CHAPTER 3

Achievement in American Schools

Herbert J. Walberg

Yield, yield per acre, and crop quality tell agronomists and farmers much about results. Reduction of mortality and morbidity chiefly concern epidemiologists and physicians. Similarly, learning or increases in achievement are or should be the chief aims of educators. Achievement here means the knowledge and skills students learn in the usual school subjects, particularly as measured on standardized multiple-choice, essay, and other examinations.

Parents, legislators, and other educational consumers are concerned with other matters as well, but their views of a school's desirability are usually based substantially on how well they think children learn. Parents know that test scores will often be among the chief determinants of success in gaining admission to selective universities. Firms employ examinations to screen for knowledgeable and skilled workers. Citizens want well-educated young people who contribute to the economy and society and who can competently vote and serve on juries.

Students should be able to read, write, calculate, and reason skillfully; they should possess deep and wide knowledge of standard subject matter. So we need to measure achievement to assess how well students perform and how well schools are preparing them for subsequent education, careers, citizenship, and other aspects of their future lives.

Assessing achievement and where American students stand are big tasks. Should we expect schools to attain higher levels of achievement? Or should we consider how much they progress in a given year? Since poverty and related factors can limit children's learning, how can we compare schools in different communities that contain differing percentages of poor and rich parents?

This chapter answers these and related questions. It shows that standardized tests provide an effective, efficient means for assessing students and schools. For this reason, nations, states, and school districts are turning to such tests to compare schools and evaluate their progress.

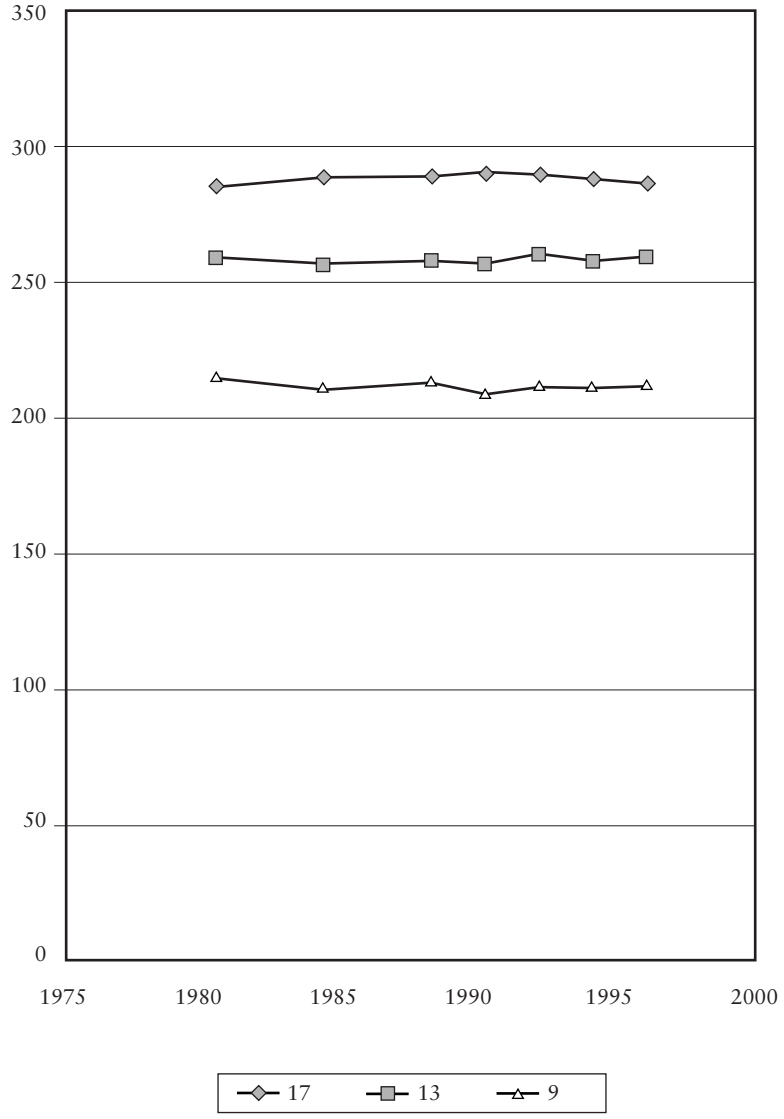
Since policymakers, parents, and citizens are keenly interested in achievement, this chapter shows how U.S. students' progress compares with that in other countries and how their achievement levels have changed for roughly the past quarter-century. In addition, we should consider the efficiency of schools in raising achievement in relation to how much they spend and to changing levels of students' abilities.

Unfortunately, as shown below, achievement has remained generally stagnant with occasional, short, and apparently random up-and-down trends despite steadily and substantially rising expenditures and greater levels of potential student abilities. Findings from international achievement surveys, moreover, show that American students compare unfavorably with those in other economically advanced countries in how much they learn in school, despite the fact that American schools expend more money on students than do schools in most other countries. The problem is not attributable to lower school graduation rates in other countries. The United States, once prideful of high school graduation rates, has fallen behind the average graduation rate of affluent countries.

Achievement Stagnation in the United States

A review of how U.S. test scores have been changing reveals little progress in solving the achievement problem. Reading scores, for example, show no upward trend 1984 through 1996 (see chart 1). Similar to the trends in chart 1, compilations of trends in various subjects and grade levels of children on National Assessment of

Chart 1. Reading Scores of 17-, 13-, and 9-Year-Olds



Source: U.S. Department of Education, *Digest of Education Statistics,* 1998, p. 129.

Education Progress scores show the same essentially flat trends. Although a small upward trend may detected for some years and grade levels, small downward trends may be found for others.[1]

In mathematics from 1988 through 1996 for students in schools with high and low concentrations of poverty, achievement scores for both groups of schools has been stagnant for the whole time period (see chart 2).[2] During this period, partly in response to *A Nation at Risk* and other alarming reports, the schools enacted many varied reforms. Yet achievement levels failed to rise. In private sector industries, we expect to see steady improvements and even breakthroughs attributable to competition, improving technology, and other innovations. Quality should rise while costs decline. Why haven't the schools similarly improved?

Rising Expenditures on K–12 Schools

Rising expenditures on public schools have long failed to increase achievement. The expenditures on public schools have risen substantially and steadily during the period from 1920 through 1997 (see chart 3). They have continued to rise in the recent period when achievement scores for random samples of students have become available and show worrisome generally flat trends of low scores with fitful blips.[3]

Achievement Gaps

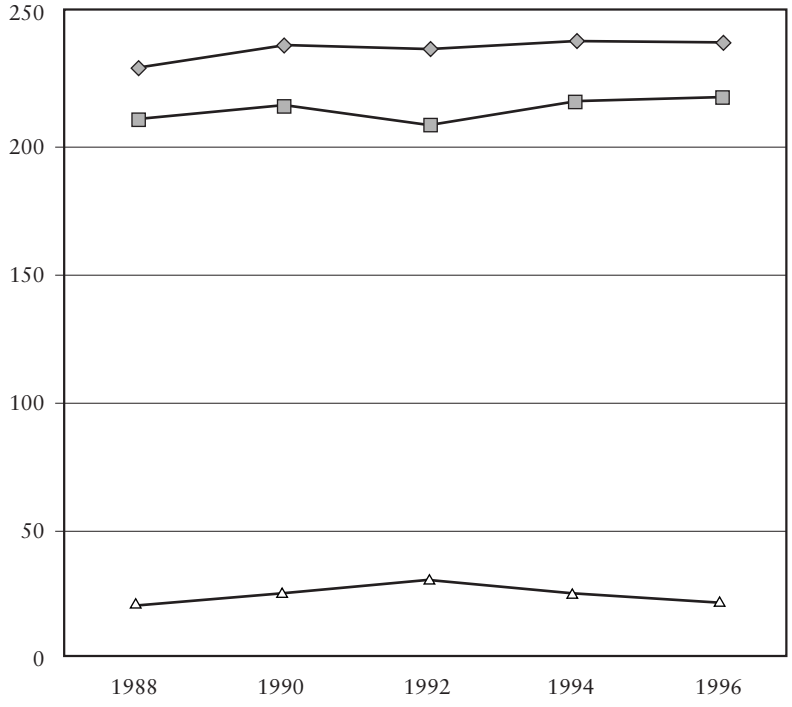
Many studies show that children in poverty often achieve less in school than children in middle-class families. For the past quarter-

1. Source: U.S. Department of Education, *Digest of Education Statistics 1998* (Washington, DC.: Department of Education, 1999), p. 129. See also Stedman's extensive compilations showing similar test score trends discussed in a subsequent section of this chapter (see note 27).

2. The time period chosen was intended to measure results of changes in Title I policy during the period. Office of Planning and Evaluation Service, *Promising Results, Continuing Challenges: The Final Report of the National Assessment of Title I* (Washington, D.C.: U.S. Department of Education Office of the Under Secretary, 1999)

3. Source: U.S. Department of Education, *Digest of Education Statistics, 1998* (Washington, DC.: Department of Education, 1999), p. 35.

Chart 2. Trends in Math Performance among 9-Year-Old Public School Students in Low- and High-Poverty Schools

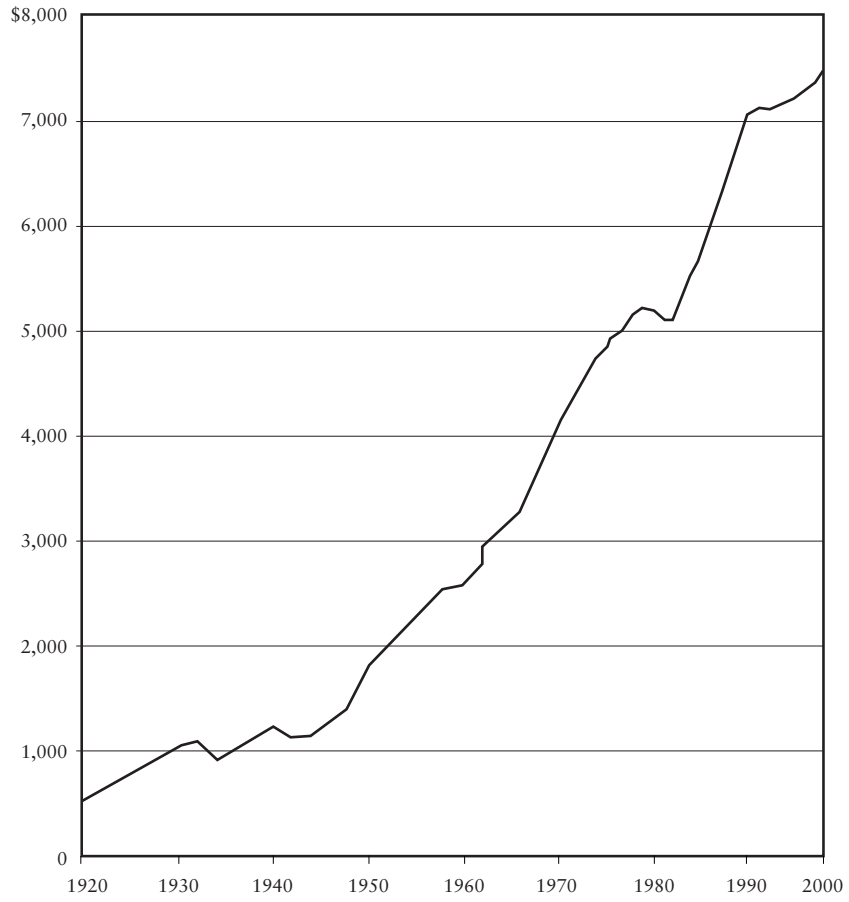


	1988	1990	1992	1994	1996
LOW	228	237	236	239	238
HIGH	208	213	206	215	217
GAP	20	24	30	24	21

Source: U.S. Department of Education, 1999, p. 6. The scale ranges from 0 to 500; high poverty schools had 76–100 percent students eligible for free lunch, low-poverty 0–25 percent. *Digest of Education Statistics*, 1998

century, the federal government has concentrated about $130 billion on Title 1/Chapter I programs on reducing the gap between children in poverty and other children. Despite expenditures at a current rate of about $8 billion, the gap between schools with high concentrations of children in poverty and other schools has remained the same (see chart 2). The huge expenditures appear to have done little good in reducing the gap.

Chart 3. Total per Pupil Expenditures in Public Elementary and Secondary Schools, 1919 to 1998 in Constant 1997–98 Dollars



Source: U.S. Department of Education, 1999, p. 35.

The Black and Hispanic Gaps in Achievement

Since the early 1970s, the National Assessment of Educational Progress (NAEP) and other groups have reported that on average black and Hispanic students have persistently and substantially scored below white students. For any given subject or group, the various trends are flat, encouraging, or discouraging; and no consistent overall trend since 1970 emerges from the many comparisons that have been made. In reading, for example, NAEP data reveal that black seventeen-year-olds had improved to the point that their scores were equal to white thirteen-year-olds by 1990, but there were no further improvements during the decade of the 1990s.[4] The patterns for different subjects and groups are similarly complex and cannot be easily summarized, but scholars agree that the differences among groups are large and enduring.

Scholars have expended much more effort in trying to explain the black gap than the Hispanic gap, but a consensus is yet to emerge. The anthropologist John Ogbu, for example, argued that the gap is ultimately attributable to the forced immigration and slavery of African Americans' ancestors.[5] Others argue that historically deficient education systems in the deep South and inferior schools for rural blacks in southern states harmed achievement, which generation after generation has yet to recover, even after blacks migrated to big industrial cities in the North during the World War I and II years.

In *Losing the Race: Self-Sabotage in Black America*,[6] University of California at Berkeley faculty member John McWhorter reviews some of these possibilities but concludes that black "victimology" is today's primary reason that blacks achieve less than whites. In his view as a black faculty observer at the university, a cult of

4. Jay R. Campbell, Catherine M. Hombo, and John Mazzeo, *NAEP 1999 Trends in Academic Progress* (Washington, D.C.: National Center for Education Statistics, 2000).

5. See, for example, "Variability in Minority School Performance: A Problem in Search of an Explanation," *Anthropology and Education Quarterly* 18, no. 4 (1987): 312–34.

6. New York: Free Press, 2000.

victimology transforms a problem to be solved into a persistent black identity of anti-intellectualism, separatism, and cultural disconnect from learning. Recent letters to the editor of *Wilson Quarterly*, however, have sharply disputed his view.[7]

Large-scale surveys suggest to me that the achievement gap for both blacks and Hispanics may be substantially attributable to poverty and the inefficiency of large-city school systems where minority groups are concentrated.[8] Although there are more poor whites than poor blacks and poor Hispanics, the poverty *rates* among these minority groups are higher. Sociologists have shown that the differences among whites and minority students of the same socioeconomic status (SES) are relatively small,[9] which supports this explanation.

In addition, socioeconomically related differences in achievement-stimulating child-rearing patterns between middle-class parents and those in poverty are huge. In a rare and careful observational study, psychologists found that higher SES parents spent more minutes per hour interacting with their children and spoke to them more frequently. On average, higher SES parents spoke about two thousand words an hour to their children; welfare parents, only about 500. By age four,

> An average child in a professional family would have accumulated experience with almost 45 million words, an average child in a working-class family would have accumulated experience with 26 million words, and an average child in a welfare family with 13 million words.[10]

Higher SES parents, moreover, used

> more different words, more multi-clause sentences, more past and future verb tenses, more declaratives, and more questions of all kinds.

7. Autumn 2000, pp. 6–7

8. Herbert J. Walberg and Herbert J. Walberg III, "Losing Local Control, " *Educational Researcher*, June/July 1994, 13 (8), 23–29.

9. For a recent collection of sociological articles on this complex subject, see Christopher Jencks and Meredith Phillips, editors, *The Black-White Test Score Gap* (Washington, D.C.: Brookings Institution, 1998).

10. Betty Hart and Todd R. Risley, *Meaningful Differences in the Everyday Experience of Young American Children* (Baltimore, Md.: Paul Brooks Publishing, 1995), p. 198.

> The professional parents also gave their children more affirmative feedback and responded to them more often each hour they were together.[11]

The researchers estimated that, by age four, professional parents encouraged their children with positive feedback 750,000 times, about six times as often as did welfare parents. The welfare parents, on the other hand, had discouraged their children with negative feedback about 275,000 times, about two and a half times the amount employed by professional parents.[12] Such parenting behaviors predicted about 60 percent of the variation in vocabulary growth and use of three-year-olds. Vocabulary is the most important single predictor of school success.

A second reason for the minority gaps is inefficiency of big-city school systems that have employed ineffective federal programs such as Title 1/Chapter 1 and bilingual education. Because of their size and greater accountability to federal and state governments than to local citizens and parents, moreover, big-city systems appear relatively indifferent to students and parents. They have gained a well-known and often deserved image of failure. Such big-city achievement gaps, apparent indifference, and inefficiency help explain minority parents' strong desire to choose their children's schools including those in the parochial and independent sectors.

Achievement Stagnates Despite Children's Rising Abilities

Long before the era when achievement stagnation was documented, students' *abilities* rose massively and steadily, and they continue to rise. From 1918 through 1995, school children's average IQ steadily rose 25 points.[13] This steep rise put the typical

11. Ibid., pp. 123–24.
12. Ibid., p. 200.
13. James R. Flynn, "IQ Gains over Time," in Ulric Neisser, ed., *The Rising Curve: Long-Term Gains in IQ and Related Measures* (Washington, D.C.: American Psychological Association, 1998), p. 37, figure 2. Test makers renormed IQ tests to make them more difficult over the period to reset them to a population average of 100. Flynn discounted these renormings and adjusted the average IQ estimates to make them comparable throughout the period for which scores are

1995 child at the 95th percentile of the 1918 distribution. These changes plus rising expenditures, better programs, and more skilled teaching should have led to ever-higher levels of academic learning. But, as shown in previous sections, this isn't so. Achievement stagnated.

International Comparisons of Achievement

The 1983 report to the U.S. secretary of education, *A Nation at Risk,*[14] first alerted policymakers and citizens that U.S. students achieve poorly compared to those in other economically advanced countries. Since then, as discussed below, more definitive international surveys have shown that the situation is even graver than originally thought. Not only do American students achieve poorly but they fall further behind the longer they are in school.

Value-Added Achievement Comparisons

The challenge in comparing schools is like comparing runners who begin a race from different starting points. To know runners' speeds or rates, we would need to measure how far they run in a given time period. Similarly, to compare the effectiveness of schools, we need to ask how much their students improve in achievement over a given time period, which is what is meant by "value-added gains." Children come to one school, for example, already reading, whereas children in another school are far less well prepared. To have the full picture, of course, we should know their gains as well as their scores at the end of any given grade. The gains, however, are particularly important for assessing schools because they are more fully attributable to the school's effectiveness in educating students than are their final achievement levels, which may have been largely determined by their family socioeco-

available. Now, as widely reported in psychological journals, the tendency for intelligence test scores to rise is called the "Flynn effect."

14. The National Commission for Excellence in Education (Washington, D.C.: U.S. Department of Education, 1983).

nomic status, child-rearing conditions in the home, and other external factors, especially those that influence the children's development before they start school.

Because of widespread interest in such comparisons throughout the world, the Paris-based Organization for Economic Cooperation and Development (OECD) began comprehensive reporting of such value-added improvements for economically advanced countries in Asia, Europe, and North America.[15] As an example, chart 4 shows the mathematics value-added achievement gains made by students in twenty-four countries from the fourth to the eighth grade.[16] Of the countries surveyed, the United States made the smallest gain.

Unfortunately, this result for U.S. schools is typical of the other comparisons of presently available data. In reading, science, and mathematics through eighth grade, U.S. schools ranked last in four of five comparisons of achievement progress. In the fifth case, they ranked second to last. Between eighth grade and the final year of secondary education, U.S. schools slipped further behind those in other countries. Because they made the least progress, U.S. secondary schools recently ranked last in mathematics attainment and second to last in science, a result that does not accord well with the National Education Goals Panel objective set about a decade ago that American students will be first in the world in mathematics and science.[17] Actually they are last or near last among students in other OECD countries.

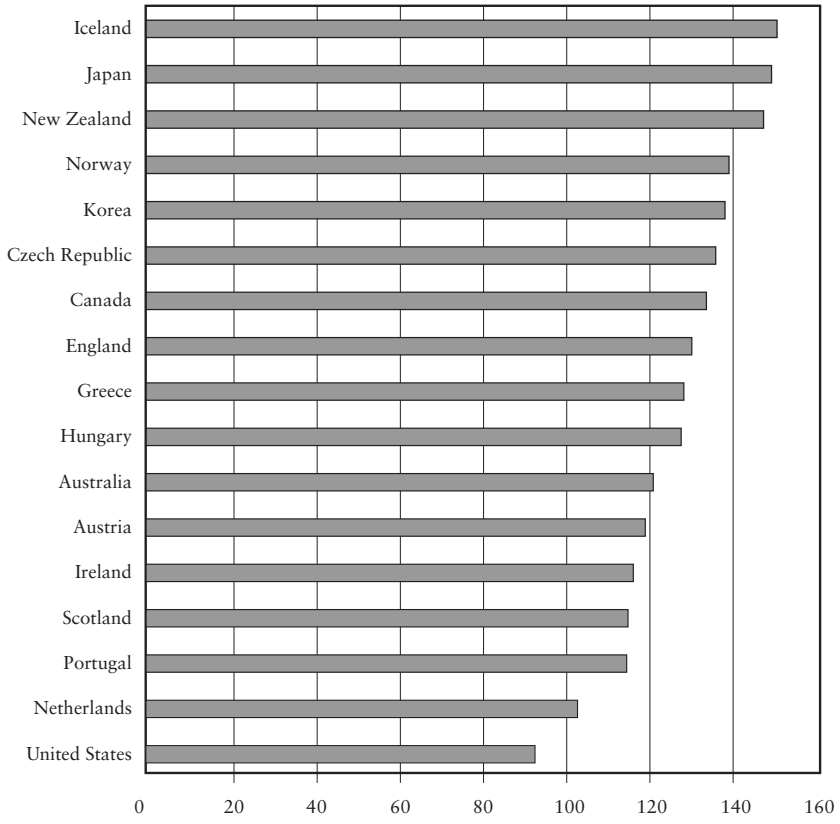
Policymakers commission international surveys of achievement

15. As of 1998, all recent value-added comparisons are in my report for the Thomas B. Fordham Foundation, *Spending More while Learning Less* (Washington, D.C.: Thomas B. Fordham Foundation, July 1998). The original data may be found in the periodic reports of the Organization for Economic Cooperation and Development's *Education at a Glance* (Paris: OECD, 1996, 1997, 1998, and 1999). The charts in this chapter illustrating the country comparisons are based on information in the 1997 report, pages 101 and 306.

16. Strictly speaking, these scores are the differences in random samples of students in fourth and eighth grade at a single point in time.

17. For the accomplishment or nonaccomplishment of various national goals, see reports of the panel such as *The National Educational Goals Report: Building a Nation of Learners*. (Washington, D.C.: U.S. Government Printing Office, 1998).

Chart 4. Mean Difference in Mathematics Achievement Scores between Fourth and Eighth Graders in Various Countries



Source: Walberg, 1998, p. 8.

in reading, mathematics, and science because these subjects are more internationally comparable than, say, civics, history, geography, and literature. Schools in various countries emphasize different aspects of these subjects, which would make the achievement scores less comparable than the subjects most often compared. Many policymakers, moreover, believe that mathematics, reading, and science are particularly important for preparedness for higher education and the workforce—reading because verbal mastery is

an essential skill in nearly all subjects and economic and social life, mathematics and science because they indicate readiness for further study in demanding fields such as engineering and medicine as well as work in high-tech, fast-growing, competitive sectors of modern economies.

Graduation and Completion Rates

When international achievement surveys began some thirty-five years ago, U.S. policymakers aimed to graduate all students from secondary school. In fact, greater percentages of U.S. students then graduated from secondary school than those from other economically advanced countries. Because greater percentages of U.S. students remained in school, they were included in the surveys, whereas more students comparable in age in other countries had dropped out, leaving a comparative elite. Thus, it might have been argued that poor American school achievement at the end of secondary school was partly excusable years ago because a less selective American group was being compared with more selective groups in other countries.[18]

In the meantime, however, European and other OECD countries made serious and successful efforts to keep more students in secondary school and graduate them. On average, schools in these countries now succeed better than do U.S. schools. The most recent OECD comparison shows that the United States ranks seventeenth among twenty-three OECD countries in the graduation rates (percentages of secondary school graduates to the total popu-

18. Perhaps the early results were not excusable because the twelfth grade comparisons were somewhat questionable in that they were based on American and other students taking college preparatory mathematics. The U.S. students seemed to be a similarly select group of students to those elsewhere. Some comparisons, moreover, were made of only the top 10 percent of students. Even in these early comparisons, our "best and brightest" still did not do well. In any case, as explained above, recent comparisons of a somewhat more select group of American students with less select groups in other countries show that American students do worst or near worst in secondary school and have fallen furthest behind during the school years.

lation at the typical age of graduation).[19] The average percentages of students aged 14–17 and 18–19 enrolled in education were also higher in OECD countries than in the United States.[20]

Thus, U.S. schools have fallen behind the graduation and enrollment rates of other economically advanced countries. This trend shows that the poor achievement progress of U.S. schools isn't attributable to educating a less selective group of students.

Productivity: Achievement in Relation to Expenditures

In considering country comparisons of achievement, it is important to know how much they spend on each student. Schools should not only be effective but efficient or productive given the amounts of money they spend. The OECD regularly reports per student expenditures on primary and secondary schools.[21] Among twenty-two countries, U.S. expenditures on primary schools[22] were third highest after Switzerland and Austria (see chart 5). They were also third highest for secondary schools. Thus, U.S. per student expenditures were among the highest, yet our achievement progress was generally worst.

For a country that leads the world in the competitiveness and productivity of many old and new industries, it is shocking that

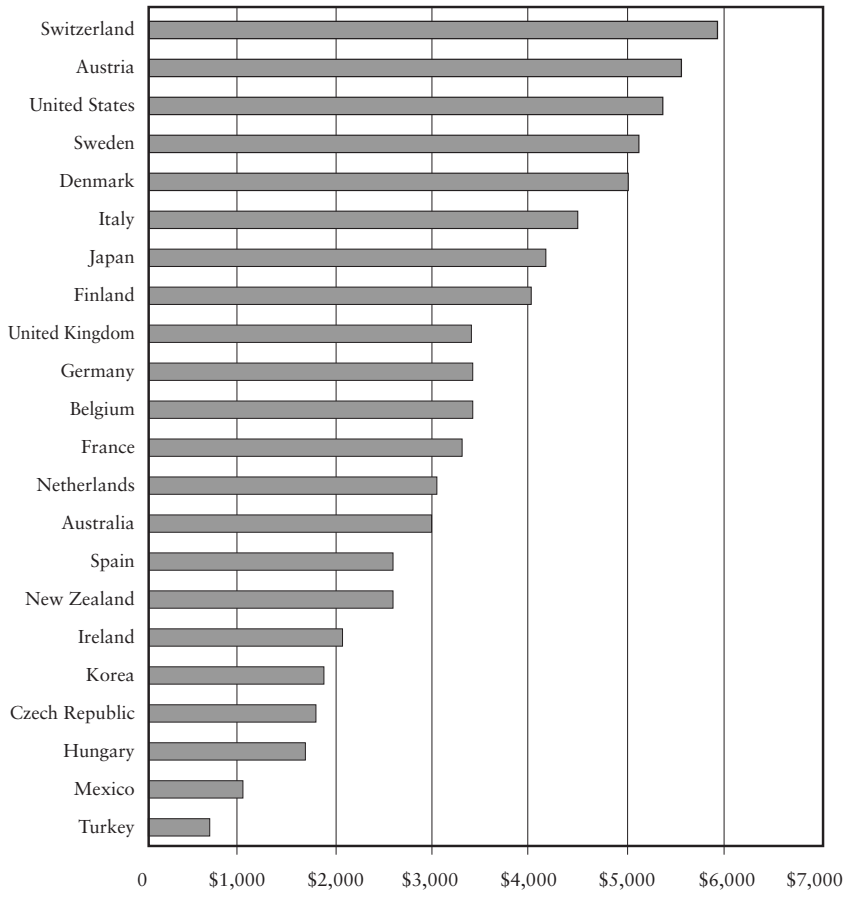
19. *Education at a Glance* (Paris: OECD), 2000, p. 147. The U.S. percentage of 74 is lower than the average of 79. The average, however, includes several less affluent, recent entrants into the OECD such as Mexico, Portugal, Spain, and Turkey with graduation rates as low as 30 percent.

20. Organization for Economic Cooperation and Development, *Education Policy Analysis* (Paris: OECD, 1997), pp. 14 and 98. OECD, *Education Policy Analysis*, 1998, p. 75.

21. See my previously footnoted report *Spending More while Learning Less* or recent editions of the OECD's *Education at a Glance*. The figures reported are OECD adjusted for purchasing power parity. See Eric Hanushek's chapter in this volume for a detailed analysis of expenditures.

22. In writing on comparative education, the term *primary* usually refers to the first school, aside from preschool, that children usually attend, which is in many countries the first six years of schooling. This roughly corresponds to what Americans often call grade schools or elementary schools, although there are many U.S. variations such as grades 1–6, 1–5, and the more traditional 1–8. Secondary schools are the second schools students attend, which in the United States are usually referred to as high schools, junior high schools, or middle schools.

Chart 5. Expenditure per Student for Primary Education in Public and Private institutions



Source: Walberg, 1998, p. 11.

American schools are so inefficient. Among the consequences is that schools fail to pull their weight in improving the quality of American life. Their graduates are less literate, less skillful, less informed as citizens, voters, and workers than they should be. Because they constitute a large sector of the economy, schools are a drag on American progress and wastefully consume resources. Citizens are not receiving value for their tax dollars. In such fields as science, mathematics, engineering, computer science, nursing,

and medicine, we must import better-prepared, well-educated workers from other countries, particularly South and East Asia.[23]

Why the Defenders of the Status Quo are Wrong

Scholars originally uncovered the substantial U.S. achievement gap about three decades ago. As discussed above, recent research has confirmed the early findings and suggests that American achievement problems are actually worse than long thought. Despite the huge amount of evidence, three writers have dismissed the validity of the unrelenting findings. Gerald Bracey[24] often takes this view in his monthly column for *Phi Delta Kappan*, a widely circulated education journal. David C. Berliner and Bruce J. Biddle wrote *The Manufactured Crisis: Myths, Fraud, and the Attack on America's Public Schools*.[25]

The Positions of the Critics

Because they have achieved some prominence, these writers deserve to be answered with respect to how their work has been published and their arguments. Gerald Bracey publishes much of his criticism of the achievement surveys in columns and annual education reports in the journal of Phi Delta Kappa, a society with the expressed purpose of promoting and defending public education. Indeed, the title of Bracey's annual reports contains the words "Public Education."[26] Neither his work nor Berliner and

23. Immigration is to be celebrated. We are of course a nation of immigrants and their descendants, and we still benefit much from new arrivals. That does not mean, however, that that American students cannot or should not have to learn or that schools should be excused for poor results.

24. See, for example, "The 10th Bracey Report on the Condition of Public Education," *Phi Delta Kappan*, October 2000, pp. 133–144. In this and other reports, Bracey describes how he has long called reporters to urge on them his dismissal of education research findings critical of public schools.

25. New York: Perseus Press, 1996.

26. Bracey doesn't hesitate to employ ad hominem argumentation. He, for example, has been allowed to give "Rotten Apple Awards" to presidents from both political parties, reporters from nationally circulated newspapers, and prominent scholars for criticizing education. Because Bracey is given the first, last, and regular word in the journal, few people correct his faulty arguments.

Biddle's book has gone through scholarly peer reviewing as would be required in journal publication in education, psychology, and the social sciences.

On the other hand, the data in reports summarized above on achievement changes in the United States over roughly the last thirty years are collected by the Educational Testing Service and similar agencies for the U.S. Department of Education. The plans for data collection and reporting undergo intense scholarly and government scrutiny. The national press covers the reports in detail, often on the front page of such respected papers as the *New York Times* and the *Wall Street Journal*.

The international comparisons come from perhaps the most widely respected international statistical data agency in the world, the Organization for Economic Cooperation and Development, of which about thirty advanced countries in Asia, Western Europe, and North America are voluntary members who seek to learn from one another through cooperative data collection and analysis how to improve the education, medical, industrial, tax, and other systems.

Could or would the many scholars, thirty national governments, and such premiere national and international agencies "manufacture" fraudulent data just to please the alleged nefarious conspirators who seek to make U.S. schools look unjustifiably bad? Would this not result in congressional hearings, exposés, and rolling heads of education malefactors?

The Substance of the Critics' Arguments

In five refereed publications of the twenty thousand-member American Educational Research Association and in a distinguished publication series of one of the oldest and most prestigious think tanks, the Brookings Institution, Lawrence C. Stedman [27] evalu-

27. "International Achievement Differences: An Assessment of a New Perspective," *Educational Researcher* 26, no. 3 (1997): 4–15; "Incomplete Explanations: The Case of U.S. Performance in the International Assessments of Education," *Educational Researcher* 23, no. 7 (1994): 24–32; "An Assessment of the Contemporary Debate over U.S. Achievement," in D. Ravitch, ed., *Brookings*

ated the contentions of Bracey, Berliner, and Biddle (BBB) that American education is unfairly or even conspiratorially portrayed by the NAEP and international achievement surveys. His many detailed examples showing they are wrong cannot be described here, but several of Stedman's major points are well worth summarizing.

The main flaw in BBB's writings, as Stedman points out, is selective evidence that suits their defense of the status quo for public schools. Actual evidence is crucial to their argument; but, instead of reviewing all of it, they select a few data points for a few years in a few subjects. They fail to point out, for example, that NAEP high school science scores remain lower than they were in 1969. BBB concentrate, moreover, mostly on trends, which they say, correctly, are fairly stable, but they ignore the levels of achievement, which are mostly low, especially the longer American students are in school.

As Stedman further points out, BBB correctly criticized the myth of a very recent general achievement decline, but they left out the sharp 1970s decline on many high school tests and ignored the contradictory evidence of the 1980s. BBB claimed that the current generation of students outperforms prior ones on "virtually all" commercial standardized achievement tests—a contention refuted by major reviews of historical trends on such tests, including Stedman's.

BBB dwelt on Scholastic Aptitude Test comparisons, but these are nearly useless for comparing achievement scores across time. The SAT, as its name implied, is an aptitude test used for college selection, not an achievement test for measuring student progress. Rather than constituting a random sample as in the NAEP, the group taking the SAT is voluntary and the fractions of students that take it at given times and states vary substantially, thereby vitiating any comparisons among them.

Papers on Education Policy (Washington, D.C.: Brookings Institution, 1998), pp. 53–121; "Respecting the Evidence: The Achievement Crisis Remains Real," review of "The Manufactured Crisis," *Education Policy Analysis Archives* 4, no. 7: http://olam.ed.asu.edu/epaa/v4n7.html; L. C. Stedman, "The Achievement Crisis Is Real," review of "*The Manufactured Crisis,*" *Education Policy Analysis Archives* 4, no. 1 (1996): http://olam.ed.asu.edu/epaa/v4n1.html.

BBB try to excuse U.S. students' poor performance on the grounds that they don't study what is on the international tests. But, as Stedman points out, they base this contention on only one study of eighth-grade math data from 1981–82, and the data did not support their claim, nor do more recent studies.

Stedman also points out that BBB claim the international assessments improperly compared the broad mass of U.S. students to an overseas elite attending highly selective schools. This is an old criticism from the earliest studies. As discussed above, the average OECD country now has greater percentages of age-eligible youth in and graduating from secondary school than does the United States.

Finally, aside from Stedman's points, BBB do not deal with any of the new research reviewed above. This research shows that, despite greater levels of spending and rising student abilities, achievement has not risen. In addition, recent value-added analyses show American students improve less than do students in other countries despite the generally higher American levels of per student expenditures.

Achievement Tests

For those who have long been away from schools, it may be useful to know about the current means of measuring achievement. In keeping with the spirit of this book, it may be useful to overview some of the major means and issues of testing students' achievement.[28]

Considerable research shows that frequent testing with essay questions, short-answer, and multiple-choice tests leads to higher achievement because students prepare more frequently and regularly to be evaluated and because frequent tests provide more information to both teachers and students about their strengths and

28. A forty-one-page booklet further explaining current testing policies is Herbert J. Walberg, Geneva D. Haertel, and Suzanne Gerlach-Downie, *Assessment Reform: Challenges and Opportunities* (Bloomington, Ind.: Phi Delta Kappa, 1994).

weaknesses.[29] Teachers may also observe and rate their students' performance in class. They may assign, for example, laboratory exercises in science, physical measurements in geometry, and essays in history and literature. Then they may judge or rate the quality of the resulting work. For additional assessment and feedback, teachers may also check their students' homework and either grade or comment on it. Such assessments may be termed *teacher-aligned* or integrated with instruction because they correspond to content of the immediate lessons being taught.

Standardized Tests

For several reasons, such teacher assessments do not serve well in large-scale surveys of achievement intended to provide information on how students, schools, districts, state, and even nations compare with one another, how they compare with established standards, or how achievement is changing over time. Tests intended for this purpose are "standardized" in that the conditions and timing of the tests are nearly identical for all students. Many sports and the international Olympics employ similar standardized conditions so that athletes' performances can be fairly compared no matter where they are and whatever the date of their performances.

Standardized tests widely sample the subject matter. In this respect, they are like national voter and consumer surveys that sample, say, a thousand people, to provide information on the entire adult population with a probable sample error of less than several percentage points.[30] Sample surveys provide information quickly,

29. With a colleague, I compiled 275 effect sizes of frequent testing, homework, mastery learning, direct instruction, and other educational methods and conditions. They serve as indicators of which methods and conditions have the largest impacts on achievement. Along with costs and other considerations, such effects can serve as the basis of formulating policies likely to improve achievement. See Herbert J. Walberg and Jin-Shei Lai, "Meta-Analytic Effects for Policy," in Gregory J. Cizek, ed., *Handbook of Educational Policy* (San Diego, Calif.: Academic Press), pp. 419–52.

30. This refers only to sampling error with respect to the time of the survey and the question asked. If the questions differ or the times differ from the intended forecast, such surveys, of course, may be grossly inaccurate.

efficiently, and cheaply. So, too, can thirty to sixty multiple-choice questions about a broad subject constituted by thousands of facts and ideas.

Sample surveys differ from a census aimed at getting information from every single member of the population. Analogously, test designers usually cannot take an achievement census by asking all conceivable questions about a subject. So they may divide the subject into various skills and areas of content, then sample within each, just as survey designers employ stratified sampling, that is, sampling within cities, suburbs, and rural areas in the several geographic regions of the nation.

Objective, Multiple-Choice Tests

So that aspects of the subject may be sampled in a short time, achievement surveys generally employ multiple-choice examinations. Thirty items may be administered in the amount of time required to answer a single essay question. Multiple-choice questions afford a much larger sample of students' knowledge and skills than do essay questions. They are also fairer to students because their scores do not depend heavily and arbitrarily on whether they happened to have concentrated or not on only one narrow aspect of the subject.

Standardized tests are fairer in another sense: skilled essayists can write impressively on many topics without really having mastered them. We may, of course, be interested in writing ability and require an essay to measure it, although multiple-choice tests can even provide a quicker, more objective, and, some argue, more accurate estimate of writing ability.[31]

Other things being equal, essays in neat handwriting get better

31. A obstacle in measuring writing ability is getting agreement from experts on definitions, especially for anything that goes beyond the conventional matters of spelling, grammar, and appropriate word usage. If there were greater consensus about cohesion, coherence, concision, organization, and elegance as components of style, graders and computers might more reliably and validly mark essays. See Joseph M. Williams, *Style: Ten Lessons in Clarity and Grace* (Reading, Mass: Addison-Wesley, 2000).

grades. Also, the same essay may be given wildly different marks by different graders or the same grader on two occasions. On the other hand, with perfect objectivity, machines can quickly and cheaply score thousands of multiple-choice tests and also produce detailed diagnostic reports on individual and group strengths and weaknesses.

A final reason that multiple-choice tests are preferred in large-scale achievement surveys is that "constructed response" tests requiring essays, laboratory equipment, calculators, and the like usually add little information value to students' scores on objective tests. The score on the multiple-choice test often serves as a better predictor of an essay grader's mark than another essay grader's mark of the same examination. So, for large-scale surveys, the large extra cost of essay examinations is usually unwarranted by the marginal information they may provide (except, as pointed out above, possibly when educators want to encourage and measure essay writing as separate from knowledge and skills in a subject such as history, literature, or science).

Test Criticism

Defenders of the status quo often wrongly criticize multiple-choice tests as "multiple-guess tests." On a test with four options per item, we would, for example, expect students to guess about 25 percent of the answers correctly if they knew nothing about the subject. Guessing, however, can be taken into consideration in scoring, either by various guessing corrections or grading relative to other students. In any case, this criticism is beside the point since standardized test scores are rarely reported as the percentage of items correct.

Some critics attack multiple-choice tests for measuring recognition rather than recall or construction of the right answer. New objective test formats allow answer-sheet recording of recalled or constructed answers that may be objectively machine scored. In any case, recognition, recall, and construction are usually very highly correlated. It may be impossible to find someone who can correctly recognize the facts in a subject such as biology or history

but can neither recall any nor reason about them. An old colleague of mine once asked Einstein's collaborator, who disliked objective tests, to find the student among a national sample of several hundred who had best answered a high school physics problem set. It turned out that that student had also attained the highest score on the objective test of physics knowledge and skills.

Little wonder that even our oldest professions such as law and medicine employ multiple-choice tests for admission and certification, as do graduate colleges and MBA programs for admission decisions. If these demanding fields employ multiple-choice tests, to what do schools aspire that cannot be similarly measured?[32]

Promising Solutions for Raising Achievement

For the reasons stated above, politicians, businesspeople, citizens, and parents are greatly concerned about the inefficiency of the public schools and their threat to the economy and society. Many business leaders describe their problems in getting competent workers, capable of reading, calculating, and learning new material.

The crisis is most acute in technology, an important growth area of the economy. The Information Technology Association of America reported that 1.6 million new information workers would be needed by the end of the year 2000 but that 850,000 positions will go unfilled. According to its survey, technical-support representatives, database developers, and programmers are the three positions in greatest demand.[33]

Achievement Standards

In response to the continuing education crisis, legislators are setting forth two kinds of accountability, both employing achieve-

32. Many examples of multiple-choice items that measure both knowledge and "higher order cognitive skills" can be found in textbooks on testing such as Anthony Nitko's *Educational Assessment of Students* (Englewood Cliffs, N.J.: Prentice-Hall, 1996).

33. Rachel Emma Silverman, "Employers Face Dearth of IT Workers as Demand Exceeds Supply, Data Shows," *Wall Street Journal*, April 10, 2000, p. A1.

ment tests to measure results. The first is more rigorous state standards with achievement tests as indicators of success and failure. Students who cannot meet these standards may be retained in grade or fail to graduate from high school. Schools with sufficiently high failure rates risk being closed. States are phasing in these standards and giving schools some time to gear curricula and instruction to the standards. Because the effects of poverty and other factors are difficult to overcome, it would also seem wise to use value-added achievement scores in any index of a school's success, at least for a time.

Such "high-stakes" achievement test schemes, however, are hardly foolproof. The tests may be technically flawed. The standards may be arbitrarily high or low. Test security may be breached. Schools may teach narrowly to the tests or only to the types of content or problems known to be tested; they may ignore other worthy education goals. Better design and administration of testing programs can overcome these problems but may require steadfastness, money, and experience.

School Choice

The second form of accountability makes schools more directly accountable to parents rather than school boards who may not represent citizens and parents well or have management skills to hold educators accountable. Among the variations of this form are charter schools that give private boards public funds to educate students that come to them. In another form, scholarships are given directly to parents, who may spend them at parochial, independent, and for-profit schools of their choice. Charter schools and scholarships provide a greater diversity of curricula, approaches to education, and means of instruction. They also provide for greater competition among providers and choice for customers, that is, parents, children, and youth.

The ideals of choice and competition have led to wondrous results for consumers in nearly all other sectors of the economy in the United States and elsewhere. Since funds for charter schools and public scholarships come from public sources, it is reasonable,

perhaps even necessary during the experimental period of, say, the next five years, to employ achievement tests to measure their progress.

Legislatures or their appointed commissions might insist that, to continue, charter schools and schools that receive publicly funded scholarship students meet a minimum level of standards as measured by independently designed achievement tests. So as to avoid the deadening hand of government regulation, the standard might be the average achievement test scores of public schools in the state in the usual school subjects plus knowledge of the Declaration of Independence, the U.S. Constitution, and important state documents.

Alternatively, the marketplace provides the ultimate accountability to customers. Unhappy ones can walk away. They can avoid providers of goods and services that make them dissatisfied. For this reason, charter and scholarship schools might only be required to engage in state-mandated testing for a limited period—say, five years—until their worth is clearly demonstrated in achieved results.

CONCLUSION

Despite huge amounts of research over many years on what best promotes learning, American schools lurch from fad to fad. They fail to make evidence the basis of practice in the manner of business, medicine, agriculture, law, and other fields. As a consequence, American schools produce the worst achievement results at the third-highest expenditures among economically advanced countries. Achievement scores have remained stagnant, moreover, despite substantially rising expenditures, rising children's abilities, and many reforms. Substantial amounts of money for special programs, more than $120 billion, have failed to reduce the achievement gap between poor and middle-class children. The schools fail to employ new technologies for learning or even to employ traditional technologies well. Current demands for standards, accountability, incentives, and choice described briefly in this chapter and at length in the other chapters are clearly warranted.