a policy introduced in 2005 that leads to an improvement of scores of graduates of one standard deviation by the end of a decade. This would be an exceptional change. An improvement of that magnitude would put U.S. student performance in line with that of students in a variety of high performing European countries, but they still would not be at the top of the world rankings. (It does, however, have a similar lofty goal to that of the governor's summit in 1989 that set a goal of being first in the world in math and science by 2000 – a goal that we did not dent during the 1990s).

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 four percent higher by 2025 and ten percent higher by 2035.

This kind of 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, Figures 3 and 4 trace out improvements in the national economy from slower and lesser changes in student outcomes.

Figure 3 retains the goal of a one standard deviation improvement in 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 \$1.4 trillion dollars, or five percent.¹⁴

the high scoring East Asian countries.

¹⁴ 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: \$16.6T in 2015; \$22.0T in 2025; and \$29.3T in 2035.

Figure 3. Growth Dividend from 1.0 s.d. Reform Begun in 2005

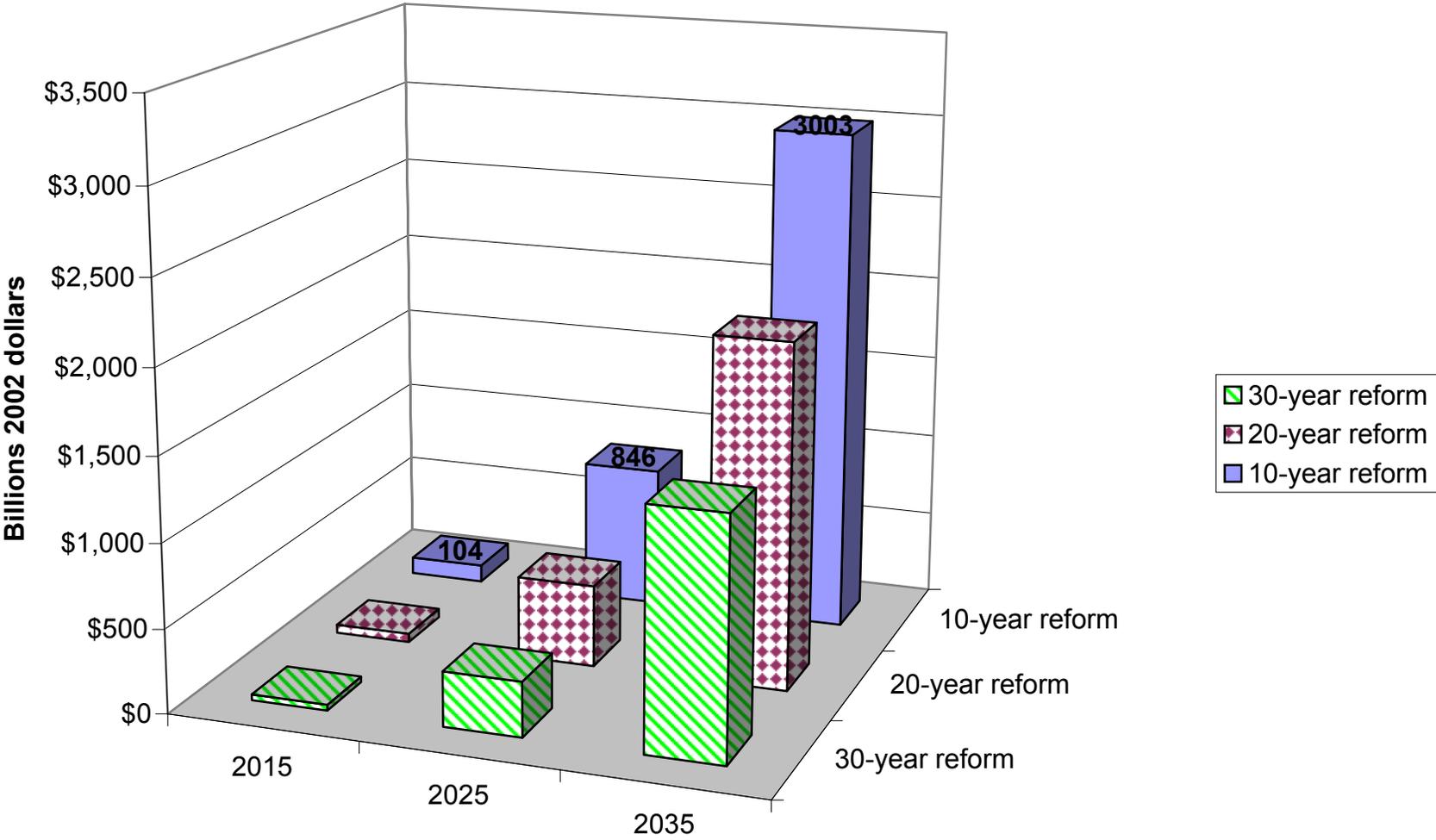


Figure 4 analyzes the outcomes of policies that achieve a more modest 0.5 standard deviation improvement in math and science, again with different time paths. While not precise, such a policy yields roughly half the gains in GDP – but the gains nevertheless remain very large and important. The \$700 billion growth dividend in 2035 that follows a 30-year reform plan remains an attractive goal.

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.

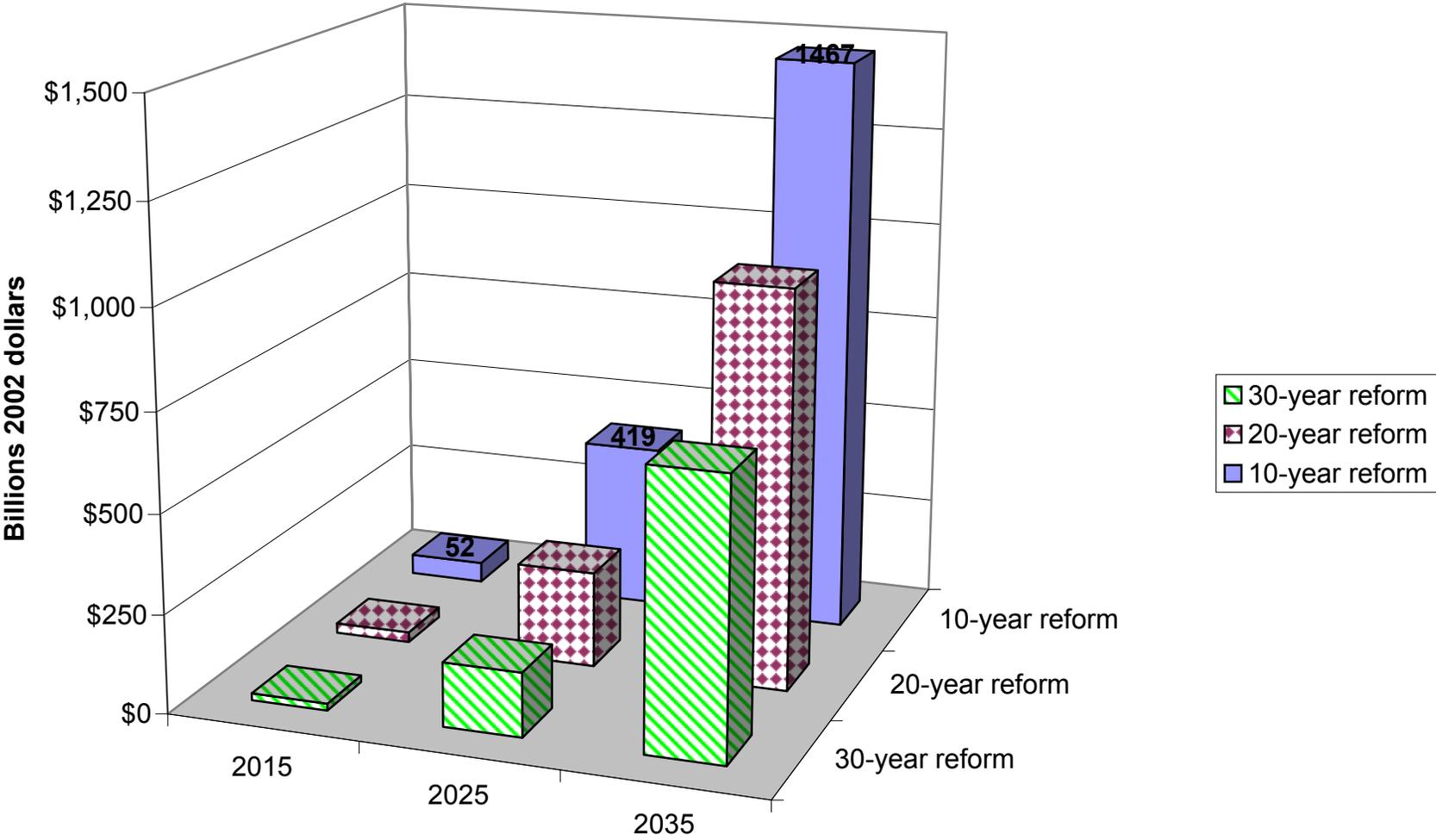
Feasible Teacher Quality Policies

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 (Appendix Figure A1).

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 (2003)). 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

Figure 4. Growth Dividend from 0.5 s.d. Reform Begun in 2005



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.¹⁵

Rivkin, Hanushek, and Kain (2001) 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, and Rivkin (2004; (forthcoming, Spring 2004 Spring 2004 Spring 2004 Spring 2004 Spring 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.¹⁶ In the end, it estimates that the differences in annual achievement growth

¹⁵ 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.

¹⁶ 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.

between an average and a good teacher are at least 0.11 standard deviations of student achievement.¹⁷

Before going on, it is useful to put this estimate of the variation in quality into perspective. 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 student not on free or reduced lunch.

A reasonable estimate (which is used throughout the following calculations) is actually that differences in quality are twice that lower bound (0.22 s.d.). This larger estimate reflects likely differences in teacher quality among schools (plus a series of other factors that bias the previously discussed estimate downwards).

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 or a 1.0 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 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 each year of the reform period. By maintaining this standard for replacement of all teachers exiting teaching (6.6 percent annually in 1994-95) but retaining all other teachers, this policy 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

¹⁷ 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.”

the 7.2 percent who change schools), a 0.5 s.d. improvement in student performance could be achieved in 10 years.

Figure 5 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 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.

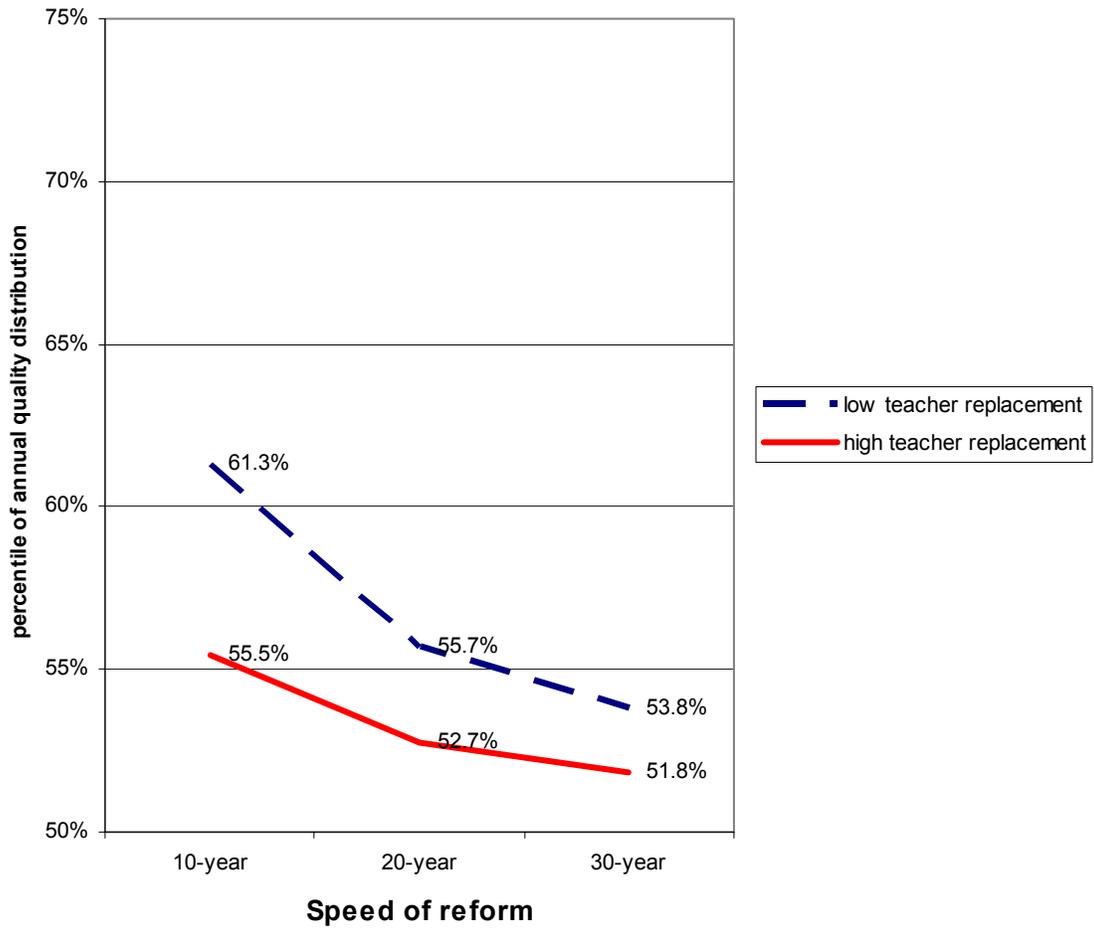
Figure 6 displays the same information for a more ambitious 1.0 standard deviation improvement. Clearly the loftier goals require higher standards for new hires. For example, a 10-year reform program with low turnover now requires annual hiring at the 72nd percentile.

These calculations demonstrate the challenge of achieving substantial improvements in achievement. It requires significantly upgrading the quality of the current teacher force.

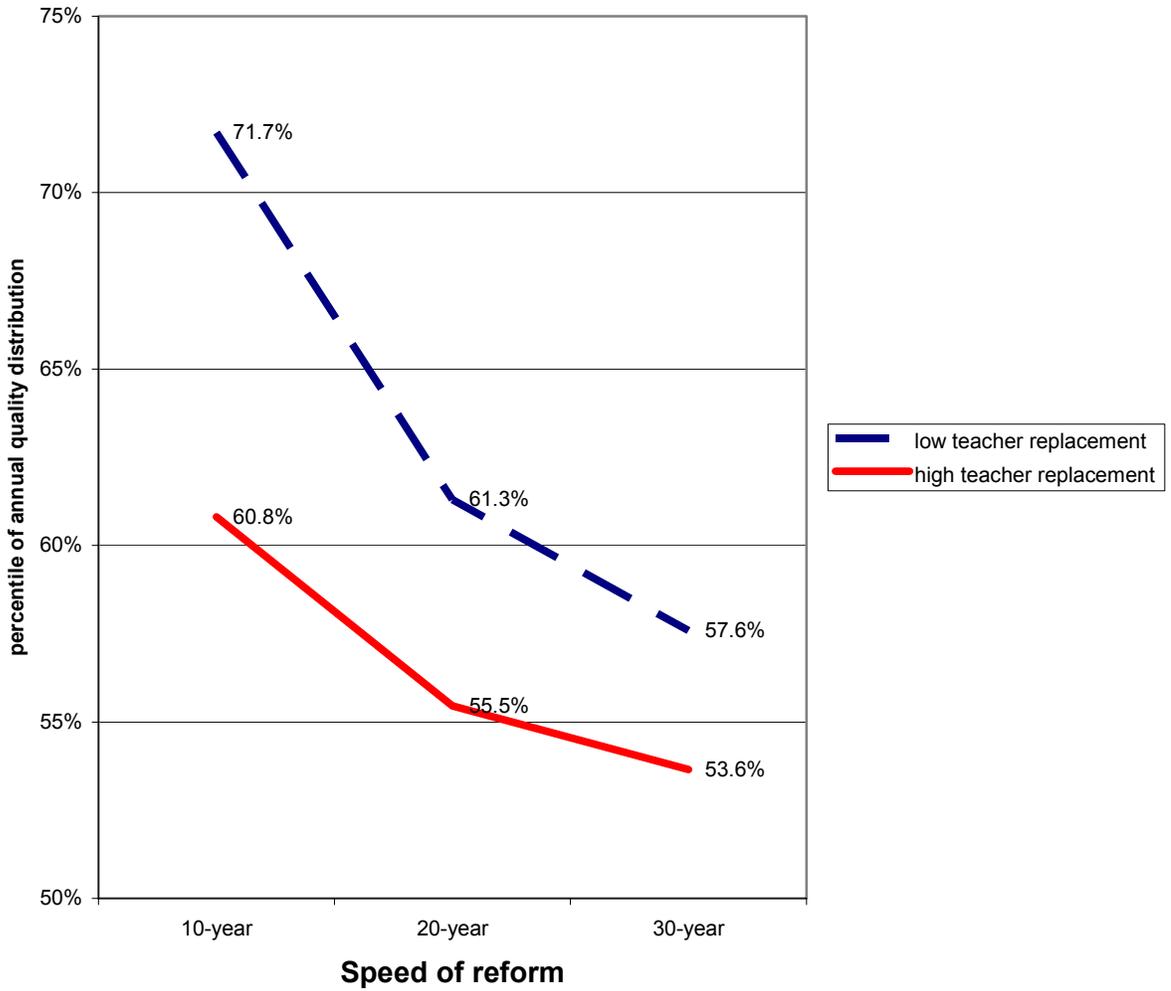
Several aspects of these scenarios deserve note. First, the improvements that are required apply to the teacher distribution that exists *each* year. In other words this standard requires continual improvement in terms of the current teachers. The continual improvement comes from the fact that the distribution 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

**Figure 5. Annual Required Hiring Percentile
(0.5 s.d. Reform)**



**Figure 6. Annual Required Hiring Percentile
(1.0 s.d. Reform)**



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.

The calculations also freeze many aspects of teaching. They assume no change in teacher turnover. Of course, teacher turnover will be affected by a variety of other policies 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 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.

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 and 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 and Rivkin (1997, (2004))).¹⁸ For a variety of reasons, however, this does not give much policy guidance for the current discussions. In simplest terms, we do not know how teacher quality responds to different levels of salaries (Hanushek and Rivkin (2004)). Moreover, policies that simply raised 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 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 2025.¹⁹ Of course, as shown previously, a reform program of this magnitude and speed would require dramatic changes in hiring of new teachers.

¹⁸ There is a current debate about how salaries of teachers compare to those in different professions; see Podgursky (2003).

¹⁹ These calculations assume that K-12 expenditures grow at 3 percent (real), implying that the current \$350 billion expenditure would grow to \$777 billion in 2025.

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.

Figure 7 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.²⁰ 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.²¹ This magnitude is identical to the annual reform dividend from growth in 2025 from a 30-year reform yielding a 0.25 standard deviation improvement.

But, teacher bonuses can be considered from another perspective. A one standard deviation improvement in performance raises individual worker salaries (not counting any growth effects) by around \$3,600 per year (figure 1); a half standard deviation reform increases earnings by \$1,800 per year. This annual addition to earnings of the smaller reform (0.5 s.d.) translates into a present value of \$30,000 for each student.²² A bonus to a teacher of \$12,500 per year could then be recouped in increased student earnings with a pupil-teacher ratio of six or more, as long

²⁰ These calculations assume that the current average teacher salary is \$50,000 – a figure close to the National Education Association survey data.

²¹ 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.

²² These calculations assume 35 years of working life and a 5 percent net discount rate. The net discount rate represents the 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.

as the bonuses elicited at least a 0.5 standard deviation improvement in student skills.²³ 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.²⁴

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).²⁵ If generalizable, this program at even half the performance result would be economically justified by either gains in individual earnings or aggregate effects.

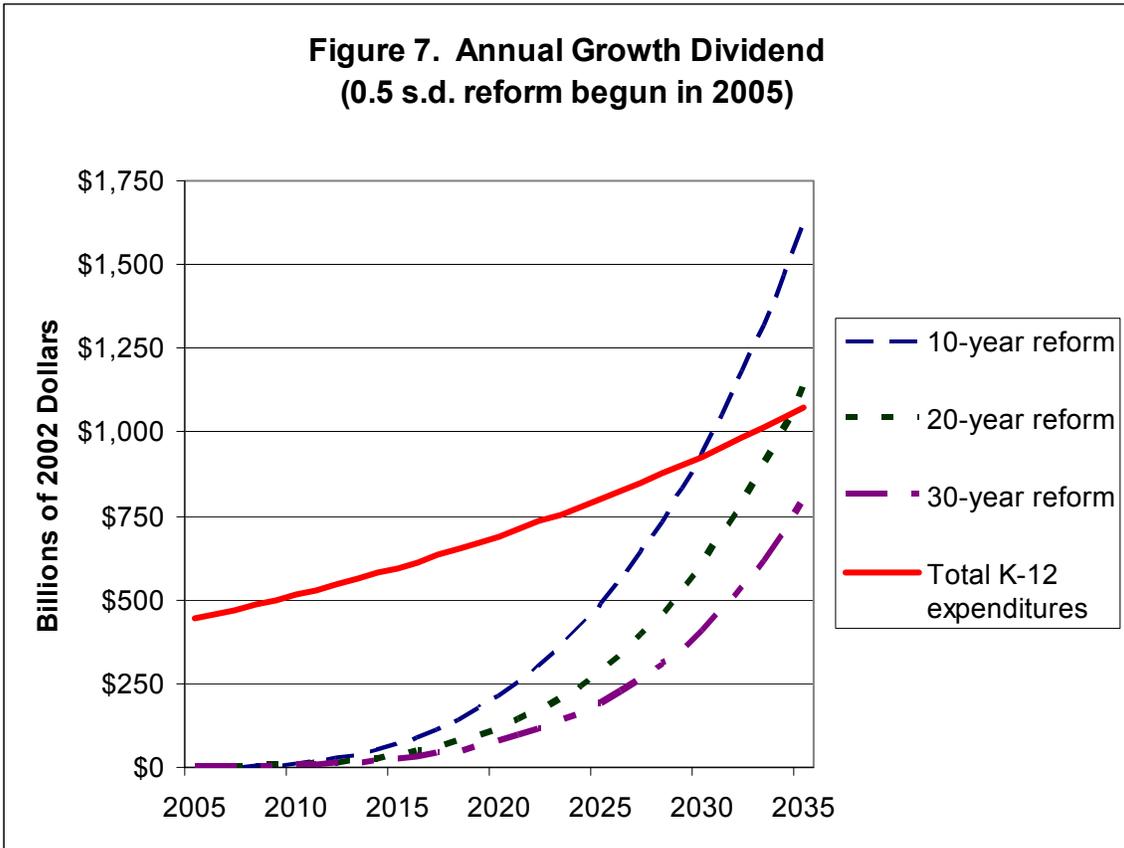
Another evaluation is found in an experiment in Israel (Lavy, 2002). Schools were placed in competition with each other, and teachers in the highest performing schools received salary bonuses. These salary bonuses, given to the entire school faculty, were rather modest

²³ 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.

²⁴ 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 and others (1994).

²⁵ This program is currently in its initial phases and the evaluation is on-going. Some preliminary results can be found in Shacter et al. (2003). Cost figures come from private correspondence.

**Figure 7. Annual Growth Dividend
(0.5 s.d. reform begun in 2005)**



(approaching three percent at the top). Nonetheless, schools competing for bonuses did better than another set of schools that just received resources.²⁶ This program shows that schools react to incentives, but it is unclear how to translate that into costs and benefits for a set of U.S. 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*. U.S. 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.

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).²⁷ 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, improvements in early reading could well lessen the costs of special education

²⁶ 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.

²⁷ See, for example, Black, Devereux, and Salvanes (2003) for a suggestive analysis of Norwegian experiences.

(Lyon and 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.

Conclusions

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*.

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 will require a significant upgrading of the teaching force. It also discusses feasible timing and speed of reform.

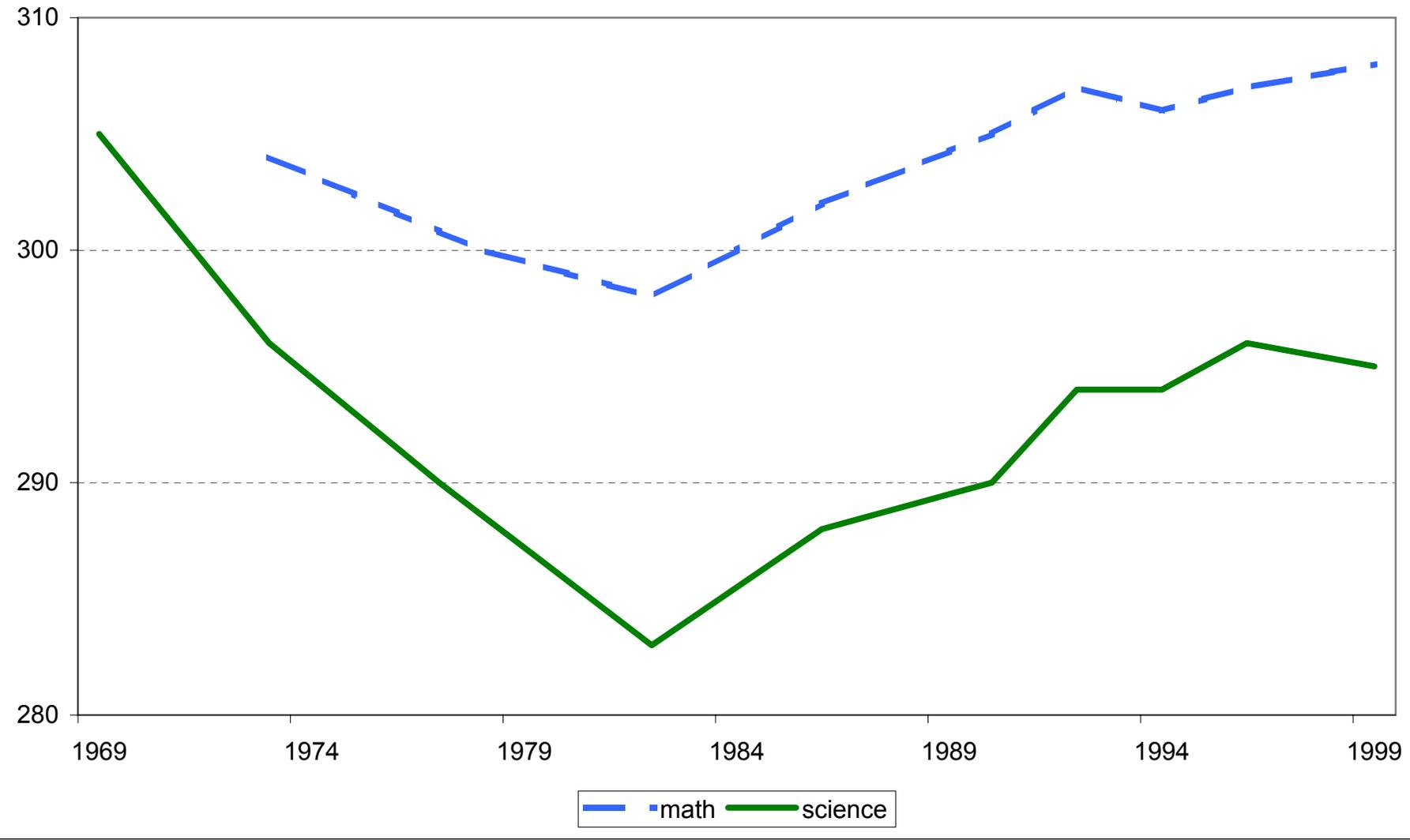
The benefit picture indicates that improvements in student performance have truly substantial impacts on individual productivity and earnings and on the growth and performance of the aggregate economy. The economic gains could in fact cover some substantial changes in expenditure on schools.

Past history, however, provides a key caution. The U.S. 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.

Figure A1.
National Assessment of Educational Progress (NAEP), age 17



References

- Altonji, Joseph G., and Charles R. Pierret. 2001. "Employer learning and statistical discrimination." *Quarterly Journal of Economics* 116,no.1 (February):313-350.
- Barro, Robert J., and Xavier Sala-I-Martin. 1995. *Economic Growth*. New York: McGraw-Hill, Inc.
- Behrman, Jere R., Lori G. Kletzer, Michael S. McPherson, and Morton Owen 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.
- Bishop, John. 1989. "Is the test score decline responsible for the productivity growth decline?" *American Economic Review* 79,no.1:178-197.
- . 1991. "Achievement, test scores, and relative wages." In *Workers and their wages*, edited by Marvin H. Koster. Washington, DC: The AEI Press:146-186.
- Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2003. "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital." NBER Working Paper No. w10066, National Bureau of Economic Research (November).
- Blackburn, McKinley L., and David Neumark. 1993. "Omitted-ability bias and the increase in the return to schooling." *Journal of Labor Economics* 11,no.3 (July):521-544.
- . 1995. "Are OLS estimates of the return to schooling biased downward? Another look." *Review of Economics and Statistics* 77,no.2 (May):217-230.
- Card, David. 1999. "Causal effect of education on earnings." In *Handbook of labor economics*, edited by Orley Ashenfelter and David Card. Amsterdam: North-Holland:1801-1863.
- Cohen, David K., and Richard J. Murnane. 1986. "Merit pay and the evaluation problem: Understanding why most merit pay plans fail and a few survive." *Harvard Educational Review* 56,no.1 (February):1-17.
- Dugan, Dennis J. 1976. "Scholastic achievement: its determinants and effects in the education industry." In *Education as an industry*, edited by Joseph T. Froomkin, Dean T. Jamison and Roy Radner. Cambridge, MA: Ballinger:53-83.
- Easterly, William. 2002. *The elusive quest for growth: An economists' adventures and misadventures in the tropics*. Cambridge, MA: The MIT Press.

- Greene, Jay 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 (September).
- Griliches, Zvi. 1974. "Errors in variables and other unobservables." *Econometrica* 42,no.6 (November):971-998.
- Grogger, Jeffrey T., and Eric Eide. 1993. "Changes in college skills and the rise in the college wage premium." *Journal of Human Resources* 30,no.2 (Spring):280-310.
- Hanushek, Eric A. 2002. "Publicly provided education." In *Handbook of Public Economics*, edited by Alan J. Auerbach and Martin Feldstein. Amsterdam: Elsevier:2045-2141.
- . 2003. "The Importance of School Quality." In *Our Schools and Our Future: Are We Still at Risk?*, edited by Paul E. Peterson. Stanford, CA: Hoover Institution Press:141-173.
- Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2004. "The revolving door." *Education Next* 4,no.1 (Winter):77-82.
- . forthcoming, Spring 2004. "Why public schools lose teachers." *Journal of Human Resources* 39,no.2.
- Hanushek, Eric A., and Dongwook Kim. 1995. "Schooling, labor force quality, and economic growth." Working Paper 5399, National Bureau of Economic Research (December).
- Hanushek, Eric A., and Dennis D. Kimko. 2000. "Schooling, labor force quality, and the growth of nations." *American Economic Review* 90,no.5 (December):1184-1208.
- Hanushek, Eric A., and with others. 1994. *Making schools work: Improving performance and controlling costs*. Washington, DC: Brookings Institution.
- Hanushek, Eric A., and Richard R. Pace. 1995. "Who chooses to teach (and why)?" *Economics of Education Review* 14,no.2 (June):101-117.
- Hanushek, Eric A., and Steven G. Rivkin. 1997. "Understanding the twentieth-century growth in U.S. school spending." *Journal of Human Resources* 32,no.1 (Winter):35-68.
- . 2004. "How to improve the supply of high quality teachers." In *Brookings Papers on Education Policy 2004*, edited by Diane Ravitch. Washington, DC: Brookings Institution Press.

- Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor. 1996. "Aggregation and the estimated effects of school resources." *Review of Economics and Statistics* 78,no.4 (November):611-627.
- Heckman, James J., and Edward Vytlačil. 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,no.1 (February):1-12.
- Lazear, Edward P. 2003. "Teacher incentives." *Swedish Economic Policy Review*.
- Lyon, G. Reid, and Jack M. Fletcher. 2001. "Early warning system: How to prevent reading disabilities." *Education Matters* 1,no.2 (Summer):23-29.
- Manski, Charles F., and David A. Wise. 1983. *College choice in America*. Cambridge: Harvard University Press.
- Mulligan, Casey B. 1999. "Galton versus the human capital approach to inheritance." *Journal of Political Economy* 107,no.6, pt. 2 (December):S184-S224.
- Murnane, Richard J., John B. Willett, M. Jay Braatz, and Yves 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,no.4 (August):311-320.
- Murnane, Richard J., John B. Willett, Yves Duhaldeborde, and John H. Tyler. 2000. "How important are the cognitive skills of teenagers in predicting subsequent earnings?" *Journal of Policy Analysis and Management* 19,no.4 (Fall):547-568.
- Murnane, Richard J., John B. Willett, and Frank Levy. 1995. "The growing importance of cognitive skills in wage determination." *Review of Economics and Statistics* 77,no.2 (May):251-266.
- National Commission on Excellence in Education. 1983. *A nation at risk: the imperative for educational reform*. Washington, D.C.: U.S. Government Printing Office.
- Neal, Derek A., and William R. Johnson. 1996. "The role of pre-market factors in black-white differences." *Journal of Political Economy* 104,no.5 (October):869-895.
- O'Neill, June. 1990. "The role of human capital in earnings differences between black and white men." *Journal of Economic Perspectives* 4,no.4 (Fall):25-46.
- Peterson, Paul E., ed. 2003. *Our schools and our future: Are we still at risk?* Stanford, CA: Hoover Press.
- Podgursky, Michael. 2003. "Fringe benefits." *Education Next* 3,no.3 (Summer):71-76.

Rivkin, Steven G. 1995. "Black/white differences in schooling and employment." *Journal of Human Resources* 30,no.4 (Fall):826-852.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2001. "Teachers, schools, and academic achievement." Working Paper No. 6691, National Bureau of Economic Research (revised)