Conclusions and Controversies about the Effectiveness of School Resources

Eric A. Hanushek

Both the U.S. public and U.S. policymakers pursue a love-hate relationship with U.S. schools. While a majority of parents believe that their children's schools are doing well, a majority also believe that the system as a whole needs help. Complicating this view is a variety of concerns about specific aspects of U.S. schools—they are too expensive, too rigid, too elitist, and too unequal.

During the past year, President Clinton has directed considerable government attention to U.S. education. This attention follows the lead of Presidents Bush and Reagan, who also focused on education policy, although the oversight of such policy is not the primary role of the federal government. President Bush, for example, in 1989 convened a historic gathering of the governors of all of the states to focus exclusively on issues of education. The governors set a series of lofty goals for the year 2000, including the goal that U.S. students should be first in the world in mathematics and science achievement. Unfortunately, we are now close to the year 2000, but we are not close to meeting the set goals.

This paper analyzes the current state of the education system in the United States. In the course of the paper, I will try to point out where controversy exists, particularly in academic discussions.

Overview

I begin with some overall observations and conclusions. The subsequent discussion will provide some of the relevant evidence and references to support my conclusions.

As a starting point, educational investments are very important to the U.S. economy, a fact that suggests there is much value in an aggressive human capital investment strategy. The U.S. economy has been built up largely by using a skilled labor force and has capitalized on the presence of skills, making human capital investments very important to the success of the overall economy. Moreover, many authors show that the labor market value of the increased skills, as measured by schooling level, has increased dramatically in

Eric A. Hanushek is a professor of economics and public policy and director of the W. Allen Wallis Institute of Political Economy at the University of Rochester.

recent years. I think this valuation demonstrates that the economy continues to need an increasingly skilled labor force. Recent work has also suggested that education is very important in boosting the growth rates of the nation as a whole and that a very important relationship exists between human capital and growth rates. Economists have recently spent considerable time and effort trying to understand why some countries grow faster than others. The majority opinion is that a nation's stock of human capital is an important component of differential growth rates. In addition, we have thought of education as a primary ingredient in providing equal opportunity to all members of society as a way of cutting down or breaking intergenerational correlations of income. Taken together, these benefits provide important and relatively uncontroversial reasons for us to continue our attention to education.

The controversies relate in small measure to how well we have been doing in providing education, but they relate more to what we should do in the future. My way of framing the issues follows.

First, U.S. students do not perform well compared with students from other countries. In international math and science exams, U.S. students have never performed very well relative to students of other countries. To compensate for this relatively low quality, the United States has historically had high levels of school attainment (years of schooling)—that is, the United States has substituted quantity for quality. Now, however, many countries that have had higher student achievement are beginning to rival the United States on quantity grounds. This suggests that the U.S. economy faces new and different levels of competition in the years ahead.

Second, the United States has made steady and large investments in human capital. The resources invested, however, have had little payoff in terms of student performance. Thus, if the United States is to be more competitive internationally in terms of student achievement, some substantially different policies will be required in the future.

Third, the most likely changes required in schools involve radically different incentives for students and for school personnel. Few direct incentives exist today for improved student achievement, and marginal changes in resources or programs are unlikely to have a noticeable effect on overall student achievement.

Fourth, improved education policies will require better measurement of student performance. In addition, such policies will probably require a period of more extensive experimentation with alternative approaches and incentive schemes.

These conclusions are roughly ordered in terms of the amount of evidence and analysis that we have on them and in terms of the amount of consensus or controversy that exists. Regardless of one's views on the underlying controversies, these conclusions indicate to me that the education sector deserves considerable attention. At the same time, the form of this attention is important. Some people have argued that the high rates of return commonly observed for individual schooling clearly justify governmental action. But the case for governmental involvement in education, as opposed to purely private decision making on schooling, requires more than that. Governmental intervention is frequently justified on the basis of external benefits, benefits that go beyond an individual's investment in schooling. Are there external benefits to investing in education? Education is often thought to be a "large externality" undertaking, but identification and measurement of those externalities have proved difficult.[1] My candidate for the most important potential external benefit from investing in education in the United States—which is new in most thinking—is the overall effect on growth rates and the potential to affect the economy. The work supporting this contention is not as refined as you might like: it does not give precise answers, and there are several qualifications. Nonetheless, I think that growth effects are likely to prove to be a very important policy issue.[2] At the same time, while establishing a role for government research, this issue does not spell out what such a role should be.

U.S. Student Performance

It is useful to begin with the performance of students in the U.S. educational system. In doing so, it is natural to contrast performance in elementary and secondary education with that in higher education. I begin with elementary and secondary education. In terms of quality of learning, U.S. schools are not now, and have never been, very competitive when judged by the performance of elementary and secondary schools around the world. Chart 1, drawn from Hanushek and Kim (1996), pre-

Chart 1

International Test Score Performance, by Test Year

Source: Hanushek and Kim (1996).

sents what we know about all international testing of math and science scores for U.S. students.

International examinations in mathematics and science have been given periodically since the 1960s. The examinations have been taken on a voluntary basis by a variable set of countries. While there was some concern about selective test taking in some countries in the early years, that concern has lessened considerably in the later years of testing. Further, Hanushek and Kim show that these tests have considerable validity in describing the quality of a country's labor force. For the analysis here, all the test scores for students in a given country in a given year are combined to produce a single country test score. The scores are placed on a scale where the world mean for each testing year is fifty.

In Chart 1, the year of testing appears along the top of the chart. Normalized scores are given on the vertical axis, making it possible to compare countries over time.

The U.S. performance moves around over time. This drift closely mirrors the average performance of U.S. seventeen-year-olds on the mathematics and science tests of the National Assessment of Educational Progress (NAEP) (see discussion below). Moreover, the key aspect of this figure is that the United States almost always falls below the median of whatever group of countries is taking the test.

The results released in the fall of 1996 for the Third International Math and Science Test placed U.S. eighth graders in the middle of world performance for 1994-95.[3] This performance, which is not included in the figure, comes even though a very wide range of forty-one countries participated in the testing. Thus, there is no real change in the latest scores.

The basic story is that the United States has not been doing particularly well in international comparisons. This result is a bit surprising, given that the United States has an economy built on a skilled labor force. You might ask, "How could that be?" While the United States is not doing well, it is producing skilled goods that one might argue require a skilled labor force.

The answer seems to be that over a long period of time, quantity of schooling has substituted for quality. Historically, the United States has had a labor force with more years of schooling, on average, than the labor forces of other countries, even if these years of schooling have been of lower quality.

That quantitative superiority is ending. Table 1 compares the percentage of students in different countries that have received upper secondary school education, essentially a high school education. These completion rates are broken down by age.

The important part of breaking these figures down by age is that they can be read as the schooling policies of countries in different years. Individuals who are twenty-five to thirty-four years old in 1992 were educated sometime in the 1980s. People aged thirty-five to forty-four were educated in the 1970s. The next group in the table was educated in the 1960s. And the final group went to school in the 1950s.

If we look at the 1980s, it is clear that a large number of countries are rivaling the United States, where 87 percent of students complete their high school education. Three other countries in the Group of Seven have completion rates exceeding 80 percent. Of the countries outside the

Table 1
PERCENTAGE OF POPULATION ATTAINING UPPER SECONDARY EDUCATION OR MORE
By Country, 1992

	Age Group			
Country	25-34	35-44	45-54	55-64
GROUP OF SEVEN				
Canada	81	78	66	49
France	67	57	47	29
Germany	89	87	81	69
Italy	42	35	21	12
United Kingdom	81	71	62	51
United States	87	88	83	73
OTHER				
Australia	57	56	51	42
Austria	79	71	65	50
Belgium	60	52	38	24
Czechoslovakia	87	79	68	51
Denmark	67	61	58	45
Finland	82	69	52	31
Ireland	56	44	35	25
Netherlands	68	61	52	42
New Zealand	60	58	55	49
Norway	88	83	75	61
Portugal	21	17	10	7
Spain	41	24	14	8
Sweden	83	76	65	48
Switzerland	87	84	78	70
Turkey	21	14	9	5

Source: U.S. Department of Education (1996b).

Group of Seven listed in the table, another five have completion rates above 80 percent. These numbers contrast sharply with those in earlier decades, when the United States had a very dramatic lead in terms of quantity of schooling. Clearly, other OECD countries and developing countries have dramatically increased the amount of schooling their youth receive. The United States' advantage in quantity of schooling is quickly disappearing.

Charts 2 and 3 provide pictures of science and mathematics achievement in the United States as measured by the NAEP, which is currently the best yardstick of student performance. The heavy line reflects the average scores of seventeen-year-olds on the NAEP over time. What we see from Chart 2 is that today our students are not doing quite as well in science as they did in 1970 (even though, as described below, we have been increasing real per pupil spending steadily over this period). Chart 3 shows essentially the same thing for math, except that instead of declining, 1996 performance is at about the same level as in 1970. This picture does not lead anyone to believe that our investment policy is soon going to address the quality concerns and to push us to the top of the international rankings. "First in the world in math and science in the year 2000" was the goal set forth by the 1989 National Governors' Conference. It does not look like we are on that path.

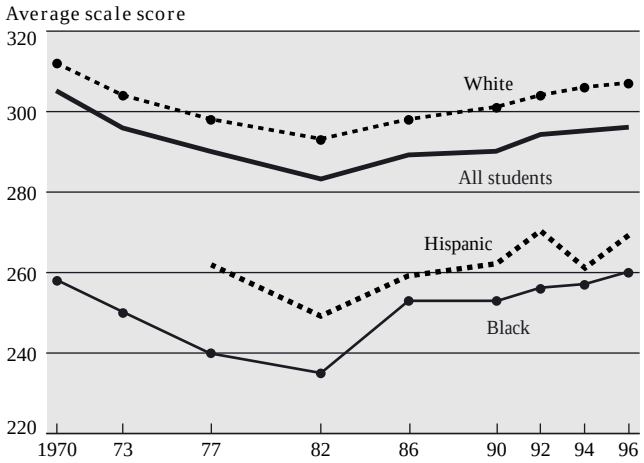
Charts 2 and 3 also suggest that there is a substantial gap between whites on the one hand, and blacks and Hispanics on the other. The gap has narrowed some, but it remains substantial and may have even widened in the most recent period. This disparity goes back to the equality of opportunity concerns; it is also consistent with several analyses that identify the importance of student achievement in explaining some of the college attendance gaps across different segments of society. Those attendance gaps exist throughout this period and seem related to quality of schooling.

The situation with higher education is very different. U.S. higher education is arguably the best in the world. Admittedly, data on higher education are not nearly as good as the data on elementary and secondary education. It is particularly hard to document quality because we do not have good, objective measures. Here is what we do know:

- U.S. business and industry are now willing to pay much more for college graduates than they were in the past, both in relative and in absolute terms;
- foreign students like to come to U.S. higher education institutions, while they do not seem to want to come to U.S. elementary and secondary schools; and
- employers seem much more pleased, at least in their public testimonials, with higher education than they are with elementary and secondary education.

Chart 2

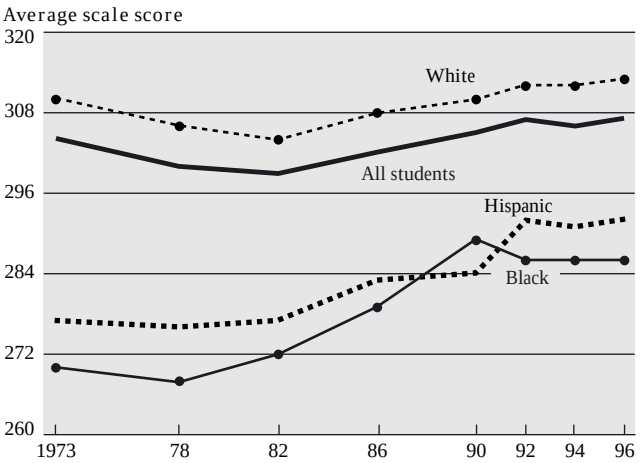
Science Achievement
Seventeen-Year-Olds, by Race/Ethnicity



Note: Science achievement is measured by performance on the National Assessment of Educational Progress (NAEP) exam.

Chart 3

Mathematics Achievement
Seventeen-Year-Olds, by Race/Ethnicity



Note: Mathematics achievement is measured by performance on the National Assessment of Educational Progress (NAEP) exam.

This adds up to a *prima facie* case that quality does not appear to be a major problem in higher education. By contrast, the data displayed before suggest that quality is the major concern in elementary and secondary education. The contrasting picture makes the recent concentration in 1997 on higher education by the President and the Congress puzzling—at least if the policy initiatives are viewed in terms of education as opposed to pure distributional politics. Perhaps the one rationalization is that the call for expanding access to schools— "making the fourteenth year the norm"—is just an extension of the historic policy of substituting quantity for quality. Without pursuing the issues of higher education, I simply assert that elementary and secondary school issues are the most important and pressing. Thus, the remainder of this discussion concentrates exclusively on these issues.

A Brief History of Spending Growth

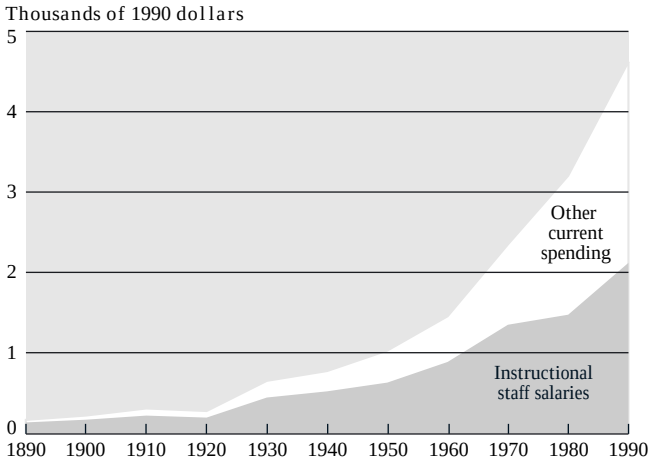
The United States has had a consistent focus on education over a long period of time. This fact surprises many people in the United States. Statements about "how important it is that President Clinton has recently focused attention on education" are common. Implicit or explicit in such discussions is the sentiment that we have been shortchanging the educational system. It may be that the President can get the attention of the population better than anybody else, but a steady policy thrust and heavy weight have been given to education and human capital investment for a long time. This focus on education, however, has not been at the federal government level.[4] Taking the long view, between 1890 and 1990, we note that real public expenditure on primary and secondary education in the United States rose from $2 billion to more than $187 billion.[5] Significantly, this almost hundredfold increase is more than triple the growth rate of GNP during the same period: current educational expenditure increased from less than 1 percent of GNP in 1890 to 3.4 percent of GNP in 1990.[6]

While increasing enrollment accounts for a portion of the rise in spending, the rise in per student expenditure explains the bulk of the change in educational outlays (Chart 4). Real per student expenditure roughly quintupled in each fifty-year period between 1890 and 1980: it went from $164 in 1890 to $772 in 1940, and to $4,622 in 1990. If we divide per student expenditure into salaries for instructional staff (teachers and principals) and all other expenditure, we find that the unmistakable pattern here is the relative growth of expenditure outside of instructional staff salaries: such spending went from 25 percent of total current expenditure in 1890 to 33 percent in 1940, and to 54 percent in 1990.

Two factors stand out as being of primary importance in explaining total instructional salary spending over the entire 100-year period: the rising price of instructional staff and the declining pupil-staff ratio. Rising teacher salaries were clearly a consequence of economywide labor productivity growth, although the extent to which teacher salaries changed relative to those of other workers is an important issue. By contrast, the decisions leading to reductions in the pupil-staff ratio despite the rise in teacher costs suggest a long-term policy of attempting to raise school quality by reducing the pupil-teacher ratio.[7] There is substantial debate over the extent to which external changes, notably the expansion of special education, contributed to the decline in the pupil-teacher ratio during the 1970s and 1980s. The analysis by Hanushek and Rivkin (1997) indicates that special education has been important but is still not the largest influence. The growth in special education over the 1980s may have accounted for one-

Chart 4

Spending per Pupil, 1890-1990



Source: Hanushek and Rivkin (1997).

fifth of the growth in spending. (Yet, because of the smaller overall spending growth in the 1990s, this percentage has almost certainly gone up.)

Price and Quality of Teachers

The desire to reduce classroom sizes has increased the demand for teachers. At the same time, rising labor market wages for both college-educated men and, particularly, college-educated women have reduced teacher supply. These supply and demand movements offer a straightforward explanation for the teacher price increase from $34 per day in 1890 to more than $177 per day in 1990, an increase that accounts for more than 40 percent of the increase in total expenditure on instructional staff over the century. But these numbers tell only part of the story. Schools have also been able to adjust the average quality of teaching personnel by shifting teacher salaries to accommodate shifts in supply and demand. Since potential teachers clearly differ in their skills and consequently their alternative wage opportunities, any increase or decrease in teacher wages beyond that occurring in other sectors reflects a change in where teachers are drawn from the distribution of workers. This fact would be expected to influence teacher quality in the long run.[8]

To trace teacher quality changes, I use annual earnings data for teachers from the six decennial censuses of population taken between 1940 and 1990.[9] Teacher earnings are compared with the earnings of those who do not teach. Specifically, our primary measure of potential teacher quality is the location of average teacher earnings in the distribution of nonteacher earnings.[10] The lower the percentage of nonteachers who earn less than the average teacher, the worse the teaching jobs when compared with alternative occupations. The use of percentile rankings as opposed to a comparison of mean earnings reduces problems associated with the census' top-coding of incomes and lessens the impact of changes in the tails of the nonteacher earnings distribution.

The movements in relative earnings of teachers have been dramatic. As shown in Table 2, however, they differ noticeably for men and women. While the average male teacher earned more than 84 percent of all males in 1940, this figure fell to 64 percent by 1990. All of this relative fall, however, occurred before 1960; following a slight dip in the 1970s, male teachers have been moving up the earnings distribution. The overall decline in the relative position of women teachers has been almost as large, although female teachers are still better positioned in the earnings distribution than male teachers. The time path of the decline for females has been very different, however, with the largest declines occurring after 1970, when the average teacher moved a full 10 percentage points down the earnings distribution.

Table 2
Relative Salaries of U.S. Teachers
By Gender, 1940-90

	Percentage of Male Nonteachers Earning Less Than Average Male Teacher		Percentage of Female Nonteachers Earning Less Than Average Female Teacher	
Year	All Workers	College Graduates	All Workers	College Graduates
1940	84.0	52.5	92.3	68.7
1950	73.4	36.2	86.7	55.0
1960	63.3	28.7	86.9	52.7
1970	62.2	25.7	85.8	47.1
1980	53.0	31.0	77.7	50.1
1990	64.0	36.5	75.1	45.3

Source: U.S. decennial census of population, public use microdata, 1940-90.

The implication of this finding is that schools, while spending increasing amounts on teachers, have also tended to let the quality of teachers slip. This story appears to be closely related to arguments such as those of Baumol (1967). He suggests that a sector subject to low rates of productivity improvement, frequently for technological reasons, will find its wage bill and costs rising relative to those of more advanced sectors. While there are reasons to be skeptical about the magnitude of any such effects (Hanushek 1997), the increasing labor costs and potentially declining teacher quality are consistent with this argument.

Recent Changes in School Resources

The search for more experienced and better educated teachers has been one of the enduring policy thrusts of the last half century. This thrust, along with the desire to reduce class sizes, has been a dominant component of aggregate changes in schools.

Table 3 tracks these changes from 1960 to 1991. Teacher education has increased dramatically, so that more

than half of all teachers had master's degrees in 1991. Moreover, teacher experience, while following some demographic cycles, has reached a very high level. Table 3 also shows the decline in pupil-teacher ratios and the increase in real spending per pupil. Because teacher salaries are closely linked to experience and education, and because variations in salaries and pupil-teacher ratios are the most important determinants in spending per pupil, the added real resources directly drive spending. Between 1960 and 1991, real spending per pupil almost tripled.

It takes little effort to see the contrast between the growing resources in Table 3 and the flat student performance in Charts 2 and 3. At the very least, this contrast suggests the possibility that something is very wrong.

Other Observations about School Efficiency

The preceding sections make a strong case that spending and policy toward schools in the United States have not been very well directed. Specifically, spending has improved dramatically, but student performance—at least over the period in which it has been measured—has not improved. This outcome is the very essence of inefficiency: more resources devoted to schools have not improved output.

The aggregate data, however, could be misleading. First, costs not directly related to the typical student—for example, costs for special education—might figure importantly in the rise in spending. Second, the other influences on student performance—families and friends—could have been less favorable over time. For example, some analysts note that single-parent families have increased in recent decades. Also, the percentage of students in families below the poverty level has increased. Factors such as these could mean that more school resources are needed to overcome existing deficits.

These issues have been debated in considerable detail elsewhere, and there is, in my judgment, a *prima facie* case against them. Nonetheless, it is also worth mentioning the detailed microlevel evidence about the impact of resources. The interpretation of the microlevel evidence has been the most controversial part of the academic debate; these controversies are less relevant to the policy debate.

The investigation of the effects of school resources began in earnest with the publication of the "Coleman Report" (Coleman et al. 1966). This congressionally mandated study by the U.S. Office of Education startled many observers by suggesting that schools did not exert a very powerful influence on student achievement. Subsequent attention was directed at providing additional evidence about the effects of resources.

Over the past thirty years, a steady stream of analyses has built up a consistent picture of the educational process. Studies of educational performance, generally following statistical analyses of the determinants of student achievement, include a variety of different measures of resources devoted to schools. Commonly employed measures include (1) the real resources of the classroom (teacher education, teacher experience, and teacher-pupil ratios); (2) financial aggregates of resources (expenditure per student and teacher salary); and (3) measures of other resources in schools (specific teacher characteristics, administrative inputs, and facilities).

The real resource category receives the bulk of attention for several reasons. First, this category best summarizes variations in resources at the classroom level. Teacher education and teacher experience are the primary determinants of teacher salaries. When combined with

Table 3
PUBLIC SCHOOL RESOURCES IN THE UNITED STATES, 1961-91

Resource	1960-61	1965-66	1970-71	1975-76	1980-81	1985-86	1990-91
Pupil-teacher ratio	25.6	24.1	22.3	20.2	18.8	17.7	17.3
Percentage of teachers with master's degree	23.1	23.2	27.1	37.1	49.3	50.7	52.6
Median years of teacher experience	11	8	8	8	12	15	15
Current expenditure per pupil (1992-93 dollars)	1,903	2,402	3,269	3,864	4,116	4,919	5,582

Source: U.S. Department of Education (1996a).

Note: Per pupil expenditures are based on students' average daily attendance.

teachers per pupil, these variables describe variations in the instructional resources across classrooms. Second, these measures are readily available and well measured. Third, they relate to the largest changes in schools over the past three decades. Table 3 displays the dramatic increase in these school inputs, with pupil-teacher ratios falling steadily, teacher experience increasing, and the percentage of teachers with a master's degree actually doubling between 1960 and 1990. Fourth, studies of growth in performance at the individual classroom level, commonly thought to represent the superior analytical design, frequently have these resource measures, but not the others, available.

These studies yield a simple conclusion, one that is supported in detail elsewhere (Hanushek 1997): there is no strong or consistent relationship between school resources and student performance. In other words, there is little reason to be confident that simply adding more resources to schools as currently constituted will yield performance gains among students. Studies of class size and pupil-teacher ratios, of teacher education, and of teacher experience give little if any support to policies of expanding these resources. This finding has obvious policy implications. Before turning to these, it is useful to clarify precisely what is and is not implied by the data.

Perhaps the most important fact to underscore is that this finding does not imply that all schools and teachers are the same. Quite the contrary. Substantial evidence suggests that there are large differences among teachers and schools. The simple fact remains that these differences are not closely related to teacher salaries or to other measured resources devoted to programs. The Coleman Report, which found that measured school resources explained a small portion of the variance in student achievement, has been commonly interpreted as implying that "schools don't make a difference." This interpretation confuses the effects of measured differences with the full effects of schools and has been shown to be wrong. There is a significant difference between measured resources (of the kind on which policy frequently focuses) and the true effects of schools.[11] In fact, it is just this difference between true effects and those of standard resources that forms the basis for the policy considerations below.

The preceding interpretations of the general ineffectiveness of school resource policies has been challenged by some researchers. Two separate challenges deserve attention before I discuss policy implications.

Labor Market Outcomes

Taken as a group, the production function studies give little indication that variations of resources have anything to do with variations in student performance. However, the widely publicized findings of Card and Krueger (1992) indicate that variations in school resources are related to earnings differences among workers.[12] The Card and Krueger analysis begins with samples of adult workers from the 1970 and 1980 censuses of population and fills in information about the schooling circumstances of individuals by using information about their year and state of birth. Card and Krueger find that labor market earnings are directly related to school resource differences.

Several factors could contribute to reconciling these conclusions: differences in levels of resources considered by each study, differences in measurement of student performance, differences in specification, and aggregation bias in the statistical analysis.

The workers in Card and Krueger's sample attended schools between the 1920s and the 1970s, a span of time encompassing variations in the level of resources going far beyond what is found today. This suggests one reconciliation of the conflicting study findings: if added resources have diminishing effects on student achievement, current school operations may be largely "on the flat" of the production function, while Card and Krueger observe ranges from the past where resources had stronger effects. A related possibility might be that the political economy of schools has changed over time. For example, with the rise of teachers unions and the resulting change in bargaining positions, resources might be used in different ways and have different student achievement implications now than in the past (see, for example, Borland and Howsen [1992], Peltzman [1993], and Hoxby [1996]). In other words, it is quite possible that the enormous changes in educational resources did have an effect on outcomes in the first half of this century, but that more recent studies are also correct in

finding "no effect" for the sorts of resource changes discussed in current schools.

A series of more technical discussions has also been introduced to "choose" between the competing views of the effects of school resources. The debate has focused on questions about the appropriateness of measuring student performance with achievement tests and on questions about specific aspects of the Card and Krueger statistical methodology. These debates go far beyond this specific paper, but they provide, in the opinion of one of the principals in the debate, a strong foundation for accepting the basic conclusion that added resources have not and are unlikely to improve student outcomes noticeably.[13]

META-ANALYSIS AND THE SUMMARY OF RESULTS

In some research areas, such as the investigation of the health effects of a certain drug therapy, there is frequently an interest in compiling results from a variety of trials. Specialized techniques to combine the results of separate studies and thus assess the magnitude and significance of some relationship have been developed. These approaches go under the general title of "meta-analysis." The previous summary of results represents one simple approach to the aggregation of results, but other researchers have attempted to do formal statistical tests.

A well-known application of formal statistical tests to education production function data is found in Greenwald, Hedges, and Laine (1996). The question they pose is whether there is any evidence that resources or expenditure differences *ever* appear to affect student performance. Their formal tests lead to rejection of this restricted null hypothesis. The most basic problem with their statistical analysis is that it addresses an uninteresting question from a policy viewpoint. Their results are sometimes interpreted as refuting the conclusion that educational inputs do not affect performance. But in my view, this work both confirms the previous substantive results and points to the same policy conclusions. As all of the analysis shows, productive results are possible, even if seldom achieved currently. This conclusion is central to much of the policy discussion.[14]

THE SCHOOLS-ARE-DOING-FINE SCHOOL

Surprisingly, in the face of evidence such as that presented here, some commentators have argued that U.S. schools look pretty good. Krueger (1998) has fallen in line with Berliner and Biddle (1995) and other writers who suggest that concerns about the performance of our education system are quite overblown and that in reality there is evidence both of high performance and of marked improvement. Because these arguments have received wide circulation—largely, it appears, from people who wish to maintain and expand the current structure—it is useful to understand how these arguments are constructed.

Krueger presents evidence about performance on the NAEP exams that he generally interprets as supporting the effectiveness of current schools. For most of his discussion, he combines scores on the reading, mathematics, and science exams for a random sample of students aged nine, thirteen, and seventeen between 1969 and 1996. He places great weight on the occurrence of a statistically significant time trend in scores and a statistically significant correlation between scores and spending per pupil in a majority of the nine tests.

First, note that these tests are not independent of each other. The cohort of students tested at age nine in any year is tested again four years later at age thirteen and again four years later at age seventeen. The same holds true for the cohort of students tested at age thirteen and again four years later at age seventeen. Thus, only two of the nine cohorts of nine-year-old students and two of the nine cohorts of thirteen-year-old students (those tested in 1994 and 1996) are not retested and included in the NAEP results for later ages. For policy purposes, we are clearly most interested in performance at age seventeen, just before students go into the job market or into postsecondary education. The combined analysis by Krueger provides equal weight to test measures at any point during the educational process, even if these measures are "superseded" by measures closer to the finish of the process and closer to the time when they have real importance.

The importance of combining the tests is immediately apparent from looking at Charts 2 and 3 in this paper and by comparing overall NAEP results on the per-

formance of students around the time that they leave high school with their performance in earlier grades. For science, the average scale score of seventeen-year-olds falls 9 points (0.2 standard deviation) between 1969 and 1996. For math, seventeen-year-olds improve 3 points (0.11 standard deviation) between 1973 and 1996. For reading (not shown in the charts), the seventeen-year-olds improve 2 points (0.05 standard deviation) between 1971 and 1996. (Writing performance, which is only available since 1984, shows a fall of 7 points, or 0.2 standard deviation, by 1996.) Only the fall in science (and in writing since 1984) is a statistically significant difference. By contrast, there are statistically significant increases in five out of six tests of the earlier age groups for science, math, and reading. In other words, our third graders are becoming better at conversation and better at making change, even if they ultimately cannot fill out a job application or participate in modern quality control functions at the workplace. It is these latter findings for early ages that permit Krueger to discuss the "surprising" finding that there is a statistically significant trend in student performance.

Imagine a scoring system for the summer Olympics where the times of runners in the 100-meter dash are recorded at 50 meters, 75 meters, and 100 meters, and where these intermediate and final times are simply averaged to make judgments about who is the fastest sprinter. This is the system Krueger suggests for evaluating the performance of U.S. elementary and secondary schools. It is possible that an Olympic track coach would want to use the information about performance at intermediate distances to assess which aspects of the race each runner should work on. Similarly, one might want to assess whether the U.S. schools are improving or lagging at different grade levels. It is doubtful, however, that the highly aggregated evidence of the NAEP scores would be good evidence for making specific resource allocations. Krueger's use is solely to support the case that U.S. schools are performing quite well, at least compared with his prior perceptions.

Second, Krueger goes on to use the same basic methodology to assess the effectiveness of school resources. He regresses spending per student in each test year on the existing nine NAEP scores. The entire exercise is peculiar in the sense that one would not normally expect the performance of seventeen-year-olds to be dictated by just the spending in the current school year. Krueger is relying on the fact that spending has moved up so steadily over the past forty years that the current spending in any year is a good index of differences in the cumulative spending over the school years for different cohorts.

It is no surprise that the results of this statistical exercise for spending are very similar to the previous investigation of trends. Chart 4 and Table 3 show that spending has moved quite steadily upward across the period of the NAEP scores, indicating that either current or cumulative spending will look very similar to a time trend. From analyzing the cumulative average spending for each of the tested cohorts (that is, averaging spending across the prior years of schooling for each age and test year group), the statistical results show the same basic pattern as the simple trends, even if there are minor changes in which estimates are labeled statistically significant. In this version of the trend exercise, neither the negative relationship between the science performance of seventeen-year-olds and spending nor the positive relationship for math performance is statistically significant at the 10 percent level. The positive correlation for reading is statistically significant at the 6 percent level. Again, there is a difference for younger cohorts in each of the tests. For age nine and age thirteen, four out of the six correlations are statistically significant at the 5 percent level, and five out of six are statistically significant at the 10 percent level. The overall results suggest, nonetheless, that good performance in earlier years has not translated effectively into higher cumulative performance at the time of graduation, when achievement counts most.

Third, the translation of the analysis into the relationship between scores and spending allows Krueger to perform a policy analysis that provides some feel for the magnitude of the results. Statistical significance helps to decide whether or not we should believe there is any relationship at all, but it does not indicate how much we might expect from an increase in spending. In order to be concrete, Krueger considers a simple increase in spending of $2,000 per student. We should first be clear about the meaning of this. With 50 million students in

1995 (45 million in public schools), this specification amounts to an increase in annual spending of $100 billion. In other words, he is proposing increasing total 1995 spending on elementary and secondary students by more than one-third. If financed entirely through appropriations by the U.S. Department of Education, this illustrative calculation would call for a quintupling of federal aid to schools.

On the basis of the education system's past performance (captured by the simple regressions of NAEP scores for seventeen-year-olds on cumulative prior spending), student achievement would be predicted to rise 0.06 standard deviation in reading and 0.10 standard deviation in math; it would be predicted to fall by 0.08 standard deviation in science. It would seem hard to get much political or public support for this pattern of results, even at substantially lower cost. These projections are obviously not based on the scaling of the charts assessing student performance. Nor does translating these projections into movements across a normal distribution of the population increase the sense that this is what we would call a successful $100 billion annual expenditure.

Betts (1998), using a different approach, calculates the rate of return to additional school district spending. His findings confirm my conclusion: on the basis of past performance and the current structure of schools, additional spending on schools appears to have a net negative rate of return.

In sum, a variety of researchers and commentators have put a spin on the performance data for U.S. schools that suggests that our schools have been doing well and have been improving as a result of past spending increases. These assertions are not supported by the data. Allowing these arguments to distract us from developing more effective policy options would be a mistake. Contrary to the assertions of Berliner and Biddle (1995), the problems of America's schools represent neither myth nor fraud but instead a series of more fundamental organizational problems.

Policy Implications

The interpretation of the overall results about school resources depends fundamentally on how the policy and decision-making process is conceived. At one level, these conclusions clearly imply that educational policymaking is more difficult than many would like. If resources had a consistent and predictable effect on student performance, policymaking would be straightforward. State legislatures could decide how much money to invest in schools and could trust local districts to apply funds in a productive manner. But the fact that local districts do not use funds effectively complicates this picture. The clearest message of existing research is that uniform resource policies will not work as intended.

Similar policy dilemmas face the courts in school finance cases. The courts have entered into education decision making by ruling on suits brought by people who believe that state legislatures are not fulfilling their constitutional obligation to provide equitable or adequate education to particular students in each state. While frequently motivated by concerns about student achievement, in reality both the judicial statement of the issue and the proposed remedies center fundamentally on the level and distribution of resources. If resource availability is not a good index of educational outcomes or if providing for overall resource levels does not ensure a desired level of performance, the courts face the same dilemma as legislatures. Simply providing more funding or a different distribution of funding is unlikely to improve student achievement (even though it may affect the burdens of school financing on the citizens of a state).

A variation of this general theme is to argue that, while resources alone may not be sufficient to guarantee achievement, adequate resources are surely necessary. Undoubtedly, this statement is accurate at some level, because a school with no funds would not be expected to add anything to student achievement. Nonetheless, as shown in Table 3, real spending per student increased by more than 70 percent between 1970 and 1991, even though student performance appears to have been essentially unchanged. Further, nothing in the previous analytical results about the effects of resources suggests that there is a level below which resources have clear and powerful effects on achievement that would be a demonstration that some schools are below the threshold of "necessity." Just asserting that there is some level of necessary expenditure does not make the case for pure resource policies in today's schooling environment. While it is not possible to define scientifically how much is "necessary," the dramatically

larger spending of today has clearly taken almost every school system in the country beyond some minimal level.

This policy conundrum is precisely what led the Panel on the Economics of Education Reform to concentrate not on the specific resources and policies of schools but on the incentive structure. Its report, *Making Schools Work*, emphasizes the need to alter current incentives in schools radically (Hanushek et al. 1994). The simple premise is that the unresponsiveness of performance to resources largely reflects the fact that very little rests on student performance. Because good and bad teachers or good and bad administrators can expect about the same career progression, pay, and other outcomes, the choice of programs, organization, and behaviors is less dependent on student outcomes than on other things that directly affect the actors in schools.

Underlying this view is a more benign opinion of school personnel. Specifically, school personnel are not just ignoring a set of policies that would lead to obvious improvements but instead are simply following existing incentives. An added part of this argument is that the kinds of policies that will work in given situations with given personnel and students vary and that these policies are not easily described and centrally regulated. The assumption is that, given better incentives, school personnel can be motivated to search out what will work in their specific situation. Under current incentives, they appear to devote more of their attention and energies elsewhere.

Earlier work on educational production has provided substantial evidence that vast differences exist among teachers and schools. It is just that these differences are not easily explained by the resources employed or by any simple set of programmatic or behavioral descriptions. The existence of effective teachers and schools, however, implies that one approach to policy is to devise ways of rewarding better performance whenever it is found. In other words, even if the details of what will work are unavailable before the fact (or even after the fact), policy can be described in terms of outcomes, and good outcomes can be rewarded.

Such a description is itself much too simple because we have limited experience with alternative incentive schemes (Hanushek et al. 1994). The alternative incentive structures include a variety of conceptual approaches to providing rewards for improved student performance; they range from merit pay for teachers to charter schools to privatization to vouchers. These are contentious proposals, in part because the introduction of performance incentives might lead to a variety of people other than current school personnel making decisions and even providing educational services. Incentive proposals also could work well or poorly, depending on the details. The purpose here, however, is not to consider the pros and cons of alternatives, but to emphasize the radically different perspective on policy that is embedded in each. Performance incentives recognize that varying approaches by teachers and schools might be productive. Thus, they avoid the centralized "command and control" perspective of much current policy. At the same time, they recognize that simply decentralizing decision making is unlikely to work effectively unless there exist clear objectives and direct accountability.[15]

Given the current lack of knowledge about the design or implications of performance incentives, an aggressive program of experimentation and evaluation seems very appropriate (Hanushek et al. 1994). Nonetheless, the lack of direct information should not be taken as support for more of what we are doing now. We actually have considerable experience with the current organization, and current approaches appear to offer little hope for general improvement.

The existing work does not suggest that resources never matter. Nor does it suggest that resources could not matter. It only indicates that the current organization and incentives of schools do little to ensure that any added resources will be used effectively. Faced with this problem, some simply declare that we should still pursue general resource policies but that we should not pursue programs that do not work. This approach would be fine if policymakers could reliably identify programs that do and do not work. We know that their judgments have not been accurate in the past.

ENDNOTES

1. The consideration of externalities is one area in which the current U.S. situation differs from that in many other countries. At the high levels of attainment in the United States, the case for strong influences on literacy, the functioning of democracy, the health of the population, or the reduction in crime—traditional items cited as externalities—seem weak. At lower levels of attainment, a better case might be made. See Hanushek (1996b) and Poterba (1996).

2. While I am perhaps biased toward the work I have done with Kim (1996), the effects of qualitative differences in student performance on national growth rates appear huge. This evidence fits nicely with concerns that have been raised in the United States about the achievement of students—concerns that are developed in the next section.

3. The performance of U.S. eighth graders was relatively better in science than in mathematics. This finding is not completely consistent with the U.S. performance over time in the separate mathematics and science tests of the NAEP.

4. The federal government is not the main actor in either the elementary and secondary or the higher education arena. Of the $260 billion in revenues for elementary and secondary schools in 1993-94, only 7 percent came from federal funds, while 45 percent came from state funds and the remainder was supplied by local school districts. Similarly, of the $179 billion in revenues for higher education, slightly more than 12 percent came from the federal government (U.S. Department of Education 1996). More important, resources to schools at all levels have shown considerable growth for a long time, even though most of these funds have come from states and localities.

5. Details of the data and analysis of costs can be found in Hanushek and Rivkin (1997). All monetary measures are adjusted by the GNP deflator to constant 1990 dollars. Moreover, most discussion concentrates on current expenditure, that is, total expenditure less capital investment. Unless otherwise noted, we also use public school expenditure, excluding that going to private schools. The proportion of students attending private schools has ranged from 10 to 15 percent for the entire century, with a vast majority of private school students attending Catholic schools.

6. Spending as a percent of GNP actually peaks around 1975 at 3.9 percent. As noted below, this percentage is affected noticeably by the demographics of the school-age population, making it a poor statistic for comparing the intensity of resources devoted to schooling.

7. An alternative explanation is that schools attempted to protect employment during periods of falling enrollment. While this may have been the initial motivation, subsequent data indicate that the reductions have been permanent ones, resistant to any increase brought about by rising enrollment. The efficacy of improving school quality by reducing pupil-teacher ratios has at the same time been seriously questioned.

8. This is not to say that nonpecuniary factors are unimportant in determining whether individuals choose to teach. Rather, we assume that nonpecuniary benefits or costs of teaching have not changed in comparison with those in other occupations, in which case changes in relative earnings function as a good index of where teachers fall in the labor force. The influence of changes in wage rates on the stock of teachers is complicated, and it depends on the choices of school districts and the behavior of prospective teachers. See Ballou and Podgursky (1995) for a consideration of various outcomes from wage changes.

9. A more complete analysis of spending changes, along with definitions of the precise data used, can be found in Hanushek and Rivkin (1997).

10. The use of annual earnings, which include money teachers receive for teaching and other occupations, obviously goes beyond comparing pure teaching salaries with salaries in other occupations. I believe that, while more common, using just teaching salaries produces the wrong comparison because teachers enjoy much longer vacations than most other workers. Overall earnings better reflect the monetary benefits of being a teacher as opposed to having a different primary occupation. Broad occupations clearly differ in a variety of nonmonetary ways, including fringe benefits and average length of workday and work year. This analysis assumes that the relative importance of these factors for teaching and other occupations has remained constant over time. Rothstein and Miles (1995) suggest that between 1967 and 1991, benefits for teachers rose faster than those for the rest of the economy, although, as they point out, such comparisons are difficult to make with a high degree of reliability. Private school and public school teachers are grouped together in my analyses; nevertheless, since a roughly constant 10 percent of students attended private school throughout the period, it is unlikely that movement in the earnings of private school teachers would have a significant impact on the overall relative wages of teachers.

11. The clearest evidence comes from a series of covariance, or fixed-effects, estimates of performance differences across teachers (for example, see Hanushek [1971, 1992], Murnane [1975], Murnane and Phillips [1981], and Armor et al. [1976]). These analyses consistently show that the differences between a "good" and a "bad" teacher in the poverty-stricken schools of Gary, Indiana, was approximately one grade level per academic year; that is, a student with a good teacher might advance 1.5 grade equivalents in a school year, while one with a bad teacher might advance only 0.5 grade equivalent (Hanushek 1992). Moreover, the consistency of individual teacher effects across grades and school years indicates that the estimated differences relate directly to teacher quality and not to the specific mix of students and the interaction of teacher and students.

12. The Card and Krueger (1992) study is the most discussed analysis of school resources and earnings, but it follows a larger line of research. For an insightful review of past studies that considers underlying characteristics of the studies, see Betts (1996).

13. Betts (1996) and Hanushek (1997) provide evidence on the generalizability of any conclusions about the effects of resources on measures other than achievement tests. Heckman, Layne-Farrar, and Todd (1996) suggest that the estimates of Card and Krueger are not robust to the database or to the model specification. They also introduce concerns about the validity of assumptions needed to identify the key parameters of the Card and Krueger model. Hanushek, Rivkin, and Taylor (1996) provide evidence that the aggregated analysis relying on state differences could bias the results toward the finding of resource effects. See, however, the discussion in Card and Krueger (1996).

14. Other factors have entered into the discussion of meta-analysis, but they are not central to the policy discussion here. See Hanushek (1996a, 1996c, 1997).

15. While the decentralization considered here refers to pure resource policies and general funding, the evidence supports this conclusion even at the level of school-based management. See Summers and Johnson (1996).

REFERENCES

Armor, David, Patricia Conry-Oseguera, Millicent Cox, Niceima King, Lorraine McDonnell, Anthony Pascal, Edward Pauly, and Gail Zellman. 1976. ANALYSIS OF THE SCHOOL PREFERRED READING PROGRAM IN SELECTED LOS ANGELES MINORITY SCHOOLS. Santa Monica, Calif.: RAND.

Ballou, Dale, and Michael Podgursky. 1995. "Recruiting Smarter Teachers." JOURNAL OF HUMAN RESOURCES 30, no. 2 (spring): 326-38.

Baumol, William J. 1967. "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis." AMERICAN ECONOMIC REVIEW 57, no. 3 (June): 415-26.

Berliner, David C., and Bruce J. Biddle. 1995. THE MANUFACTURED CRISIS: MYTHS, FRAUD, AND THE ATTACK ON AMERICA'S PUBLIC SCHOOLS. Reading, Mass.: Addison-Wesley Publishing Company.

Betts, Julian R. 1998. "The Two-Legged Stool: The Neglected Role of Educational Standards in Improving America's Public Schools." Federal Reserve Bank of New York ECONOMIC POLICY REVIEW 4, no. 1.

———. 1996. "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature." In Gary Burtless, ed., DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS, 141-91. Washington, D.C.: Brookings Institution.

Borland, Melvin V., and Roy M. Howsen. 1992. "Student Academic Achievement and the Degree of Market Concentration in Education." ECONOMICS OF EDUCATION REVIEW 11, no. 1 (March): 31-9.

Card, David, and Alan B. Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." JOURNAL OF POLITICAL ECONOMY 100, no. 1 (February): 1-40.

———. 1996. "Labor Market Effects of School Quality: Theory and Evidence." In Gary Burtless, ed., DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS, 97-140. Washington, D.C.: Brookings Institution.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. EQUALITY OF EDUCATIONAL OPPORTUNITY. Washington, D.C.: U.S. Government Printing Office.

Greenwald, Rob, Larry V. Hedges, and Richard D. Laine. 1996. "The Effect of School Resources on Student Achievement." REVIEW OF EDUCATIONAL RESEARCH 66, no. 3 (fall): 361-96.

Hanushek, Eric A. 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." AMERICAN ECONOMIC REVIEW 60, no. 2 (May): 280-8.

———. 1992. "The Trade-Off between Child Quantity and Quality." JOURNAL OF POLITICAL ECONOMY 100, no. 1 (February): 84-117.

———. 1996a. "A More Complete Picture of School Resource Policies." REVIEW OF EDUCATIONAL RESEARCH 66, no. 3 (fall): 397-409.

———. 1996b. "Rationalizing School Spending: Efficiency, Externalities, and Equity, and Their Connection to Rising Costs." In Victor R. Fuchs, ed., INDIVIDUAL AND SOCIAL RESPONSIBILITY: CHILD CARE, EDUCATION, MEDICAL CARE, AND LONG-TERM CARE IN AMERICA, 59-91. Chicago: University of Chicago Press/NBER.

———. 1996c. "School Resources and Student Performance." In Gary Burtless, ed., DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS, 43-73. Washington, D.C.: Brookings Institution.

———. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." EDUCATIONAL EVALUATION AND POLICY ANALYSIS 19, no. 2 (summer): 141-64.

Hanushek, Eric A., and Dongwook Kim. 1996. "Schooling, Labor Force Quality, and the Growth of Nations." Unpublished paper, University of Rochester Department of Economics, December.

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.

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): 611-27.

Hanushek, Eric A., et al. 1994. MAKING SCHOOLS WORK: IMPROVING PERFORMANCE AND CONTROLLING COSTS. Washington, D.C.: Brookings Institution.

Heckman, James S., Anne Layne-Farrar, and Petra Todd. 1996. "Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship." In Gary Burtless, ed., DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS, 192-289. Washington, D.C.: Brookings Institution.

REFERENCES (*Continued*)

Hoxby, Caroline Minter. 1996. "How Teachers' Unions Affect Education Production." QUARTERLY JOURNAL OF ECONOMICS 111, no. 3 (August): 671-718.

Krueger, Alan B. 1998. "Reassessing the View That American Schools Are Broken." Federal Reserve Bank of New York ECONOMIC POLICY REVIEW 4, no. 1.

Murnane, Richard J. 1975. IMPACT OF SCHOOL RESOURCES ON THE LEARNING OF INNER CITY CHILDREN. Cambridge, Mass.: Ballinger.

Murnane, Richard J., and Barbara Phillips. 1981. "What Do Effective Teachers of Inner-City Children Have in Common?" SOCIAL SCIENCE RESEARCH 10, no. 1 (March): 83-100.

Peltzman, Sam. 1993. "The Political Economy of the Decline of American Public Education." JOURNAL OF LAW AND ECONOMICS 36 (April): 331-70.

Poterba, James M. 1996. "Government Intervention in the Markets for Education and Health Care: How and Why?" In Victor R. Fuchs, ed., INDIVIDUAL AND SOCIAL RESPONSIBILITY: CHILD CARE, EDUCATION, MEDICAL CARE, AND LONG-TERM CARE IN AMERICA, 277-304. Chicago: University of Chicago Press/NBER.

Rothstein, Richard, and Karen Hawley Miles. 1995. WHERE'S THE MONEY GONE? CHANGES IN THE LEVEL AND COMPOSITION OF EDUCATION SPENDING. Washington, D.C.: Economic Policy Institute.

Summers, Anita A., and Amy W. Johnson. 1996. "The Effects of School-Based Management Plans." In Eric A. Hanushek and Dale W. Jorgenson, eds., IMPROVING AMERICA'S SCHOOLS: THE ROLE OF INCENTIVES, 75-96. Washington, D.C.: National Academy Press.

U.S. Department of Education. 1996a. DIGEST OF EDUCATION STATISTICS 1996. Washington, D.C.: National Center for Education Statistics.

———. 1996b. EDUCATION INDICATORS: AN INTERNATIONAL PERSPECTIVE. Washington, D.C.: National Center for Education Statistics.