2

Little Gain in Student Achievement

Paul E. Peterson

Except for what is learned at school, very little in American life has remained unchanged since *A Nation at Risk*. Homes are bigger, appliances niftier, and gardens more lush. E-mail, Websites, faxing, and cell phones have transformed communication. Recreational equipment has become so sophisticated that sports can reach new extremes.

Nor is it just technology that is improving. Most social indicators have also moved in a positive direction. The average child today is growing up in a more learning-friendly family environment than ever before. For instance, parents are, on average, more educated. In 1970, 52 percent of the population over age twenty-five had a high school diploma or its equivalent; by 1999, 83 percent did. The share of the population holding a college degree increased from 11 percent in 1970 to 25 percent in 1999. Families are smaller, allowing parents to con-

Samuel Abrams provided valuable research assistance for this essay. Thomas Polseno and Antonio Wendland provided staff assistance.

centrate more attention on each child. The percentage of families with three children or more declined from 36 percent in 1970 to 15 percent in 2000. True, immigration rates have risen and more children live in homes that are non-English-speaking (17 percent in 2000, compared with 8 percent in 1980). Likewise, more children do not live with both their parents (now 31 percent, versus 15 percent three decades ago). Yet the poverty rate has remained essentially unchanged, average income has risen steeply, welfare dependency has declined, the murder rate has attenuated, and drug dependence has abated. A much higher percentage of the population of very young children are enrolled in some kind of preschool program, an increase among four-year-olds from 29 percent to 69 percent between 1970 and 2000. On balance, the positive trends seem to outweigh the more problematic ones.

Despite all these favorable changes in society, few demonstrable achievement gains have been realized in elementary and secondary schools.

Upon issuing *A Nation at Risk*, the National Commission on Excellence in Education appeared to mobilize new focus and energy within America's schools. The report motivated calls for reform by governors, presidents, lawmakers, task forces, and national committees. Citizens regularly tell pollsters that education is, for them, a leading national issue. Yet, if one looks at what students learn in school, nothing much has altered for more than a third of a century.

We know this because we have a variety of yardsticks that help us track key trends in American education. The most familiar of these measuring devices is the SAT, the test given to those high schoolers planning to attend college. Also well regarded is the National Assessment of Educational Progress (NAEP), which administers tests in math, reading, science, and other subjects to a cross-section of all students nationwide at ages nine, thirteen, and seventeen. A more traditional measure of progress can be gleaned by looking at high school graduation rates, long used as an indicator of educational im-

provement in the United States. Finally, it is possible to compare U.S. students with their peers abroad. The International Association for the Evaluation of Educational Achievement (IEA) has administered tests in mathematics and science on several occasions since the early sixties, to samples of students in many countries at ages nine, thirteen, and seventeen. More recently, the Organization for Economic Cooperation and Development's (OECD) Program for International Student Assessment (PISA) and the International Adult Literacy Survey (IALS) have provided comparative estimates of literacy skills.

None of these measuring devices supplies, by itself, an indisputably exact measure of what is happening in American education. But when all yield similar results, it is unlikely that the overall findings can be attributed to peculiar defects in specific measuring rods. It is therefore especially significant that, no matter what instrument is used, the results are roughly the same: America's schools are stagnating, showing little improvement since *A Nation at Risk* was written. In fact, by some measures, educational performance has fallen below the standards set by previous generations.

To compare findings from a variety of assessments of American education, it is helpful to introduce a statistical concept, the standard deviation. This concept provides a measure of how much scores (or some other measured variable) are spread around their average. It allows one to place into a single metric results from different tests that contain unlike questions and employ alternative scoring conventions. It is not necessary here to describe all the properties of a standard deviation, but its magnitude needs characterization. Generally speaking, a full standard deviation (generally presented as 1.0 standard deviations) is very large. When it comes to the NAEP test, for example, a standard deviation is equivalent to roughly the four years of learning that take place between the fourth and eighth grades. If a group of fourth-graders score one standard deviation higher than average on the NAEP fourth-grade test, then they are performing as well as the average eighth-grader would have. Conversely, if low-performing

eighth-graders score a standard deviation below average, then they score no better than the average fourth-grader would have.

A standard deviation change of 1.0 in test-score performance is estimated to raise annual earnings by 15 percent to 20 percent later in life.[1] Making an adjustment for all other factors that affect earnings, it is estimated that young adults in their late twenties who earn $30,000 a year could have earned $4,500 to $6,000 more if their test-score performance had been one standard deviation higher. Quite apart from dollars and cents, higher performance is correlated with happier lives, including a reduction in the risk of incarceration, welfare dependency, and the bearing of a child out of wedlock.[2]

There are other ways of characterizing the magnitude of a one standard deviation difference in test performance. Japanese middle schoolers scored about this much higher on the IEA math test than did their peers in America. The large size of this difference has prompted many to search for the secret of the Japanese success. If the Japanese-American comparison seems too abstract, consider a home-grown example, the black-white test-score gap, which in math is also approximately one standard deviation. A variety of explanations have been offered for the size of this gap—the legacy of slavery, poverty, family life, peer-group influence, school resources, low expectations, and minimal school choice for black students. If there is little agreement on which of these factors is most important, there is no doubt that the discrepancy itself is large.

Even a change in test scores of just half a standard deviation (also referred to as 50 percent of a standard deviation) constitutes a substantial change, particularly if the change occurs within a few years. Changes of as little as 10 percent of a standard deviation are usually considered quite small. But even small changes can become big ones if they accumulate over time. If test scores in America had increased by an average of 4 percent of a standard deviation each year over the twenty years since *A Nation at Risk* was written, today's scores, compounded over time, would have been nearly 120 percent of a standard

deviation higher than in 1983. Had such small changes been occurring regularly, from year to year, the U.S. education system would have been recognized as a dynamic element in American society, just as innovative as many other sectors of the American economy.

But what really happened? By using the standard deviation as a statistical measure of changes over time, we are able to address this question by looking at overall trends in American educational outcomes, as measured by a range of surveys of student achievement and high school graduation rates.

SAT Scores

We first pick up the best-known ruler used to assess the state of American education—the SAT. The letters once stood for the Scholastic Aptitude Test, but its makers decided to drop any substantive meaning from the letters once it became apparent that the SAT measured cognitive skill attained at a point in time—not just some underlying propensity to learn. Despite its peculiar moniker, the SAT has several advantages as a measuring device. For one thing, it has been used to measure student performance for decades, during which time psychometricians have refined the test's precision.[3] For another, it is a high-stakes test. Performance on the SAT affects the likelihood that high schoolers will be admitted to the college of their choice and the probability that they will win a scholarship. Because real consequences flow from performance on this test, most students take it seriously, studying sample questions to familiarize themselves with the test's general format and approach. Private entrepreneurs claim they can raise a student's test scores, and in some cases they do, but usually not by more than a modest amount—unless the student makes the special kind of effort usually associated with genuine learning. In short, the SAT reveals how well students can demonstrate their command of a body of material on an occasion when high performance is expected.

It is not just students who take the test seriously. So does the College Entrance Examination Board, which recommends to colleges that they accept this test as a criterion for college admission, and the Educational Testing Service, the private firm that designs and administers the SAT. Questions are carefully prepared and pretested to ensure precise measurement of a student's demonstrated ability. Special efforts are taken to ensure accurate scoring and to guard against cheating. Items are closely held, and the test is administered under the direction of trained proctors.

Offsetting these advantages are certain limitations. The SAT is generally taken only by college-bound high schoolers. Indeed, not even all of these students take the test. Some colleges accept the American College Testing Assessment (ACT) instead.[4] And if a student plans to attend junior or community college, most of the time neither the SAT nor the ACT is demanded. Nor is the SAT a requirement for admission to many four-year colleges. In 2000, only 46 percent of high school seniors took the SAT.

To the extent that the percentage of seniors taking the test varies from year to year, scores can fluctuate for this reason alone, making it difficult to draw straightforward conclusions from changing SAT test scores. Critics of school reform often ascribe the drop in SAT scores to the fact that the share of high school seniors taking the test has been increasing during the post–World War II period, potentially diluting the skills of the test-taking pool. However, studies have found that the decline in SAT scores during the sixties and seventies was only partially caused by a change in the social composition of the test-taking pool. The percentage of high school seniors taking the test hovered at around 33 percent between 1972 and 1984, after which the share climbed to about 45 percent in 1990 and essentially leveled off. Meanwhile, the drop in SAT scores occurred during the seventies, when participation rates were stable. The modest revival in SAT scores occurred during the eighties, when participation rates were actually *rising*.[5] Figure 1 presents the trend in average SAT scores between

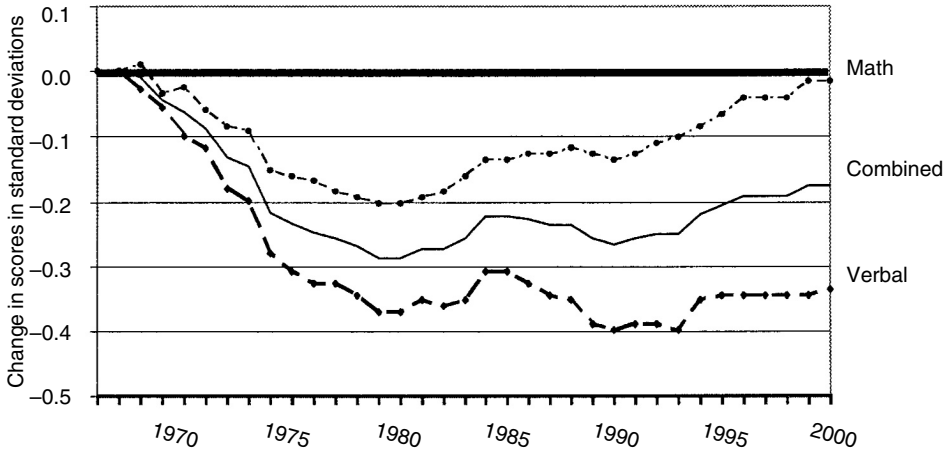

Fig. 1. SAT math, verbal, and combined scores, 1967–2001

Note: Initial score set to zero; subsequent scores constitute changes in standard deviations relative to scores in 1967. The standard deviation used in this table is the average standard deviation from years 1972–2001 (110.6 in verbal and 118.2 in math).

Source: College Entrance Examination Board, National Report on College-Bound Seniors.

1967 and 2001. In the figure, average SAT scores in 1967 are set at zero, allowing one to ascertain how much of a standard deviation the scores have changed in the years since 1967. If the score climbs above zero, students, on average, are doing better than students did in 1967. If the score is below zero, they are doing worse. All comparisons are calculated as percentages of a standard deviation.

The overall picture can be grasped by looking at the trend in combined math and verbal scores displayed in figure 1. These combined scores fell by nearly 30 percent of a standard deviation between 1967 and 1982, a serious sign of deterioration that helped prompt the writing of *A Nation at Risk*. Those who drafted this document can take heart—perhaps even credit—for halting a further decline. But the gains since 1982 are modest, only about 15 percent of a standard

deviation, or less than one percent per year, leaving the country well below its standing in 1967.

When these combined scores are broken down separately into their math and verbal components, noticeable differences may be observed. The math decline was never as steep, just 20 percent of a standard deviation, and by 2001 it had fully recovered. Meanwhile average verbal scores collapsed by 35 percent of a standard deviation, and they have not recovered.

The more positive trend in math scores, compared with verbal scores, may or may not be related to the specific reforms promoted by *A Nation at Risk*. On the one hand, there is little sign that schools adequately addressed the shortage of qualified math and science teachers. As Caroline Hoxby reports in her essay "What Has Changed and What Has Not" (chapter 3), teachers were even less likely to have a degree in these subjects in 1999 than in 1982. On the other hand, more high school graduates were taking academically oriented math and science courses in 1999 than at the time *A Nation at Risk* was penned.

But if more academic courses in math had positive effects on student learning in this subject area, why did we not obtain a similar impact after *A Nation at Risk* from the introduction of more English courses that were given an academic label? The answer to this question remains elusive. Perhaps, as E. D. Hirsch Jr. suggests in his essay "Neglecting the Early Grades" (chapter 9), the replacement of phonics with whole word instruction in elementary school left high schoolers unequipped to read challenging material. Perhaps high school teachers have given up trying to provide rigorous instruction in reading comprehension, letting students focus instead on their own personal responses to the material. Perhaps the syntax and range of expression to which students are exposed in their textbooks has been unduly simplified. Perhaps instruction in math, with its more structured curriculum and clearer set of standards, has been kept relatively intact,

though Paul Clopton and Williamson Evers suggest otherwise in their essay "The Curricular Smorgasbord" (chapter 8).

About such matters, one can only speculate. But this much is known: Even SAT math scores were no better in 2001 than in 1967. And when math and verbal SAT scores are considered together, they reveal a decline of nearly 20 percent of a standard deviation since 1967 and less than a 10 percent uptick since 1982. On the whole, there is little sign of substantial improvement. If the United States was at risk in 1982, it remains so now.

The National Assessment of Educational Progress

The SAT may not be the best tool for assessing the overall state of elementary and secondary schools in the United States. Less than half of all high school seniors took the test in 2001. Nor does the SAT tell us anything about what is happening to children at an earlier age. But if the SAT is a less than perfect measure of school performance, the National Assessment of Educational Progress (NAEP) should be the answer. After all, it is known as the nation's report card, the product of an official, government-administered survey of student learning located within the Department of Education's National Center for Education Statistics. The test is designed by the Educational Testing Service, the same firm that designs the SAT, ensuring, once again, that the skills of leading psychometricians are utilized in its preparation. And the NAEP surveys a representative sample of all students in school, not just those planning to go to college. NAEP math, science, and reading tests have been regularly administered to a representative sample of students at ages nine, thirteen, and seventeen since the early seventies.

Not every student takes the NAEP. Instead, a carefully identified representative sample of schools are selected for testing, and, within each school, designated students are invited to participate in testing

sessions. Unlike the SAT, the test is a low-stakes test. Results are not reported by schools, nor are individual scores released. In fact, no one student answers all the questions on the NAEP, though it uses a sophisticated sampling frame that most psychometricians believe yields reliable results for the population that is sampled. The short test each student takes reduces test fatigue and constitutes less of an intrusion on the time of the school and the student.[6]

Nonetheless, the NAEP, too, has certain limitations. Most important, participation in the NAEP by individual schools has been voluntary. And even when a school agrees to participate, not every student does. Unfortunately, as can be seen in figure 2, participation rates have fallen noticeably since the 1970s, especially among older students.

The joint school-student participation rate for seventeen-year-olds dropped from 68 percent in 1973 to only 58 percent in 1999. For those at age thirteen, the drop over this time period was from 79 percent to 74 percent. If one assumes that low-performing schools are the most likely to refuse to participate and that low-performing students within schools are the most likely to be excused from participation or absent that day, then a decline in participation rates could by itself raise average score results, lending the false appearance that student performance was improving. Although the Educational Testing Service makes statistical adjustments in NAEP scores for nonparticipation by schools, the adjustments only take into account geographical and demographic characteristics of a school, such as whether it is in an urban or nonurban location and the share of a school's students who are minorities. These adjustments cannot correct for variation in school quality unrelated to these characteristics. Similarly, statistical adjustments are made for students, but apart from adjustments for school characteristics, the ethnic background of the student is the only student demographic characteristic for which an adjustment is made. No adjustment is made for any other variation in student performance unrelated to these characteristics. Among seventeen-year-olds, year-to-year changes in participation rates are sig-

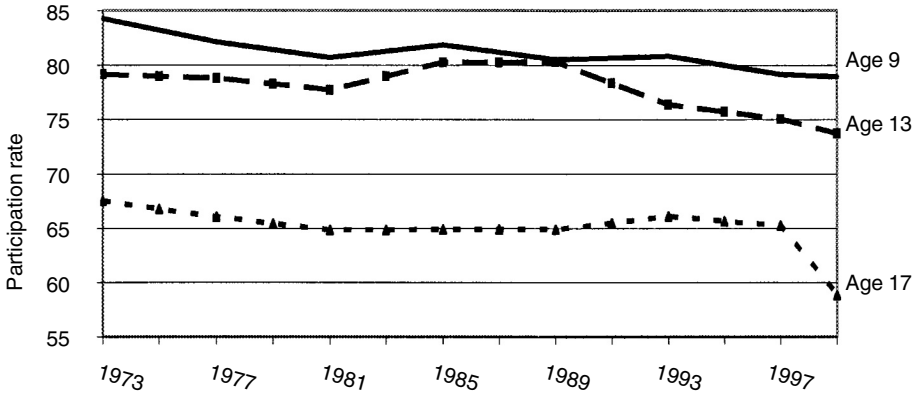

Fig. 2. Combined school-student participation rates in the National Assessment of Educational Progress (NAEP), 1973–1999

Note: Participation rates occasionally differed by science, math, and reading exams. The data points present the average participation rate for all three NAEP exams by age cohort. Participation rates are the probability that a student participates in a particular subject-matter testing-session times the probability that the original school selected participates.

Source: J. R. Campbell, C. M. Hombo, and J. Mazzeo, *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance*. (NCES #2000469, October 13, 2000).

nificantly correlated with year-to-year changes in average test score performance, suggesting that falling rates have inflated recent NAEP test scores.[7]

In other words, NAEP biases run in a direction exactly opposite of SAT biases. If potentially upward trends in American education could be underestimated by the SAT simply because SAT participation rates have risen, then such trends are likely to be exaggerated by the NAEP because its participation rates have been falling.

Bearing this in mind, let's examine the pattern of change in NAEP test scores of those aged seventeen, most of whom were in their last year of high school, about the time many were taking the SAT. As figure 3 shows, among this group of students, NAEP math scores climbed by about 10 percent of a standard deviation between 1973

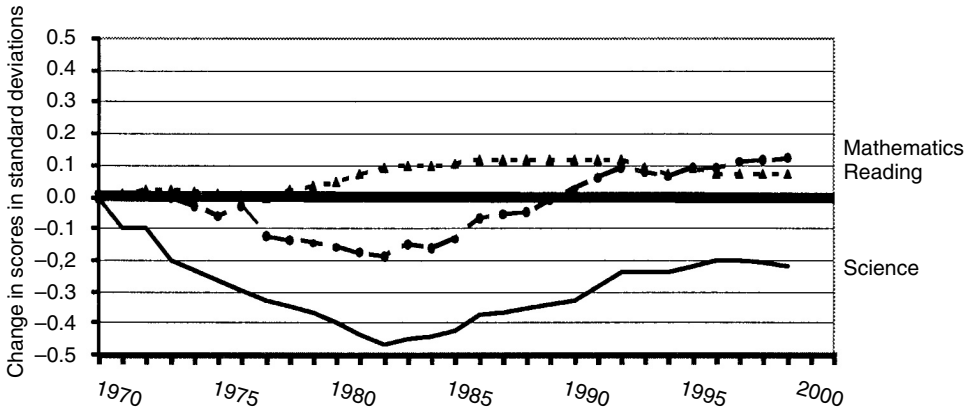

Fig. 3. Math, reading, and science scores on the National Assessment of Educational Progress (NAEP), seventeen-year-olds, 1970–1999

Note: Initial score set to zero; subsequent scores constitute changes in standard deviations relative to initial score. Initial science and math scores obtained in 1970; in reading, initial score obtained in 1973. The standard deviation used here is the average standard deviation from years 1970–1996 (45.4 in science, 31.4 in math, and 42.2 in reading).

Sources: NAEP test scores: J. R. Campbell, C. M. Hombo, and J. Mazzeo, *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance.* (NCES #2000469, October 13, 2000). NAEP standard deviation data: N. L. Allen, J. E. Carlson, and C. A. Zelenak, *The NAEP 1996 Technical Report.* (NCES #1999452, October 19, 2000).

and 1999. Over about the same time period, reading scores rose by about 8 percent. But given the 10 percentage point decline in participation rates, even this slight gain may be more apparent than real.

Science scores dropped by 20 percent, however. The downward shift occurred during the seventies, falling by as much as 50 percent of a standard deviation. Though they recovered subsequently, students in 1999 were still not performing as well in science as they had in the early seventies. And some of the apparent recovery may have been artificial, the simple by-product of falling participation rates. Taken as a whole, NAEP results for seventeen-year-olds, like the SAT scores, reveal a system unable to revive itself.

If one looks at thirteen-year-olds, the picture is only slightly

brighter. Figure 4 shows that reading scores barely improved between 1970 and 1999, a gain so small it may simply be due to the slip in participation rates. Science scores dropped by 20 percent of a standard deviation during the seventies, recovered in the eighties, only to slip again in the nineties. In the end, they were about where they had been in 1970. While reading and science scores were essentially flat, math scores over roughly the same period were as much as 30 percent of a standard deviation higher. But in the three subjects taken together, not much of an upward trend in test score performance can be detected over the time period or as a result of the issuance of *A Nation at Risk* once declining participation rates are taken into account.

The brightest picture appears when one examines the youngest cohort, the nine-year-olds. As displayed in figure 5, these math scores rose by nearly 40 percent of a standard deviation between 1973 and 1999. However, reading and science gains were much smaller, only 5 percent and 10 percent of a standard deviation. Still, the 40 percent uptick in math among this youngest cohort is encouraging. Yet the gains in these early years are not sustained as the child continues into middle and high school.

Some have attributed the minimal sign of educational progress to negative changes in society, rather than to the school.[8] The most important factor that might be depressing student test performance is the rising number of immigrants from Third World countries. If these families have weak educational backgrounds and speak little English, their children may perform poorly on standardized tests. Perhaps the performance of immigrant children is offsetting the gains that American schools are otherwise realizing. To explore this possibility, we examined the trend in average test scores of white students, a group which contains only a few immigrants, most of whom have migrated from advanced industrialized societies whose schools are at least comparable with those in the United States. As can be seen in figure 6, trends in NAEP performance among seventeen-year-old white students track closely the overall trends reported above. The reading

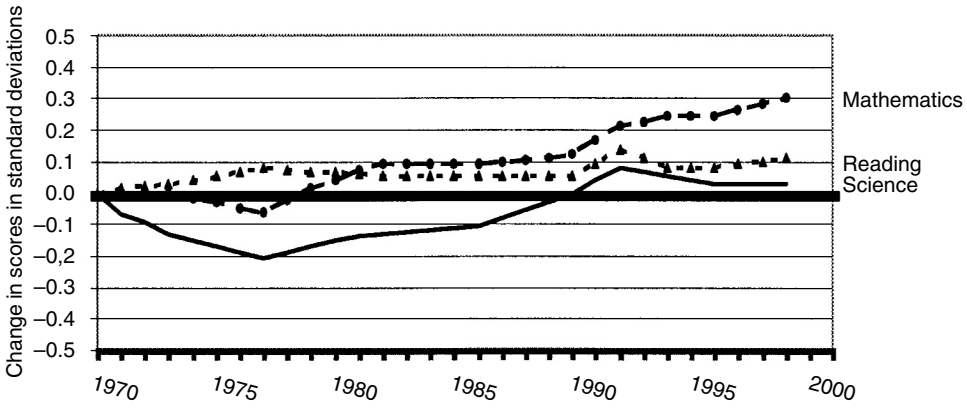

Fig. 4. Math, reading, and science scores on the National Assessment of Educational Progress (NAEP), thirteen-year-olds, 1970–1999

Note: Initial score set to zero; subsequent scores constitute changes in standard deviations relative to initial score. The standard deviation used here is the average standard deviation from years 1970–1996 (38.4 in science, 32.7 in math, and 36.8 in reading).

Sources: NAEP test scores: J. R. Campbell, C. M. Hombo, and J. Mazzeo, *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance*. (NCES #2000469, October 13, 2000). NAEP standard deviation data: N. L. Allen, J. E. Carlson, and C. A. Zelenak, *The NAEP 1996 Technical Report.* (NCES #1999452, October 19, 2000).

scores of seventeen-year-old white students rose slightly (10 percent of a standard deviation) before *A Nation at Risk* was issued but have not increased since. Math scores declined noticeably (20 percent of a standard deviation) before *A Nation at Risk* was issued, but have recovered so that they are now somewhat above their 1970 benchmark. Science scores fell sharply before *A Nation at Risk* and have recovered only about half of their losses. Taking the three subject areas together (and taking into account declining participation in NAEP testing sessions), the test scores of white seventeen-year-olds remained essentially flat over the last thirty years of the twentieth century. One cannot attribute educational stagnation to the influx of immigrants from Third World countries.[9]

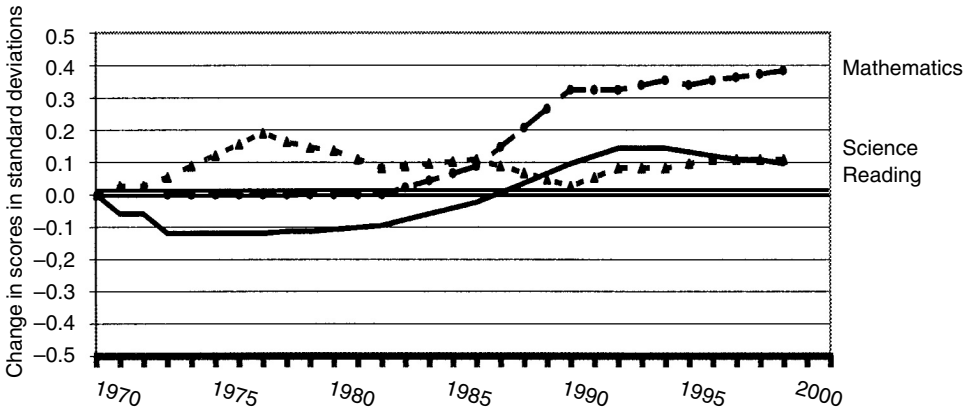

Fig. 5. Math, reading, and science scores on the National Assessment of Educational Progress (NAEP), nine-year-olds, 1970–1999

Note: Initial score set to zero; subsequent scores constitute changes in standard deviations relative to initial score. The standard deviation used here is the average standard deviation from years 1970–1996 (41.5 in science, 33.9 in math, and 36.8 in reading).

Sources: NAEP test scores: J. R. Campbell, C. M. Hombo, and J. Mazzeo, *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance*. (NCES #2000469, October 13, 2000). NAEP standard deviation data: N. L. Allen, J. E. Carlson, and C. A. Zelenak, *The NAEP 1996 Technical Report*. (NCES #1999452, October 19, 2000).

An optimist might interpret these results as showing that reforms recommended by *A Nation at Risk* are slowly taking hold, having their initial impact on younger students who will sustain—and perhaps accelerate—these gains as their schooling continues. Unfortunately, there is little basis for such optimism. The science scores of nine-year-olds shifted upward between 1986 and 1992, but these did not translate into gains when these students reached the age of thirteen. The math scores of nine-year-olds rose quite dramatically between 1986 and 1990, but the improvement for this cohort at age thirteen was much less. When the same cohort became seventeen in 1996, no gain could be detected. What had been achieved by age nine had been lost altogether by age seventeen.

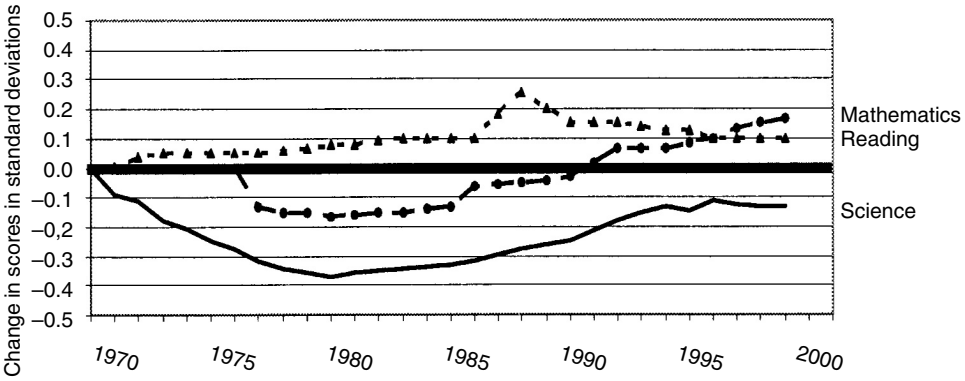

Fig. 6. Math, reading, and science scores on the National Assessment of Educational Progress (NAEP), white seventeen-year-olds, 1970–1999

Note: Initial score set to zero; subsequent scores constitute changes in standard deviations relative to initial score. The standard deviation used here is the average standard deviation from years 1970–1996 (45.4 in science, 31.4 in math, and 42.2 in reading).

Sources: NAEP test scores: J. R. Campbell, C. M. Hombo, and J. Mazzeo, *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance*. (NCES #2000469, October 13, 2000). NAEP standard deviation data: N. L. Allen, J. E. Carlson, and C. A. Zelenak, *The NAEP 1996 Technical Report.* (NCES #1999452, October 19, 2000).

To the pessimist, these results reveal a school system in decline. Although student performance could be depressed by today's greater prevalence of single-parent families, this should be more than offset by the higher levels of parents' educational attainment and smaller family sizes, the two most important family characteristics affecting a child's ability level. Were the deterioration in family life the main cause of educational stagnation, then one would expect to find the worst results among the youngest cohort, the nine-year-olds. But it is this young group that has actually shown the most improvement, perhaps as a result of the higher levels of their parents' educational attainment. As the child ages, however, schools are not adding to what has been learned by the age of nine.

In sum, NAEP scores do not show much more improvement in American education than SAT scores did. Gains apparent when a child is young are not sustained through high school. Those in their last year of schooling score about the same as those a generation and more earlier. If the National Assessment of Educational Progress is supposed to indicate the country's "progress," then the survey's title is quite ironic. It might be better termed the National Assessment of Educational Stagnation.

Graduation Rates

Thus far, we have examined trends in test score performance over time. Perhaps the landscape changes when one looks at real-world outcomes, such as high school graduation rates. Just as higher test scores positively affect a student's well-being later in life, so does the acquisition of a high school diploma. High school graduates are more likely to remain married and avoid incarceration. They are also less likely to bear children out of wedlock or become welfare-dependent. Even if test scores have not improved, perhaps a growing percentage of students are at least remaining in school. After all, the National Education Goals Panel, a group appointed in 1989 to push the *A Nation at Risk* agenda forward, established a 90 percent high school completion rate as a key objective to be reached by the end of the century.

Unfortunately, one cannot even report good news on this front. Not only have high school graduation rates failed to continue to move upward; they have actually declined since *A Nation at Risk* was written.

Surprisingly, the deterioration in the high school graduation rate has not been given the public attention the subject deserves—mainly because the accounting system regularly used to track the graduation rate is quite misleading. Displayed in figure 7 is information from the *Digest of Education Statistics*, the official record keeper on the state of American education. It shows an 86 percent graduation rate in

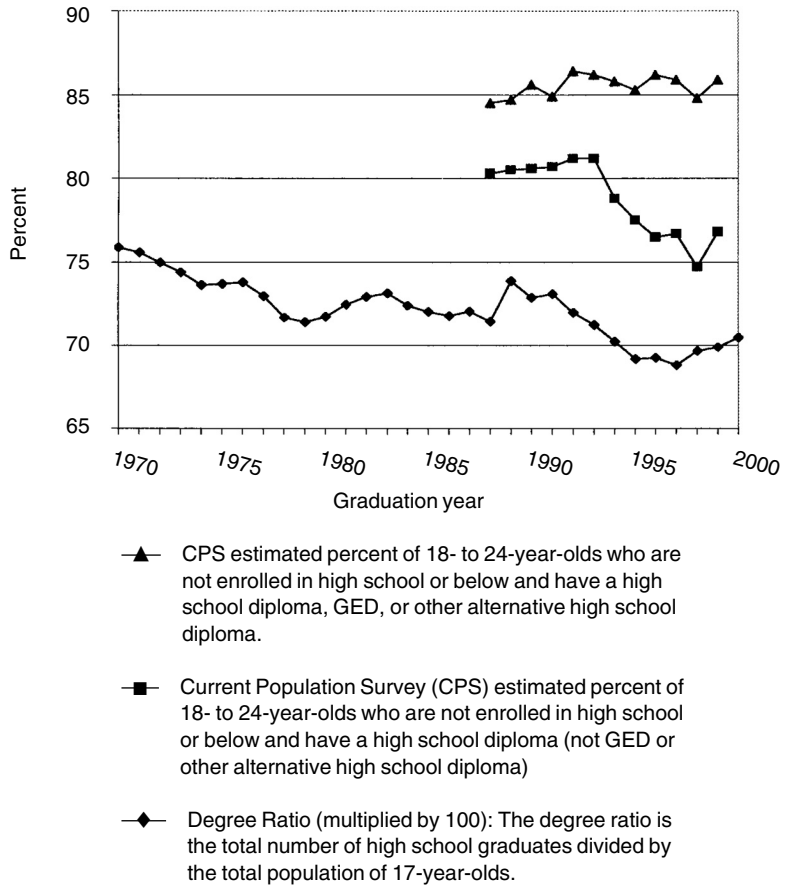

Fig. 7. Estimated high school graduation rates

Note: CPS data indicating whether high school credentials were obtained through a regular diploma or through an alternative route were first collected in 1988.

Sources: *The Digest of Education Statistics*, 2001; CPS data from NCES *Dropout Rates in the United States*, 2001.

1998, the same percentage as in 1985. The 86 percent figure has been accepted as accurate, repeated in numerous official reports, including the one issued by the National Education Goals Panel. If it indicates little progress in graduation rates, at least it shows no decline.

The statistic overstates what is being accomplished in U.S. high

schools, however. The data themselves are collected by the Census Bureau in its annual Current Population Survey of 50,000 households. Each household is asked whether any member in the household between the ages of 18 and 24 has received a high school diploma, is currently in school, or has received the equivalent of a diploma. Approximately 86 percent of young people of this age are said to be high school graduates, so defined.

Relying on this source for information about the quality of the American high school has several problems. For one thing, Census Bureau surveys tend to undercount those households at greatest risk of having a high school dropout. Just as schools can't retain some at-risk students, so census takers cannot find them. It also counts as graduates those still in high school, even at age 18 or older.[10] Even more important, the metric treats the certificate of General Educational Development (GED) as equivalent to a high school diploma, when it is not.

A GED is awarded to students who can pass an examination that apparently demonstrates they have learned as much as is expected of a high school graduate. By law in thirty-five states, the GED is treated as if it were a high school diploma. As the GED has become increasingly promoted as the equivalent of a high school diploma, the number of certificates awarded has risen, climbing from 227,000 in 1971 to 419,000 in 1980 to 501,000 in 2001.

The rising tide of GED bearers is routinely counted in official statistics as part of the high school graduation rate. On the face of it, no test can be thought to be equivalent to a training program. Were it so, then GED recipients would be paid as well as high school graduates. But according to two studies that adjusted for other factors affecting wages, a regular high school diploma is worth 11 to 18 percent more than the GED.[11]

A case for the GED can still be made. For one thing, it provides some students with the opportunity to circumvent an unsuitable high school and still pursue further education. Also, there is some evidence

that the GED, particularly when coupled with work training programs, can enhance earnings.[12] And older immigrants may use the GED as a way to get ahead in their new land. But the growth in the number of GEDs cannot be attributed simply to its utilization by older immigrants. The biggest growth in test-taking has been among nineteen-year-olds, who accounted for 42 percent of all test-takers in 2000 versus just 36 percent in 1990. As figure 8 shows, the numbers of teenagers taking the GED examination grew from around 120,000 in 1989 to around 220,000 in 2000.

Regardless of the merits of the GED, there is little reason to include recipients of a GED when one is assessing the progress made by U.S. high schools. GED programs of study are offered by community colleges, prisons, and private entrepreneurs. They constitute an alternative to, not a component of, the high school. For our purposes, then, GEDs should be excluded from estimates of trends in the graduation rate.

When one puts those holding a GED to one side, the declining state of the American high school becomes manifest. As can be seen in figure 7 (see page 56), Census Bureau estimates that the graduation rate for eighteen- to twenty-four-year-olds fell from 81 percent in 1990 to 77 percent in 1999.[13] Other indicators of high school graduation rates reveal a similar decline. If one calculates the rate by calculating the number graduating from public schools as a proportion of the number of public school students in eighth grade five years earlier, the public school graduation rate declined from more than 78 percent in 1991 to less than 75 percent in 2000.[14] This is a higher-bound estimate of public school graduation rates, however. Many eighth-graders in the private sector transfer into the public sector in their high school years, thereby boosting this estimate. Immigration of adolescents from abroad also boosts this measure of graduation rates. Perhaps a more precise measure is to calculate the number of all high school diplomas handed out each year as a percentage of all seventeen-year-olds. According to this measure the ratio of graduates to those in the appro-

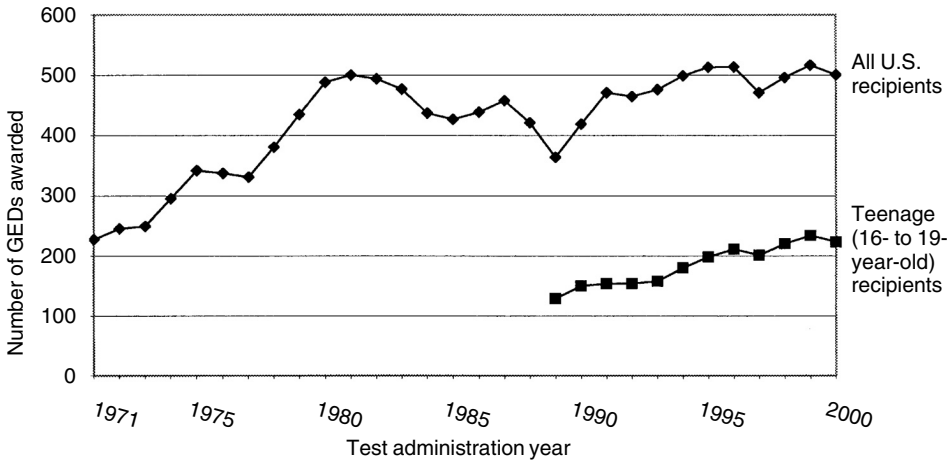

Fig. 8. Number of students receiving the General Educational Development certification, 1971–2001
Source: Digest of Education Statistics, 2001

priate age group fell from 77 percent in 1970 to 74 percent in 1990 and down to 70 percent in 1999—a drop of seven percentage points over the last third of the century (see figure 7, page 56). In other words, in a society where the demand for human capital has been constantly rising, America's schools have responded by turning out a smaller percentage of high school graduates. On the face of it, the nation, far from making progress, remains at risk. Not only are high schoolers not learning as much as their predecessors, a smaller percentage are completing their program of study. Looking at these same data, Paul Barton of the Educational Testing Service asks whether there has been a "closing of the education frontier."[15]

International Comparisons

We have seen few, if any, gains in American education over the past three decades. But how does the country compare with other advanced democracies? Does it remain the world's educational superpower? Is it just one of the pack? Or is it trailing most of the industrial world?

Historically, the U.S. education system has had an outstanding record. It was the first country to achieve universal elementary education, the first to expand its secondary system so as to include the vast majority of students, and the first to establish a broadly encompassing, highly competitive, and world-renowned system of higher education. As late as 1970, a higher percentage of U.S. young people completed their secondary education than did their peers in any other country in the world.

But by the late nineties, the United States no longer led the world in the quantity of secondary schooling received. Instead, the U.S. secondary school completion rate was only about average among the advanced industrial democracies that are members of the OECD, ranking behind Japan, Korea, Germany, France, Ireland, and other European countries (see figures 3 and 4 in the chapter 5 essay by Eric Hanushek, "The Importance of School Quality," for data on high school and college completion rates worldwide). Nor is it just in quantity that the United States lags other advanced industrial democracies. On quality indicators, too, its record is less than sterling.

The best information on school quality has been collected by the IEA, whose tests have been administered to countries throughout the world on several occasions, beginning in the sixties.[16] The IEA tests students only in math and science, forgoing tests of verbal skill, because the designers of the IEA tests thought language differences might invalidate international comparisons of verbal ability. The number of participating countries has changed over the years; during the mid-nineties, thirty-eight countries agreed to have their students tested. Most of the advanced democracies participate in the survey, and several developing countries do so as well.

Hanushek summarizes the results from the IEA surveys in chapter 5 of this volume. As can be seen in figure 1 of that chapter, U.S. average scores in math and science fall below those of the world leaders. In the mid-nineties, for example, the performance of U.S. students ranked closer to the international average (among participating

nations) than to that of students in the highest-performing countries—Singapore, Korea, Japan, Belgium, Hong Kong, the Netherlands, and Austria. Nor are there signs of improvement in U.S. test score performance since the 1970s. Even though some fluctuation over time can be observed, average scores were almost exactly the same in the mid-nineties as they had been in the early seventies. In other words, as with the SAT and the NAEP, IEA tests reveal little sign of educational progress in the United States.

The U.S. standing in the world deteriorates the further along the students are in their educational careers. As can be seen in figure 5 in the Hanushek essay, the ranking of U.S. children at age nine is reasonably satisfactory. Admittedly, U.S. students, on average, scored about 60 percent of a standard deviation below the highest-performing country, Korea, and 20 percent behind the Netherlands, a higher-performing European country. But their overall standing is nonetheless higher than the international average. In other words, just as elementary school children's NAEP scores seemed to be improving, so it appears that their IEA scores reveal an elementary educational system that is functioning adequately, if not brilliantly.

But the standing of the United States slips downward to the international average among students taking the IEA at age thirteen. At this age, U.S. students are 50 percent—no longer just 20 percent—of a standard deviation behind their peers in the Netherlands. And they have fallen more than two-thirds of a standard deviation behind Korean and Japanese thirteen-year-olds, as well as a full standard deviation below world-leading Singapore.

The Unites States' place in the world deteriorates further in the last year of high school. Among seventeen-year-olds, the United States ranks at the bottom of participating countries, save for Lithuania, Cyprus, and South Africa. Although Japan, Korea, and Singapore are not in this survey, students in the Netherlands now outrank those in the United States by nearly a full standard deviation.

In other words, the IEA data, when examined separately by age

group, yield results quite consistent with the NAEP data reported in a previous section: a U.S. education system that is somewhat strong at the elementary level, but weakens as students age and move through school. These results cannot be explained away, as some critics like to do, by claiming that the United States is testing, at age seventeen, a broader array of students than other nations are. IEA data analysts have adjusted for intercountry differences in participation rates. Though critics say the IEA corrections are less than perfect, it's not clear that this point, if valid, would be decisive. Among nations participating in the IEA assessments, the United States no longer has the largest percentage of students in school at age seventeen. As discussed earlier, the graduation rate in the United States now trails those of many other industrialized nations.

The IEA results are taken from tests of math and science ability and knowledge. But according to SAT and NAEP data, U.S. schools are doing even more poorly at inculcating verbal and reading skills. Is there any international evidence on this score? Until recently, no international testing student survey had been conducted in reading. But in 2000, the OECD's PISA conducted its own international survey of what they refer to as reading, math, and science literacy. They surveyed students at the age of fifteen in thirty-one countries, almost all of them advanced industrial democracies. As a supplement to the IEA's surveys, the PISA has many advantages. It provides information at age fifteen (rather than nine, thirteen, and seventeen, the age of students surveyed by the IEA), yielding data at another point in a student's progression through high school. By focusing on age fifteen, the PISA surveys students at an age when hardly any have dropped out of school and before students begin taking highly differentiated courses. Also, the PISA, as a distinct survey by another organization, can reveal whether the IEA's findings are robust, not simply the result of some unknown quirk in the IEA's methodology. Furthermore, by including math, science, and reading literacy in one survey, the PISA

provides for direct comparisons across the three domains of knowledge.

Though PISA results differ from those reported by the IEA in some particulars, overall, the findings from the two international surveys are very similar: As can be seen in figure 9, the average combined scores of U.S. students in all three subjects fall at about the international average of all participating countries, more than 40 percent of a standard deviation behind Japan, Korea, and Finland, the highest-performing countries.

Figures 10 and 11 break out test score results by subject matter. In reading, the United States ranks fifteenth among the thirty-one participating countries, 45 percent of a standard deviation behind Finland, the highest-performing country. One might excuse U.S. performance on the grounds that English is a difficult language to learn, except for the fact that the United States was the lowest-scoring of all the English-speaking countries. Canada, New Zealand, Australia, the United Kingdom, and Ireland all ranked higher (in the order presented). So did Korea, Japan, Sweden, Austria, Belgium, Norway, and France.

In math, the United States scored slightly below the international average and 60 percent of a standard deviation behind Japan, the world leader (see figure 11). On this test, eighteen of thirty-one countries outranked the United States. Most of those with scores below the United States scores were developing countries in Eastern Europe or Latin America. Among well-established, industrial democracies, only Italian schools trailed those of the United States. On the science test, the United States performed only slightly better, this time outranked by thirteen other countries.

Still another international comparison, by the IALS, provides a way of assessing educational quality in the United States from a comparative vantage point. Unlike the IEA and PISA surveys, the IALS was administered not to students but to adults. It was given during the mid-nineties to a cross-section of those aged sixteen to sixty-five

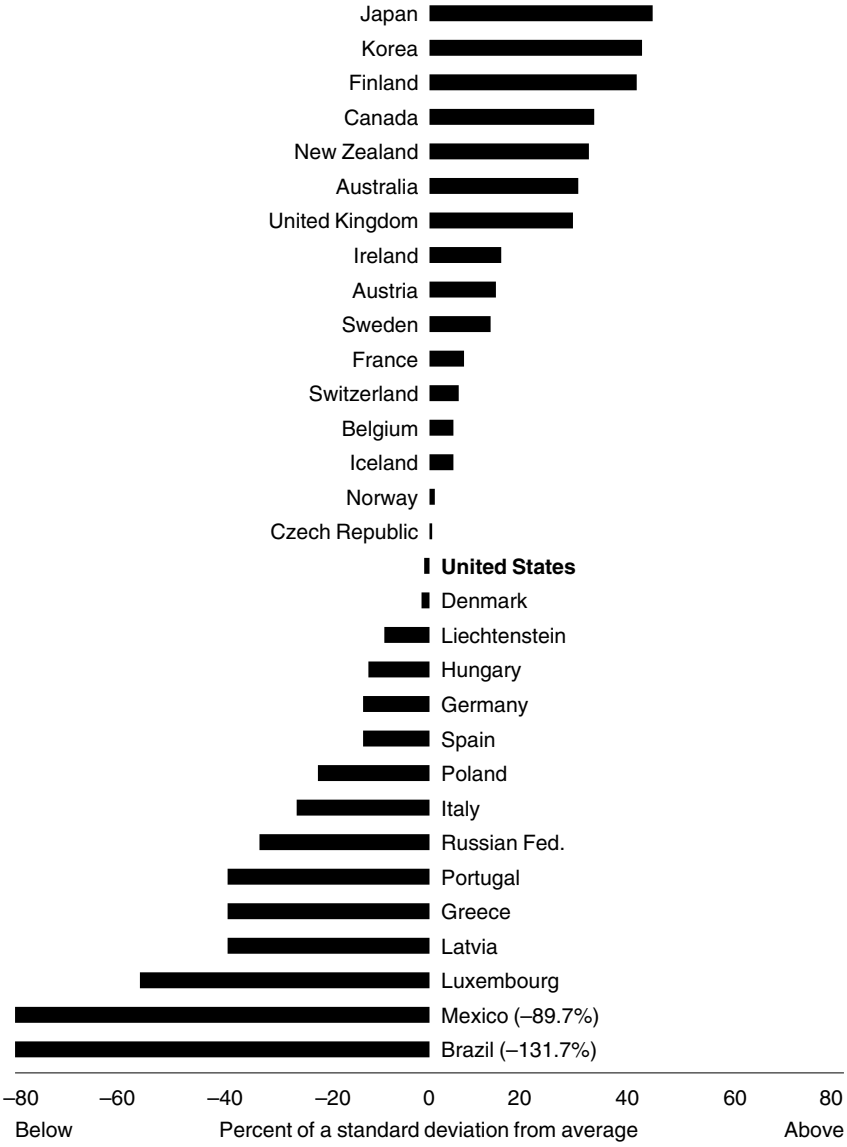

Fig. 9. Student combined test score performance on Program for International Student Assessment (PISA), 2000

Note: Percentage exceeds graph boundary.

Source: Organization for Economic Cooperation and Development, Program for International Student Assessment, Knowledge and Skills for Life, 2000.

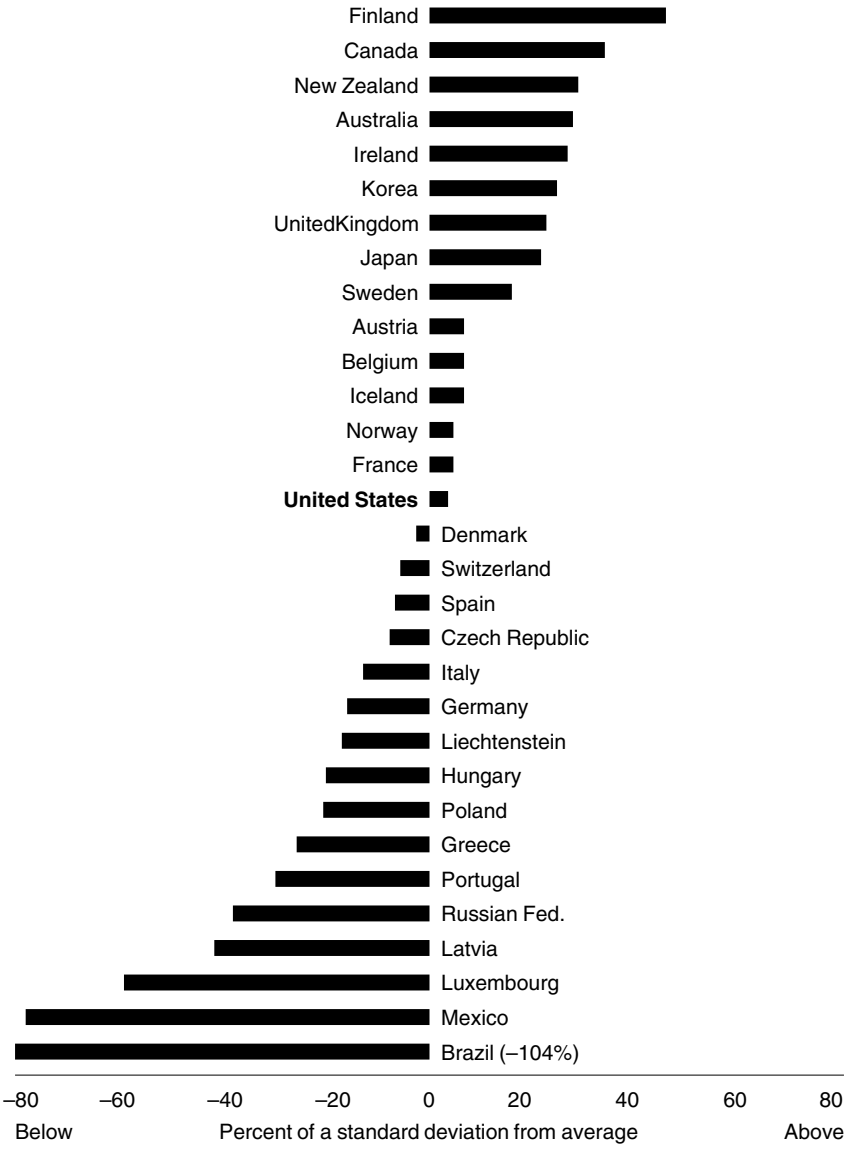

Fig. 10. Student test performance in reading on Program for International Student Assessment (PISA), 2000

Source: Organization for Economic Cooperation and Development, Program for International Student Assessment, Knowledge and Skills for Life, 2000.

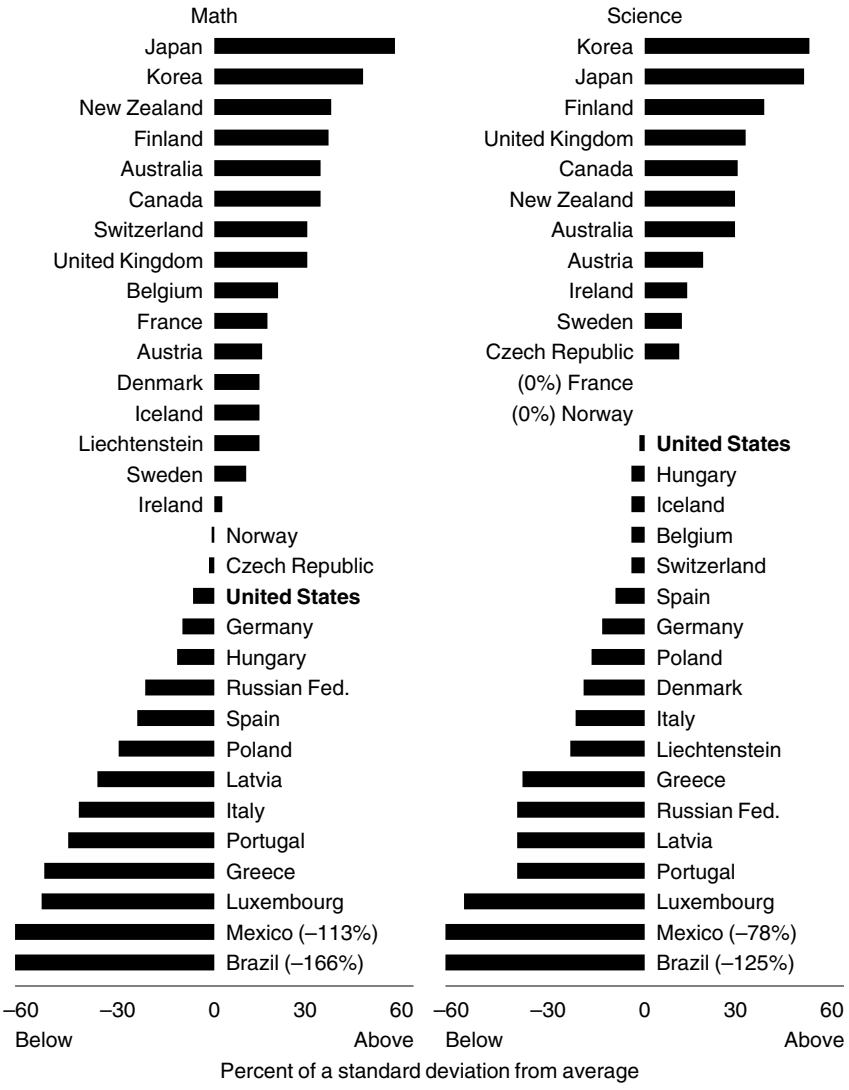

Fig. 11. Student test performance in math and science on Program for International Student Assessment (PISA), 2000

Note: Percentage exceeds graph boundary.

Source: Organization for Economic Cooperation and Development, Program for International Student Assessment, Knowledge and Skills for Life, 2000.

in fourteen European and North American countries. On this test, American adults barely attained the international average set by these advanced democracies, says a report issued by the Educational Testing Service.

The United States ranked twelfth on the test, trailing Sweden, Norway, Denmark, Finland, the Netherlands, and Germany by significant margins. The news gets even worse. The United States is living on its past. Among the oldest group in the study (those aged fifty-six to sixty-five), U.S. prose skills stood in second place. These oldsters had attended school in the fifties, a time when U.S. schools were no-nonsense learning institutions, SAT scores reached heights to which they have never since returned—and Europeans were still trying to put together an educational system that could serve more than an elite cadre.

The younger the age group, the lower the place of the United States on this international list. Americans who went to school during the sixties captured a respectable bronze medal in literacy, while those schooled in the seventies ran a respectable race, so as to get fifth place. But those aged sixteen to twenty-five, that is, those who were wandering school hallways during the nineties, ranked fourteenth. The IALS records a simple, steady progression downward.

Apologists will find excuses for these outcomes, of course. The downward U.S. trajectory is due more to gains elsewhere than to slippage within the United States, some will say, as if we should then not care that it is happening. Or, it will be said, U.S. scores are pulled down by its immigrants and ethnic diversity, overlooking the fact that other countries have immigrants, too. Lifelong learning opportunities are greater in the United States than elsewhere, it will be claimed, so young folks will eventually reach the levels the oldest group has achieved. No matter that schools are bad; catch-up time will come later on.

But such excuses don't ring true, especially when the literacy test only confirms results from the IEA, the PISA, the NAEP, and the SAT.

All signs point to stagnation, perhaps even deterioration, in the quality of American schools.

Conclusions

What is the best way to assess the state of American elementary and secondary education? Should we look at the SAT scores of those high schoolers planning to go to college? Or the NAEP scores of all seniors? Or the PISA performance of fifteen-year-olds, an age when fewer students have left school? Or perhaps the NAEP scores of thirteen-year-olds? Or is it best to get an estimate at age nine? Should we rely on SAT or NAEP results, which look at U.S. trends over time? Or use IEA, PISA, and IALS surveys to compare the United States with other countries? Or should we place greater weight on high school graduation rates than test-score performance?

Any one of these measures may be defective. SAT scores may not have climbed as much as they would have, had participation rates not increased. NAEP scores might have been even lower, had participation rates not declined. There are several ways to measure the graduation rate, and, depending on how it is done, one can obtain somewhat different results. International assessments have their own challenges. Making sure that sampling techniques are equivalent requires great care and sophistication. In addition, reading and literacy assessments may unfairly equate languages that are noticeably different. Nor is this problem eliminated altogether when making science and math comparisons.

So no one measure should be accepted as providing the definitive assessment of American education. But when multiple measures yield similar results, the story they tell becomes ever more compelling. By all accounting devices available since the National Commission on Excellence in Education published *A Nation at Risk* twenty years ago, the United States has not responded adequately to the challenge set forth in the commission's report. The picture the statistics portray is

not such an unrelieved span of gray that it could be taken as a modern artistic masterpiece. A tinge of blue appears here and there. Some gains at the elementary level can be detected. Math scores have not slipped much since the seventies—in fact they may have improved somewhat. But the overall effect is unmistakably grim. The United States has always trailed many other countries in math instruction, and there is no sign it has closed the gap. Verbal skills are even worse. Here there are multiple signs of a downward trend. Most disturbing, all signs of declining quality in test performance among high schoolers are accompanied by a decline in the percentage of students finishing high school. Students are walking away from public schools, choosing other ways of getting the apparent equivalent of a diploma. They seem to understand, better than anyone else, that the American schoolhouse is badly in need of repair.

Notes

1. The 15 percent estimate is drawn from Richard Murnane, John B. Willet, Yves Duhaldeborde and John H. Tyler (2000), "How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?" *Journal of Policy Analysis and Management*, Vol. 19 (4): 547–68. See also Duncan D. Chaplin, 1998, "Raising Standards: The Effects of High School Math and Science Courses on Future Earnings." *Virginia Journal of Social Policy and Law*, 6:1, 111–26. The 20 percent estimate is made by Susan E. Mayer and Paul E. Peterson, "The Costs and Benefits of School Reform," in Susan E. Mayer and Paul E. Peterson, eds., *Earning and Learning: How Schools Matter* (Brookings Institution Press, 1999), 341–54.
2. Sanders D. Korenman and Christopher Winship, "A Reanalysis of the Bell Curve," in Kenneth Arrow, Samuel Bowles, and Stephen Durlauf, eds., *Meritocracy and Economic Inequality* (Princeton: Princeton University Press, 2000), 137–78.
3. Unfortunately, the design of the SAT is being altered in important respects, so it is not clear whether it will remain a useful yardstick for measuring change in American schooling.

4. According to the American College Testing Assessment Web site, the test is an "examination designed to measure academic achievement in four major curriculum areas: English, mathematics, reading, and natural sciences. Materials covered on the four tests that make up the ACT Assessment correspond very closely to topics covered in typical high school classes." On this Web site, ACT, Inc., describes itself as an "independent, not-for-profit organization that provides more than a hundred assessment, research, information, and program management services in the broad areas of educational planning, career planning, and workforce development." ACT Incorporated Home Site. 2002. Accessed 5 July 2002. www.act.org.
5. For a comparison of the two trend lines, see David W. Murray, "Waiting for Utopia," *Education Next* (Summer 2002), 75. Also see Daniel Koretz, "What Happened to Test Scores, and Why?" *Educational Measurement: Issues and Practice* (Winter 1992), 7–11; Daniel Koretz, *Educational Achievement: Explanations and Implications of Recent Trends* (Congressional Budget Office, 1987).
6. The original purpose of this complicated testing design was to preclude measurement for any particular school, school district, or state. When originally proposed, state school officials were concerned that the NAEP would lead to national accountability standards that would intrude on state and local control. Diane Ravitch, *National Standards in American Education: A Citizen's Guide* (Brookings Institution Press, 1995), 70–71.
7. To see whether or not changing participation rates were correlated with test score performance, regression analyses were conducted in order to see whether participation rates were correlated with test score performance when controls were introduced for the year in which the test was administered, and, specifically, whether or not the testing date preceded or postdated the issuance of *A Nation at Risk*. We found a statistically significant impact of participation rates on the test score performance of seventeen-year-olds. For seventeen-year-olds, regression results were as follows: participation rate –63.439, annual trend 55.39, issuance of *Nation at Risk* report –181.15. Constant: 4417.27 Participation was significant at .1 level, one tail test. Other variables not significant. Number of observations: 27. Similar results were obtained from a regression that included a term that controlled for the interaction between the annual

trend and the issuance of *A Nation at Risk*. Although a negative relationship between participation rates and average test score performance is also observed for the younger cohorts, the relationship is not statistically significant in similar estimations of average test scores of nine-year-olds and thirteen-year-olds.

8. Richard Rothstein, "Out of Balance: Our Understandings of How Schools Affect Society and How Society Affects Schools," 30th Anniversary Conference: Traditions of Scholarship in Education, Spencer Foundation, Chicago, Illinois, January 2002. Available at Spencer Foundation Web site.
9. The trends among white students in the younger cohorts also track overall trends fairly closely. For white nine-year-olds, math scores are up by 40 percent of a standard deviation since 1970, with the gains coming after *A Nation at Risk* was issued. Math and science scores are also up slightly since 1970 (10 percent of a standard deviation). The change for thirteen-year-old white students is a gain between 1970 and 1999 of roughly 30 percent of a standard deviation in math, 15 percent of a standard deviation in reading, and 5 percent of a standard deviation in science. As with the nine-year-olds, the math gain occurs after 1982.
10. Although this inflates graduation rates, there is no reason to believe that the prevalence of older students remaining in high schools has changed over time.
11. Stephen Cameron and James Heckman, "The Nonequivalence of High School Equivalents," *Journal of Labor Economics*, 11, no. 1, (1993); Richard J. Murnane, John B. Willett, and Kathryn Parker Boudett," "Do High School Dropouts Benefit from Obtaining a GED?" *Educational Evaluation and Policy Analysis* 17(2), 1995, 133–47.
12. Murnane, Willett, and Boudett.
13. Duncan Chaplin, "Tassels on the Cheap: The GED and the Falling Graduation Rate," *Education Next* 2, no. 3 (Fall 2002), 24–29. Figures 6 and 7 are drawn from this essay.
14. For a full discussion and application of a variation on this estimation technique, see Jay P. Greene, *High School Graduation Rates in the United States*, Center for Civic Innovation at the Manhattan Institute and the Black Alliance for Educational Options, April 2002.
15. Paul Barton, "The Closing of the Education Frontier?" Policy Informa-

tion Report (Research and Development, Policy Information Center, Educational Testing Service, Princeton University, September 2002). Available at www.ets.org/research/pic.

16. Both the title of the TIMSS survey and the number of countries participating in the survey have changed over the decades.