

# Reassessing the View That American Schools Are Broken

---

*Alan B. Krueger*

A growing number of scholars and political commentators have concluded that the U.S. public school system is flawed, and that it can only be corrected by fundamental changes in the institutions that govern education. Chubb and Moe (1990, p. 3), for example, argue that the “existing [educational] institutions cannot solve the problem, because they are the problem.” Widespread belief that the current educational system is flawed, rather than any concrete or systematic evidence indicating that an alternative system performs better than the current one, has motivated frequent calls for radical “institutional reforms” of schools.

The view that the U.S. school system has failed, or is “broken,” is commonly supported by three arguments: (1) there has been a steady decline in the performance of American students on standardized tests, (2) American

children perform worse on international comparisons than foreign children, and (3) the existing system fails to convert school resources (such as smaller classes) into school outputs (such as better test performance).<sup>1</sup>

This paper reassesses the claim that American schools are broken. The first section examines trends in National Assessment of Educational Progress (NAEP) test scores, and the relationship between average test score performance and school resources on an aggregate level. Although the aggregate data show a surprisingly strong, *positive* relationship between educational spending and student achievement, one should be cautious about drawing any causal inference from such a relationship because of changes in the composition of students over time and changes in the focus of educational spending. More convincing evidence comes from the randomized experiment on class size, which I describe in the subsequent section. Next, I infer the influence of schooling on student performance by considering gains in student achievement by socioeconomic status (SES) during the school year and during the summer months. The paper’s final

---

*Alan B. Krueger is the Bendheim Professor of Economics and Public Affairs at Princeton University and a research associate of the National Bureau of Economic Research.*

section summarizes evidence on the increasing economic rewards associated with completing high school.

The main conclusion from this review is that the widely held belief that American schools have failed—that they are performing worse today than they have in the past, that a high school degree is no longer valuable, and that additional resources yield no benefits in the current system—is not supported by the evidence. The evidence suggests that the perceived crisis in education has been greatly exaggerated, if indeed there is a crisis at all. Nonetheless, major changes in U.S. schooling might produce more desirable results. However, it would not be prudent to radically restructure the U.S. education system out of misplaced frustration that the current system has failed miserably or out of an unsupported presumption that progress cannot be made in the context of the existing system. In light of these findings, the conclusion offers incremental proposals to improve our schools.

## WHAT DO THE AGGREGATE ACHIEVEMENT TEST DATA TELL US?

### AGGREGATE TIME TRENDS

Concern over the deteriorating performance of U.S. students is often based on time-series trends in the Scholastic Aptitude Test (SAT). For example, Chubb and Moe (1990, pp. 7-8) write, “the single most important symbol of the underlying problem came to be the monotonic decline, from the mid-1960s through 1980, in the scores of high school students on the national Scholastic Aptitude Test, or SAT.” The emphasis on the average SAT score is odd because the exam is not designed to measure students’ current levels of acquired skills, but instead their potential to perform well in college. Even more important, the students who take the SAT are a self-selected lot, and the selection has changed dramatically over time. As a wider segment of American students has attended college, the percentage of high school seniors taking the exam has increased. This increase has been particularly strong among students who rank in the bottom half of their high school class (see Berliner and Biddle [1995]). Because the composition of students taking the SAT has changed over time, the College Entrance Examination Board, which publishes the test, has repeatedly warned against

inferring trends in school or student performance from the SAT (see, for example, College Entrance Examination Board [1988]).

To the extent that one can correct for the changing mix of students who take the SAT, there is little cause for alarm. For example, Berliner and Biddle (p. 22) show that between 1976 and 1993, the average SAT score has gone up for every demographic group except whites, and it declined only slightly for whites. The authors (p. 32) also summarize evidence that shows an upward trend in the 1980s in the California Achievement Test (CAT), the Stanford Achievement Test, the Comprehensive Test of Basic Skills, and other commercial tests. There is little support in these data for the claim made by the National Commission on Excellence in Education (1984, p. 8) that “average achievement of high school students on most standardized tests is now lower than 26 years ago when Sputnik was launched.”

Most analysts probably agree that the NAEP exam provides a more meaningful assessment of trends in student performance than the SAT. Like the SAT, the NAEP is conducted by the Educational Testing Service. But unlike the SAT, the exam is administered to a representative sample of students and is intended to assess progress on basic math, reading, and science skills. The NAEP exam has been administered to nine-, thirteen-, and seventeen-year-olds in selected years since 1970. There are a total of nine time trends that can be analyzed with the NAEP data. Chart 1 presents the average NAEP exam scores for each year, after age and subject main effects have been removed.<sup>2</sup> For most of the subjects and age groups, the NAEP data display a modest upward time trend after an initial dip in the early 1970s. Indeed, the correlation between the average NAEP score and time (that is, the year in which the test was given) is positive for eight of the nine age-by-subject cases, and it is statistically significant at the 10 percent level for seven of the nine cases. The median of these nine linear trends indicates that test scores are rising by .06 standard deviation per decade.<sup>3</sup> It is also possible that the unadjusted NAEP data understate the upward trend in student performance because the composition of students has changed over time. In particular, the rising proportion of students who are immigrants and minorities,

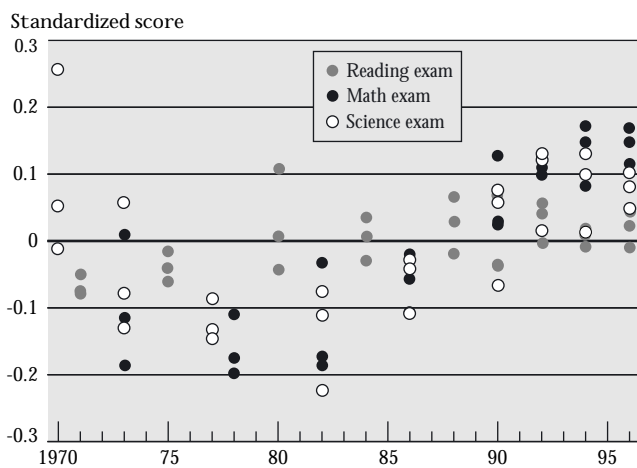
and raised in poverty and by single parents, might be expected to lower average test scores over time.<sup>4</sup>

Chart 2 displays trends in average NAEP mathematics test scores for seventeen-year-old black students and for all students who live in disadvantaged urban communities.<sup>5</sup> The scores are expressed as deviations from the 1973 overall NAEP score, divided by the 1996 cross-sectional standard deviation. Perhaps surprisingly, the chart shows that the most disadvantaged students have made the greatest gains. The gap in math scores between students in disadvantaged communities and all communities narrowed by approximately one-half of one standard deviation in the 1980s. Moreover, between the early 1970s and 1990, the black-white NAEP mathematics test-score gap for seventeen-year-olds decreased by nearly half, although the gap has expanded in the 1990s. These findings are inconsistent with the popular stereotype that inner-city schools are in decline.

Is the upward trend in the aggregate NAEP scores big or small? To some extent, the significance of the trend is in the eye of the beholder. Hanushek (1996, p. 51), for one, argues that “there is no way to conclude that aggregate performance has increased significantly over the past quarter-century.” The following calculation, however, suggests that

Chart 1

Standardized NAEP Scores over Time

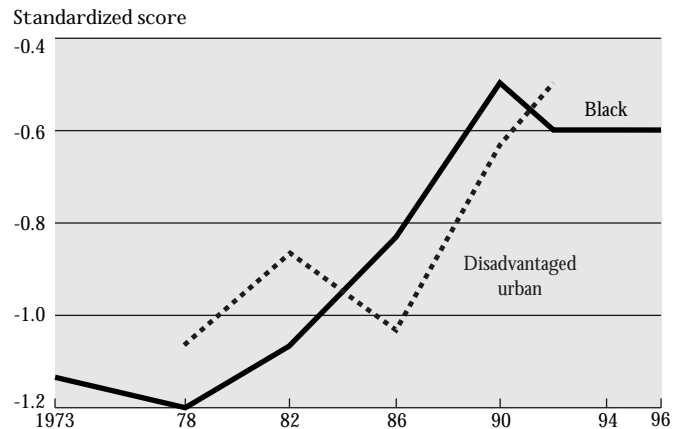


Source: Author's calculations, based on National Center for Education Statistics (1997).

Note: Each point corresponds to an average adjusted for subject and age group; see endnote 2 for more details.

Chart 2

Standardized NAEP Math Scores for Seventeen-Year-Old Black and Disadvantaged Urban Students



Source: National Center for Education Statistics (1994, 1997).

Note: Each score was standardized by subtracting the 1973 score for all students and dividing by the 1996 standard deviation across all students.

the time trend is not trivial. Over a twenty-five-year period, the average NAEP score is predicted to have increased by .15 standard deviation, based on the median of the nine linear trends for all subjects and age groups. What does it mean for the average test score to rise by .15 standard deviation? If the distribution of scores is normal, an increase of .15 standard deviation implies that the average (or median) student would have advanced six percentile ranks. In other words, the student scoring in the fiftieth percentile today would perform as well as the fifty-sixth-percentile student did twenty-five years ago. Although this is not a dramatic improvement, it is difficult to find well-evaluated, large-scale educational innovations that have produced equally large gains for the average student.

### THE RELATIONSHIP BETWEEN AGGREGATE TEST SCORES AND SCHOOL RESOURCES

Hanushek (1996) presents two notable figures. The first shows a near-exponential growth in expenditures per student from 1890 to 1990. The second shows the average NAEP score for seventeen-year-olds on the math, reading, and science tests for available years since 1970. On the basis of these figures, Hanushek (p. 51) concludes, “the aggregate data provide a *prima-facie* case that school spending and school resources are not linked to performance.”

To my surprise, a straightforward statistical analysis of these data is more supportive of the opposite conclusion (Table 1). Specifically, I pooled the NAEP data across the three age groups and three subject tests and estimated an ordinary least squares (OLS) regression of the form:

$$Y = a + b(\text{spending/student}) + \text{subject dummies} + \text{age dummies},$$

where Y is the average score on the NAEP exam measured in standard deviation units, and spending/student is current school spending in constant 1995-96 dollars divided by the number of enrolled students.<sup>6</sup> In some specifications, dummy variables are also included for the age of students (nine and thirteen, with seventeen omitted) and for the subject (math and reading, with science omitted).

The regression results in column 1 exclude the subject and age dummies, while the results in column 2 include these explanatory variables. In either model, an increase in expenditures per student has a positive and statistically significant association with average test scores. A \$2,000 increase in expenditures per student is associated with an increase of about .11 standard deviation in the average NAEP score.<sup>7</sup>

The science exam may be more difficult to compare over time than the math or reading exams because of major breakthroughs in basic science in the last thirty years and the diversity of science curricula across schools. In column 5, the sample is limited to the math and reading exams. Here, we find a larger effect of school spending: a \$2,000 increase is associated with an increase of .14 standard deviation in the mean achievement score.

A great deal of work on “educational production functions” has focused on class size. Therefore, in columns 3, 4, and 6 of Table 1 the pupil-teacher ratio is used as a measure of school resources instead of expenditures per student.<sup>8</sup> These results are also consistent with the view that resources matter. According to the model in column 4, a reduction in the pupil-teacher ratio of eight students—from, say, twenty-three to fifteen—would be associated with an increase in the average score of .176 standard deviation. This is the equivalent of the average student moving up seven percentile ranks, again assuming normality.

To increase the sample size, I pool together all three subject tests and all three age groups in the results reported in Table 1. Perhaps aggregating across age groups and subjects

*Table 1*  
EFFECT OF SCHOOL RESOURCES ON STANDARDIZED NAEP SCORES: POOLED SAMPLE OF MATH, READING, AND SCIENCE SCORES  
ACROSS NINE-, THIRTEEN-, AND SEVENTEEN-YEAR-OLDS  
OLS Coefficient Estimates with Standard Errors in Parentheses

Variable	(1)	(2)	(3)	(4)	Math and Reading Only	
					(5)	(6)
Spending per enrolled student	.054 (.013)	.057 (.011)	—	—	.072 (.012)	—
Pupil-teacher ratio	—	—	-.015 (.005)	-.022 (.006)	—	-.030 (.007)
Math	—	-.080 (.025)	—	-.078 (.027)	-.132 (.021)	-.131 (.023)
Reading	—	.051 (.024)	—	.051 (.026)	—	—
Age nine	—	-.053 (.024)	—	.032 (.035)	-.060 (.026)	.059 (.038)
Age thirteen	—	-.032 (.024)	—	-.032 (.026)	-.043 (.026)	-.042 (.028)
R <sup>2</sup>	.182	.439	.118	.361	.622	.548
Sample size	78	78	78	78	51	51

Sources: National Center for Education Statistics (1997); U.S. Department of Education (1997, Tables 63 and 166).

Notes: Scores have been scaled by subtracting the 1996 score and dividing by the 1996 cross-sectional standard deviation. Spending per enrolled student is in thousands of 1995-96 dollars. Each equation also includes an intercept.

distorts the results. If a separate regression of test scores on expenditures per student is estimated for each of these nine series, however, a positive association is found for eight of the nine cases. Even with a short time series of data, this relationship is statistically significant at the .10 level for seven of the nine cases. The weakest relationships arise for seventeen-year-olds, especially in science (which is negative, with a t-ratio of  $-.77$ ).

I must confess to being surprised by the consistently positive association between NAEP test scores and school resources. To see if my priors were unusual, I e-mailed a short questionnaire to each of the eight other presenters at the “Excellence in Education” conference to assess their expectations of these correlations. Six presenters replied. The median respondent expected six of the nine correlations between NAEP scores and expenditures per student to be positive, with four and a half of the nine correlations statistically significant and positive, and one statistically significant and negative.<sup>9</sup> Thus, the actual correlations are somewhat more supportive of the view that resources are associated with achievement than this small sample of experts anticipated.

How did Hanushek conclude from the aggregate NAEP data that achievement and school resources are not linked? First, he displayed the NAEP data on a scale ranging from 280 to 310. This is a very wide scale, equivalent to 1 standard deviation on the 1996 NAEP math exam. Under normality, if the average student increased his or her performance by 1 standard deviation, the student would move up thirty-six percentile ranks in the distribution. With such a wide scale, any change in the NAEP score appears visually attenuated. Second, Hanushek only displayed trends for seventeen-year-olds; these students exhibit a weaker relationship between test scores and resources than the other age groups. If the model in column 2 of Table 1 is estimated just for nine- and thirteen-year-olds, for example, the coefficient on spending per student is 33 percent larger.

An analysis of the NAEP scores more thorough than mine—conducted by Grissmer et al. (1997) and based on regional-level data over time—reaches the same qualitative conclusion as that suggested by Table 1. But I

do not wish to extol the findings based on the aggregate NAEP data very much, if at all. Obviously, many relevant factors have changed over time that may bias (either upward or downward) the relationships estimated in Table 1. In addition, to the extent that the generosity of resources is partially determined by low test performance (as in compensatory education), simultaneity bias will attenuate the relationships found in the table. Suffice it to say that my interpretation of the aggregate data is that they provide *prima facie* evidence that student achievement may be linked to school resources. In my view, a far more compelling test of whether resources matter in the current system is discussed in the next section.

#### THE TENNESSEE STUDENT-TEACHER ACHIEVEMENT RATIO EXPERIMENT

There has been considerable debate over whether devoting more resources to schools in the current system would improve student outcomes.<sup>10</sup> Research has been unable to resolve this debate, in part because it is unclear which variables (family background, innate ability, and so forth) should be held constant when the effect of school resources on student performance is estimated. Additionally, when education production functions are estimated with observational data, there is concern about reverse causality: more resources may be assigned to some schools or classes because of low achievement. Finally, there is no consensus as to the appropriate specification of the education production function. For example, some researchers have related the change in test scores to the level of resources in any given year, some have related the change in test scores to the change in resources, and others have related the level of test scores to the level of resources.

An experiment in which children are randomly assigned to classes with high and low levels of resources would help to overcome many of these statistical problems. Because children are already assigned to teachers and schools, controlled experimentation is more feasible in education than in many other fields. Yet the education field lags behind medicine, job training, and agricultural research in the extent to which controlled experiments are utilized. The Food and Drug Administration requires convincing

evidence from a well-designed experiment before new drugs that influence life and death can be put on the market; but when it comes to new educational innovations, weaker standards of evidence are required.

The Tennessee Student-Teacher Achievement Ratio (STAR) experiment is an exception in the education field. In this experiment, 11,600 Tennessee students in eighty participating schools were randomly assigned to varying sized classes in kindergarten and grades 1 through 3. Mosteller (1995) describes Project STAR as “a controlled experiment which is one of the most important educational investigations ever carried out and illustrates the kind and magnitude of research needed in the field of education to strengthen schools.” Although the experiment was not perfect (what study is?), the results strongly suggest that smaller class sizes help students, especially low-income and minority students.

The key features of the experiment include the following.<sup>11</sup> The experiment began in 1986 and included the wave of students who were enrolled in kindergarten that year. Within each participating school, kindergarten students were randomly assigned to a small class (an average of 15.1 students), a regular-size class (an average of 22.4 students), and a regular-size class with a teacher’s aide (an average of 22.8 students). The original plan called for the students to remain in their original class-size assignment until the third grade. After the third grade, the experiment was concluded and all students were assigned to regular-size classes. As noted below, with one important exception, the experiment went largely as planned.

Another feature of the experiment is that additional waves of students entered the experiment in the first grade, the second grade, and the third grade. In particular, because kindergarten attendance was not mandatory in Tennessee at the time of the study, many new students entered the program in the first grade. Moreover, students were added to the sample over time because they repeated a grade or because their families moved to a school zone that included a participating school. Some 2,200 new students entered the project in the first grade and were randomly assigned to the three types of classes. More than 1,000 new students entered the experiment in both the second and the

third grades. Newly entering students were randomly assigned to one of the three class types. This feature of the experiment enables the estimation of class-size effects for each wave of students who entered the experiment in various grades.

The students were given a battery of tests at the end of each school year. I focus on the results of the Stanford Achievement Test. Specifically, I measure student performance by the average percentile rank on the math, reading, and word recognition tests.

The Tennessee STAR experiment is the best designed large-scale educational experiment to date. Nonetheless, it had four important limitations:

- Because of parental complaints, students in the regular-size and regular-size/teacher’s aide classes were randomly reassigned between these two types of classes between kindergarten and first grade, while the students in small classes continued in small classes. Note that results from the kindergarten year are uncontaminated by this deviation from the original experimental design. In addition, my analysis (see Krueger [1997]) suggests that the reassignment of students in regular-size classes in first grade did not invalidate the main results of the experiment.
- The experiment did not collect baseline test scores. These data would have been useful to assess whether the students were uniformly distributed across class types by initial achievement level. Nonetheless, the students’ background characteristics (such as age, race, and probability of receiving free or reduced-price lunch) appear to be uniformly distributed across class types, which suggests that random assignment was carried out successfully.
- In grades 1 through 3, each regular class had the services of a part-time teacher’s aide 25 to 33 percent of the time on average, so the variability in aide services between groups was restricted.<sup>12</sup> Because the present focus of my analysis is primarily on the effect of class size, this feature of the experiment is of less concern.
- Attrition from the sample was high, in part because some students repeated grades and were not tracked, and in part because some students moved to other school districts.<sup>13</sup>

## RESULTS OF THE STAR EXPERIMENT

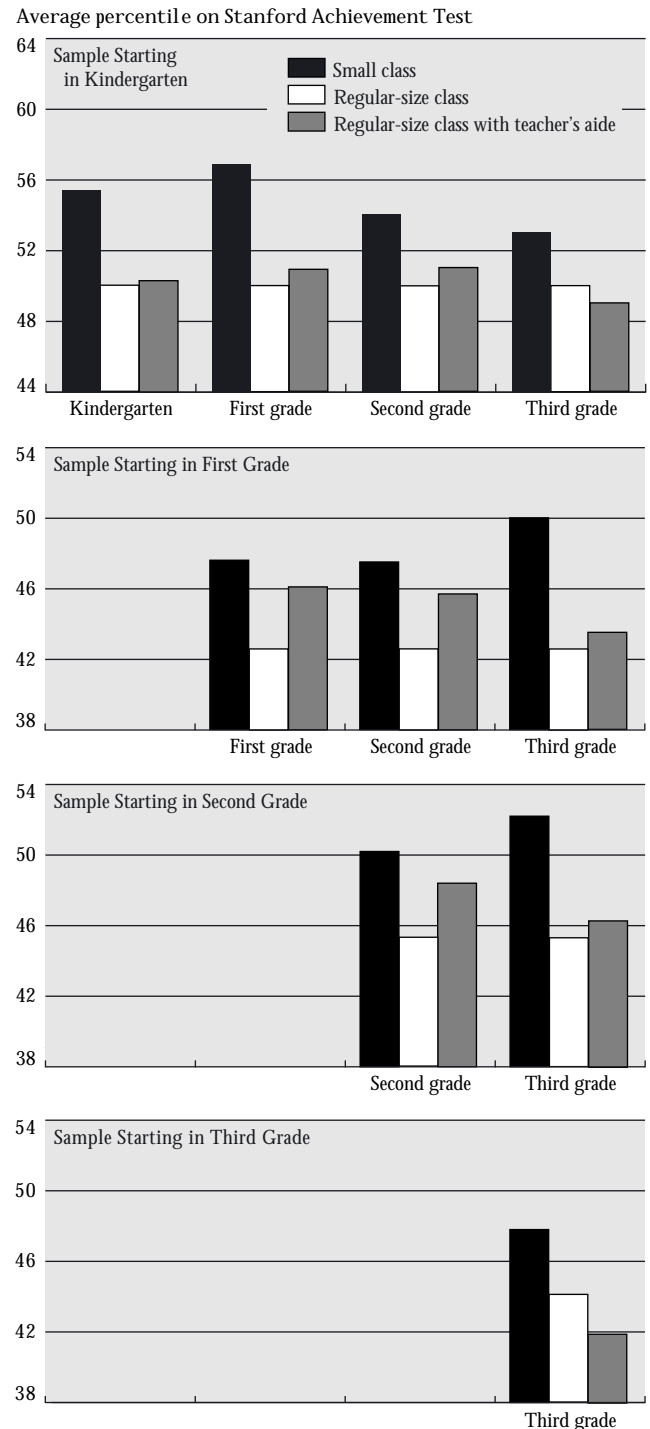
The STAR experiment has been analyzed by Folger and Breda (1989), Word, Johnston, Bain, et al. (1990), Finn and Achilles (1990), and Krueger (1997). The main results of the experiment are summarized below. This summary draws heavily on Krueger, to which the reader is referred for further elaboration of the statistical results.

- The main results for the first four years of the experiment are summarized in Chart 3. For each entry wave of the experiment and grade level, the chart shows the average percentile ranking of students assigned to small classes, regular-size classes, and regular-size classes with a teacher's aide.<sup>14</sup> At the end of the initial year in which students were assigned to small classes, their average performance exceeded that of students in the regular-size and regular-size/teacher's aide classes by about five to eight percentile points. It is an interesting coincidence that this range encompasses the estimated effect of reducing class size by seven to eight students that I obtained from the regression model with the aggregate NAEP data (Table 1).
- For the largest wave, which entered the experiment in kindergarten, the relative advantage of students assigned to small classes tends to grow between kindergarten and first grade, and then is relatively stable. For students who entered in the first or second grade, the advantage of attending a small class tends to grow in subsequent grades.
- In most grades, students assigned to classes with a full-time teacher's aide perform about as well, or only slightly better, than students assigned to regular-size classes without a full-time aide.
- As in all experiments, it is possible that the "treatment" group worked in some way to prove the treatment effective (so-called Hawthorne effects), or that the "control" group worked extra hard to overcome the deficit of being assigned to a small class (so-called John Henry effects).<sup>15</sup> Because there was variability in size even among the classes in the control group, it is possible to explore the likely impact of such "reactive" effects to the experiment. Specifically, I divided the students assigned to regular-size classes into a relatively small class-size group (an average of twenty-one students) and a relatively large class-size group (an average of twenty-five students). I then estimated the difference in average test scores between students in the smaller and larger classes, the results of which appear in Chart 4.<sup>16</sup>

Students in the smaller classes among the controls scored higher on the tests than students in the larger classes. Because the benefit of reducing class size is of roughly comparable magnitude in Chart 4 and Chart 3

Chart 3

### Effect of Class Size on Test Scores: Tennessee Student-Teacher Achievement Ratio Experiment



Source: Krueger (1997, Figure 2).

and because students (and their teachers) in the smaller classes in Chart 4 did not know they were in a treatment group, there is little support for the view that the main experimental results described earlier are contaminated by Hawthorne effects.

- The effect of attending a smaller class tends to vary systematically across certain groups of students. For example, in the STAR experiment minority students and students on free or reduced-price lunch tended to receive a larger benefit from being assigned to a small class. This pattern is consistent with Summers and Wolfe's (1977) finding that attending a small class is more beneficial for low-achieving students than for high-achieving ones.
- The effect of attending a small class also tends to vary across schools. Notice that in the STAR experiment a separate class-size effect can be estimated for each school, because each school had at least one of each class type. I estimated the effect of attending a small class for each school. The standard deviation of these eighty effects (after adjusting for sampling variability) was 6.9 percentile points. At the average school, the assignment of students to a small class raised performance by 4.6 percentile points. For two-thirds of the schools, the effect of attending a small class was positive, while for one-third it was negative. Furthermore, 30 percent of the schools

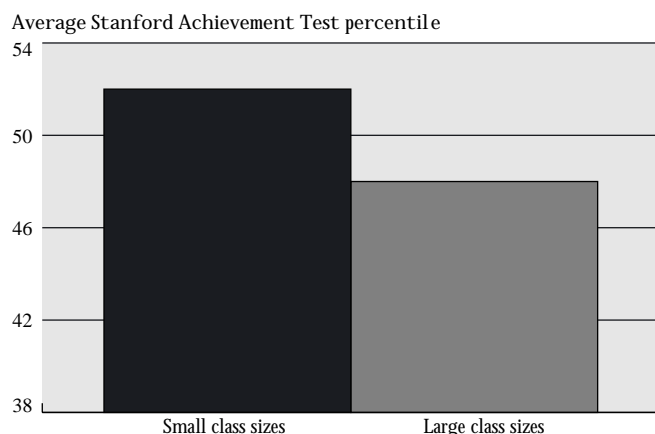
had t-ratios on the small-class effect exceeding 2, while 2.5 percent had t-ratios of less than -2. Smaller classes seem to help student performance at the average school and, indeed, at most schools, although there appears to be a wide distribution of the effect of class size on performance across schools.

If researchers and administrators could determine which schools manage to translate resources more effectively into student performance than others, we could target resources to those types of schools, and try to emulate their practices elsewhere. Consequently, I related the school-level class-size effects to variables such as the racial composition of the students, the urbanicity, and the percentage of students receiving free lunch. Although some of these variables were related to the effect sizes in bivariate regressions, they were all individually insignificant when they were included in a multiple regression.

- The students who participated in Project STAR were returned to regular classes after the third grade and have been tracked since then. Nye, Zaharias, Fulton, et al. (1994) find that students who were placed in small classes had lasting achievement gains through at least the seventh grade, although the later benefits are difficult to compare in magnitude with those at earlier grades because of changes in the tests that were administered. Since the STAR students are currently finishing high school, it would be desirable to learn more about their long-term academic—and just as important, nonacademic—outcomes as they enter early adulthood.

Chart 4

#### Hawthorne or John Henry Effects: A Comparison of Test Scores within Regular Classes



Source: Author's calculations, based on Tennessee Student-Teacher Achievement Ratio data.

Notes: All grade levels have been pooled together. Small classes have an average of twenty-one students; large classes have an average of twenty-five students.

#### SUMMER AND SCHOOL-YEAR TEST SCORE GAINS

Another way to infer the impact of schooling on educational achievement is to compare student progress during the school year and during the summer months. Entwisle, Alexander, and Olson (1997) provide a particularly careful application of this approach. Specifically, they collected data on 790 first-time first-grade students from a stratified sample of twenty Baltimore public schools in 1982. These students were tracked for several years. They were given the California Achievement Test at the beginning and end of each school year.<sup>17</sup> Consequently, test score gains could be tracked during the school year and during the summer months when schools were not in session.



Entwisle, Alexander, and Olson divide students into three groups based on their families' socioeconomic status. This categorization involved weighting five standardized variables: mother's and father's occupations and education, and subsidized meal status. The top panel of Chart 5 summarizes their findings concerning seasonal gains on the math CAT for children from the lowest and highest SES groups. The test-score gains during the school year are remarkably similar for the two groups. The bottom panel of Chart 5, however, indicates that during the summer months, children from low-SES families lose ground, while children from high-SES families gain ground. Table 2 summarizes the cumulative gains on the reading and math CATs over the first five years during which the children were tracked. These results suggest that children from high- and low-SES households make roughly the same progress during the school year, although the gap between high- and low-SES children expands during the summer months because

Table 2  
CUMULATIVE FIVE-YEAR GAINS ON CAT EXAMS,  
BY FAMILY SOCIOECONOMIC STATUS

Socioeconomic Status	Over the School Year	Over the Summer
	Math	
Low	185.8	-8.0
High	186.3	24.9
Reading		
Low	193.3	0.8
High	190.9	46.6

Source: Entwisle, Alexander, and Olson (1997, Table 3.1).

children from low-SES families fall behind when school is not in session.<sup>18</sup> One interpretation of these results is that the public schools overcome whatever learning deficits are associated with low socioeconomic status when schools are in session, but during the summer months the family environmental effects dominate.

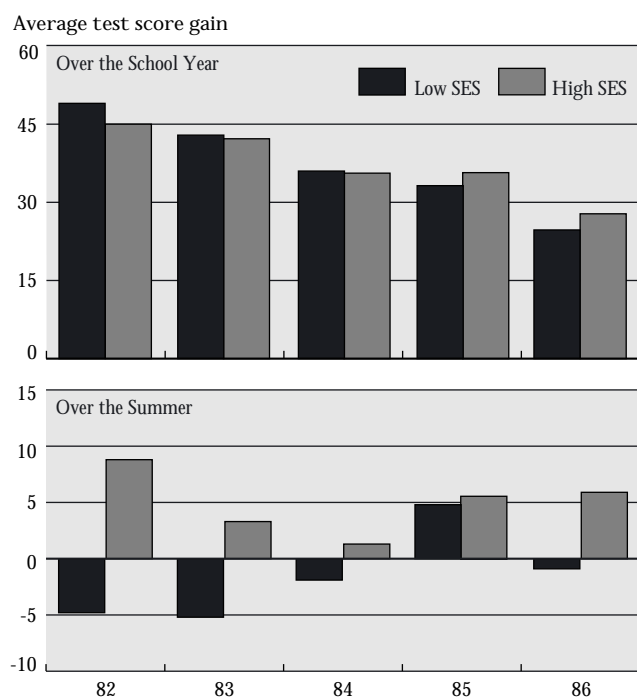
In 1967, Martin Luther King, Jr., wrote, "the job of the school is to teach so well that family background is no longer an issue." At least insofar as test-score gains are concerned, the schools may have achieved this goal. During the school year, students make roughly the same gains regardless of family background. The challenge is to overcome the relative decline in performance of low-income children that occurs when school is not in session, and before school begins.

### FINISHING HIGH SCHOOL PAYS OFF

Looking across workers, researchers have found that average earnings tend to be higher for those with more years of schooling.<sup>19</sup> Moreover, a large literature documents that the "monetary return to education"—that is, the earnings of highly educated workers relative to those of less educated workers—rose substantially in the 1980s (see Levy and Murnane [1992] for a survey). Even when the same cohorts are tracked through time, the return to education appears to have increased, suggesting that changes in the skill composition of workers across education levels do not explain this phenomenon. The most compelling interpretation of the rise in the return to education is that the labor market now values skills more highly than it used to, probably because of the combined effects of skill-biased technological changes, globalization, and institutional changes.

Chart 5

#### Seasonal Gains on CAT Math Exam, by Family Socioeconomic Status

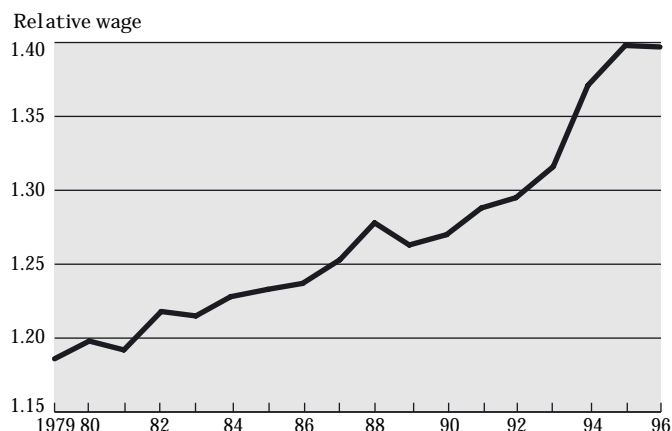


Source: Entwisle, Alexander, and Olson (1997, Table 3.1).

Notes: Sample consists of 498 Baltimore public school students who entered first grade in 1982. SES = socioeconomic status.

Chart 6

### Ratio of Median Weekly Wage of High School Graduates to that of High School Dropouts



Source: Author's calculations, based on unpublished Bureau of Labor Statistics tables.

Note: The chart presents the ratio of median weekly wages of workers with exactly a high school education to those with less than a high school education for full-time workers aged twenty-five and older.

Although much of the literature on the return to education focuses on the college–high school wage ratio, Chart 6 documents that it is also true that the ratio of earnings of those with exactly a high school degree to those with less than a high school degree increased over time. Between 1979 and 1996, for example, the earnings advantage of high school graduates relative to that of high school dropouts roughly doubled, from 19 percent to 40 percent. Although the level of earnings of high school graduates has declined over the last two decades once we adjust for inflation, the relative earnings of high school graduates has increased substantially. This finding suggests that—compared with the alternative—completing high school is more valuable today than it used to be, which is inconsistent with the view that the secondary education system fails to produce something of value.<sup>20</sup>

### CONCLUSION

The evidence reviewed in this paper suggests that the U.S. public school system has not deteriorated and may in fact be reasonably efficacious. It is nonetheless possible that school resources could be deployed more effectively and educational innovations could improve student outcomes. Moreover, the seemingly ever-increasing demand for skilled workers in the

United States underscores the potential economic benefits of enhancing the skills of future generations of Americans by improving public schooling. But the conclusion that the current educational system is not inherently flawed leads to policy proposals of a more incremental nature. My personal view is that policymakers should be risk-averse when it comes to changing the public school system. To alter the institutional structure of U.S. schools radically without sufficient evidence that the “reforms” would be successful is to put our children at risk. With this in mind, I would propose that the following incremental reforms be given serious consideration.

- Careful experimentation and evaluation should proceed on a limited basis before wide-scale institutional changes are introduced. To the extent possible, randomized assignment to treatment and control groups should be the research design used to evaluate educational initiatives, such as vouchers, magnet schools, and charter schools. Long-term follow-up focusing on concrete outcomes, not just test scores, would also be desirable. The research base on which educational policy is formulated should be greatly expanded. More experimentation and evaluation are needed. A risk-averse strategy would require fairly convincing evidence that any new educational initiative is effective before it is implemented on a wide scale.
- Demographic projections indicate that public school enrollment will rise in the coming decade. This rise, in turn, will lead to pressure to increase class size or school spending. The research suggests to me that an increase in class size, especially in the early grades, would lower the average student's performance. Society may choose to increase class sizes rather than incur additional educational expenses, but this decision should be made with the expectation that student achievement will be affected by the level of school resources. Decision makers and the public should not bank on there being a “free lunch” when it comes to raising class size.
- In most U.S. public schools, the school year runs about 180 days. This is shorter than in many other industrialized countries. One of the strongest findings in the economics of education literature is that, on average, students who attend school longer earn higher incomes in the labor market. It is reasonable to hypothesize that more days of schooling would have a similar effect. In addition, Entwisle, Alexander, and

Olson's (1997) work suggests that disadvantaged children fall behind in the summer months when schools are closed. I would propose experimenting with an expanded school year of, say, 210 days. To facilitate evaluation of such a change, it would be desirable to phase in an extension of the school year across randomly selected communities within a state.

- The school day in the United States is a legacy of a bygone era when most women did not work for pay.

Times have changed. Probably because of lax supervision in the afternoon when most parents are working, the juvenile violent crime rate peaks between 2 p.m. and 4 p.m. on school days.<sup>21</sup> On nonschool days, the crime rate peaks in the evening. This finding suggests possible gains from extending the school day, or from providing a richer set of after-school activities. Again, experimentation and evaluation would be desirable.

---

## ENDNOTES

*The author thanks Michele McLaughlin for very helpful research assistance and Lisa Barrow, Michele McLaughlin, and Cecilia Rouse for helpful comments.*

1. For example, these arguments are expressed in National Commission on Excellence in Education (1984) and Chubb and Moe (1990).
2. The NAEP scores in the chart were first standardized by subtracting the 1996 test score (for that age group and subject) from each observation, then dividing this quantity by the 1996 cross-sectional standard deviation. The transformed scores were then regressed on two age dummies and two subject dummies. The residuals from this regression are displayed. This regression adjustment was used to remove average age and subject effects.
3. The linear trends were estimated from a bivariate regression of the NAEP exam on the year in which the test was taken. The NAEP score is scaled relative to the cross-sectional standard deviation on the exam in 1996. See also National Center for Education Statistics (1997) for a discussion of trends in NAEP scores.
4. However, these factors may be partially or fully offset by the rise in average educational attainment of parents and the decline in the number of children per family.
5. Disadvantaged urban students are defined as those who live in metropolitan statistical areas and attend schools where a high proportion of the students' parents are on welfare or are not regularly employed. Unfortunately, NAEP data are available for this group only from 1978 to 1992.
6. More precisely, if the average test score in year  $t$  is  $Y_t$ , the average score in 1996 for that subject and age group is  $Y_{96}$ , and the standard deviation of scores across students in 1996 for that subject and age group is  $s_{96}$ , then the dependent variable is  $(Y_t - Y_{96})/s_{96}$ . Thus, each year's test score is measured in standard deviation units from the 1996 average test score for that age group and subject.
7. Hanushek (1996) focuses on expenditures per pupil in average daily attendance instead of expenditures per enrolled pupil. Because enrollment is arguably less endogenously determined than attendance, I use expenditures per enrolled student. If the attendance measure is used, however, the coefficient in Table 1, column 2, row 1 is similar: .053 (standard error of .010).
8. A number of arguments could be made that the pupil-teacher ratio is a superior measure for these specifications. Expenditures will be heavily influenced by teacher pay, which varies over time in part because of external labor market forces that influence the pay of college-trained women. Perhaps more important, in these results the pupil-teacher ratio is more closely tailored to the grade level of the students, whereas the expenditures per student pool all grade levels. But I would not want to identify any particular estimate in Table 1 as "the" correct estimate.
9. Another question on my survey was, "Suppose that public school spending in the U.S. were to permanently increase by \$2,000 per student in 1997 dollars. Give your best point estimate of the probability that student achievement would increase, on average." The median respondent thought there was a 75 percent probability that average achievement would rise in this hypothetical situation. Some of the respondents indicated that they expected any increase to be small, however.
10. Even quantitative summaries of the literature reach contrasting findings (see, for example, Hanushek [1986] and Hedges, Laine, and Greenwald [1994]).
11. Project STAR was funded by the Tennessee legislature, at a total cost of approximately \$12 million. The research was designed and carried out by research teams from Tennessee State University, Memphis State University, the University of Tennessee, and Vanderbilt University. See Word, Johnston, Bain, et al. (1990) and Folger (1989) for detailed descriptions of the experiment.
12. Regular classes often had a teacher's aide because the ethic underlying the study was that students in the control group (the regular classes) would not be prevented from receiving resources that they would ordinarily receive.
13. Straightforward adjustments for attrition (for instance, assigning the last test score to students who leave the sample for each subsequent year) did not indicate that attrition distorted the results (Krueger 1997).
14. The results are based on separate OLS regression models estimated for each entry wave and grade level. The dependent variable in the regression model is the student's percentile ranking on the Stanford Achievement Test, and the independent variables are class-type assignment dummies, school dummy variables, race, sex, free-lunch status, teacher sex, teacher race, and teacher education (see Krueger [1997, Table 6, column 4]).
15. In the current case, one may be more concerned about Hawthorne effects because the treatment appears to have been effective, although John Henry effects may lead to an underestimate of the benefit of attending a small class. For an interesting study that casts doubt on the presence of Hawthorne effects in the original Hawthorne experiments, see Jones (1992).

---

## ENDNOTES (*Continued*)

16. To increase the sample size, I pooled students across all grade levels and entry waves. The results shown in Chart 4 are from a regression of the students' percentile rank on a dummy variable for class size, student characteristics, and teacher characteristics.

17. Students were tested regardless of whether they were held back a grade. In principle, the CAT is scaled so that scores are comparable across grade levels.

18. Murnane (1975) similarly finds that test scores for inner-city children increase during the school year and either stagnate or decline during the summer months. Grossman and Sipe (1992) find that achievement levels remain constant for disadvantaged youth who are randomly selected to participate in summer school, while test scores decline for the control group over the summer. They find no long-term benefits of the summer program, however.

19. A long-standing concern has been that the observed earnings-education gradient might reflect unobserved factors such as inherent ability or family background. Most of the literature that tries to estimate the payoff to years of education, however, concludes that omitted variables do not seriously bias the OLS estimates (see, for example, Griliches [1977] and Angrist and Krueger [1991]).

20. For studies of how school resources relate to the return to education, see Card and Krueger (1992) and Heckman, Layne-Farrar, and Todd (1996).

21. See, for example, U.S. Department of Justice (1996), which is based on FBI data on violent crime in South Carolina.

---

## REFERENCES

- Angrist, Joshua, and Alan Krueger.* 1991. "Does Compulsory Schooling Affect Schooling and Earnings?" *QUARTERLY JOURNAL OF ECONOMICS* 106 (November): 979-1014.
- 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.
- 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 (February): 1-40.
- Chubb, John E., and Terry M. Moe.* 1990. *POLITICS, MARKETS AND AMERICA'S SCHOOLS.* Washington, D.C.: Brookings Institution.
- College Entrance Examination Board.* 1988. *GUIDELINES ON THE USES OF COLLEGE BOARD TEST SCORES AND RELATED DATA.* New York.
- Entwisle, Doris, Karl Alexander, and Linda Olson.* 1997. *CHILDREN, SCHOOLS AND INEQUALITY.* Boulder, Colo.: Westview Press.
- Finn, Jeremy D., and Charles M. Achilles.* 1990. "Answers and Questions About Class Size: A Statewide Experiment." *AMERICAN EDUCATIONAL RESEARCH JOURNAL* 27 (fall): 557-77.
- Folger, John.* 1989. "Editor's Introduction: Project STAR and Class Size Policy." *PEABODY JOURNAL OF EDUCATION* 67, no. 1 (fall): 1-16.
- Folger, John, and Carolyn Breda.* 1989. "Evidence From Project STAR About Class Size and Student Achievement." *PEABODY JOURNAL OF EDUCATION* 67, no. 1 (fall): 17-33.
- Griliches, Zvi.* 1977. "Estimating the Returns to Schooling—Some Econometric Problems." *ECONOMETRICA* 45 (January): 1-22.
- Grissmer, David, Stephanie Williamson, Sheila Kirby, and Mark Berends.* 1997. "Exploring the Rapid Rise in Black Achievement Scores in the United States (1970-1990)." Unpublished paper, RAND Institute on Education and Training, Washington, D.C.
- Grossman, Jean Baldwin, and Cynthia L. Sipe.* 1992. "Summer Training and Education Program (STEP): Report on Long-term Impacts." Philadelphia: Public Private Ventures, winter.
- Hanushek, Eric A.* 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *JOURNAL OF ECONOMIC LITERATURE* 24 (September): 1141-77.
- . 1996. "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.
- Hekman, James J., Anne Layne-Farrar, and Petra Todd.* 1996. "Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship." In Gary Burtless, ed., *THE LINK BETWEEN SCHOOLS, STUDENT ACHIEVEMENT, AND ADULT SUCCESS.* Washington, D.C.: Brookings Institution.
- Hedges, Larry V., Richard Laine, and Rob Greenwald.* 1994. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *EDUCATION RESEARCHER* 23, no. 3 (April): 5-14.
- Jones, Stephen.* 1992. "Was There a Hawthorne Effect?" *AMERICAN JOURNAL OF SOCIOLOGY* 98, no. 3 (November): 451-68.
- Krueger, Alan B.* 1997. "Experimental Estimates of Educational Production Functions." NBER Working Paper no. 6051.
- Levy, Frank, and Richard Murnane.* 1992. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations." *JOURNAL OF ECONOMIC LITERATURE* 30, no. 3 (September): 1333-81.
- Mosteller, Frederick.* 1995. "The Tennessee Study of Class Size in the Early School Grades." *THE FUTURE OF CHILDREN: CRITICAL ISSUES FOR CHILDREN AND YOUTHS* 5 (summer/fall): 113-27.
- Murnane, Richard.* 1975. *THE IMPACT OF SCHOOL RESOURCES ON THE LEARNING OF INNER CITY CHILDREN.* Cambridge, Mass.: Ballinger Publishing Company.
- National Center for Education Statistics.* 1994. *NAEP 1992 TRENDS IN ACADEMIC PROGRESS.* Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, July.
- . 1997. *NAEP 1996 TRENDS IN ACADEMIC PROGRESS.* Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, August.
- National Commission on Excellence in Education.* 1984. *A NATION AT RISK: THE IMPERATIVES FOR EDUCATIONAL REFORM.* Washington, D.C.: U.S. Department of Education.

---

## REFERENCES (*Continued*)

- Nye, Barbara, Jayne Zaharias, B.D. Fulton, et al.* 1994. THE LASTING BENEFITS STUDY: A CONTINUING ANALYSIS OF THE EFFECT OF SMALL CLASS SIZE IN KINDERGARTEN THROUGH THIRD GRADE ON STUDENT ACHIEVEMENT TEST SCORES IN SUBSEQUENT GRADE LEVELS. Seventh-grade technical report. Nashville, Tenn.: Center of Excellence for Research in Basic Skills, Tennessee State University.
- Summers, Anita A., and Barbara L. Wolfe.* 1977. "Do Schools Make a Difference?" *AMERICAN ECONOMIC REVIEW* 67, no. 4 (September): 639-52.
- U.S. Department of Education.* 1997. DIGEST OF EDUCATION STATISTICS 1996. Washington, D.C.
- U.S. Department of Justice.* 1996. JUVENILE OFFENDERS AND VICTIMS: 1996 UPDATE ON VIOLENCE. Washington, D.C.: Office of Juvenile Justice and Delinquency Prevention.
- Word, Elizabeth, J. Johnston, Helen Bain, et al.* 1990. "The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report, 1985-1990." Nashville, Tenn.: Tennessee State Department of Education.

*The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.*