

10

Real Accountability

Herbert J. Walberg

By 1985, the concerns of policy makers and the public were increasingly shifting to achievement outcomes. *A Nation at Risk* showed American students lagged behind those in other countries. *A Nation at Risk* argued that the best jobs and industries of greatest growth required general knowledge, language mastery, and mathematical, scientific, and technical skills. It seemed obvious that voting, serving on juries, and other duties of citizenship require such knowledge and skills as well as mastery of American history, civics, and geography.

To meet the crisis of mediocrity, legislators and school boards continued to spend substantially more money on schools,¹ and educators reformed policies and practices. But ever more pointedly, legislators, citizens, and parents asked how much students were actually learning. They wanted accountability for results.

These concerns were warranted. Indeed, *A Nation at Risk* underestimated the problem because we now know that American students fall further behind the longer they remain in U.S. schools even though,

when they begin school, they are just as able as students in other countries.² Before and after *A Nation at Risk*, moreover, both per-student spending and students' mental abilities rose steadily and substantially, but students' achievement in the standard school subjects stagnated at low levels. Perhaps most worrisome, more than \$125 billion dollars of federal money spent on students in poverty failed to eliminate or even reduce the gap in achievement between them and more-advantaged students.

Failed schools have debilitating effects on the economy: An estimated 78 percent of our nation's institutions of higher learning offer remedial courses for first-year students who are unready for college work. About half of American firms provide training to make up for inadequate schooling, perhaps a considerable fraction of the estimated annual \$55 billion spent on employee training. A U.S. Department of Labor study estimated that illiteracy in one year cost eight southern states \$57.6 billion in lost productivity, substandard work, unrealized taxes, unemployment claims, and social problems.³

Accountability Milestones

A Nation at Risk encouraged national, state, and local deliberations and many policy reforms. A consequence was an effort to hold educators accountable for outcomes, particularly achievement test scores. Policy makers, businesspeople, the public, and parents increasingly insisted on having better measures and improvements of actual outcomes. As shown in table 1, accountability reforms grew slowly at first but later swelled into the crescendo of state and federal legislation we now see.

One significant accountability milestone took place in 1989, at the National Governors' Association Education Summit in which then President George H. Bush, the nation's governors, and business leaders gave impetus to business-style accountability for schools. "Systemic

Table 1. Significant accountability events

Year	Event
1983	<ul style="list-style-type: none"> • During President Reagan’s term of office, a panel of citizens issues <i>A Nation at Risk</i> that declares a “rising tide of mediocrity in education.” Influenced by poor U.S. standings on achievement tests, the panel calls for core curriculum subjects, higher learning expectations, more time in school, and better teaching. Business leaders complain of poor employee academic-skill preparation for work.
1988	<ul style="list-style-type: none"> • Twenty-five states pass legislation intended to raise education spending. • Congress creates the National Assessment Governing Board, composed of elected state officials, school board members, business leaders, scholars, and others, to develop assessments and standards for national, regional, and state comparisons of achievement in reading, mathematics, science, and other subjects.
1989	<ul style="list-style-type: none"> • President Bush calls National Education Summit of state governors to establish education goals for 2000. Arkansas governor and future president Bill Clinton plays a leading role.
1991	<ul style="list-style-type: none"> • The U.S. Department of Education funds efforts to draft national curriculum standards for core curriculum subjects. • National Assessment Governing Board releases first-ever valid state achievement comparisons. • The Governing Board also releases the first-ever percentages of students meeting the standards of Advanced, Proficient, Basic, and Below Basic levels; poor performance levels continue in subsequent years.
1994	<ul style="list-style-type: none"> • President Clinton signs Goals 2000: Educate America act, mandating creation of the National Education Standards and Improvement Council. Congress voices opposition and no one is appointed to the council.
1995	<ul style="list-style-type: none"> • National curriculum standards are released to widespread criticism from Congress and other groups.
1996	<ul style="list-style-type: none"> • The second National Education Summit of governors pledges to set standards at the state and local levels. • The Southern Regional Education Board releases report showing that states around the country have much lower standards than the National Assessment.
1998	<ul style="list-style-type: none"> • Thirty-eight states have adopted state standards in core academic subjects.
2001	<ul style="list-style-type: none"> • A report commissioned by Congress shows that approximately \$120 billion in spending and detailed regulations of state and local districts over more than 20 years have failed to reduce the gap between children in poverty and other children.

Table 1. (continued)

<i>Year</i>	<i>Event</i>
2002	<ul style="list-style-type: none"> • President Bush signs No Child Left Behind act, which is to approximately double spending and impose further regulations; states to develop challenging standards for students to meet by twelfth grade. • U.S. Department of Education reports that only nineteen states meet the 1994 federal Elementary and Secondary Act requirements. • All states but high-scoring Iowa have adopted curriculum standards in core subjects, but most are neither well measured nor enforced. • U.S. history again stumps seniors; almost 60 percent score Below Basic.

reform,” as recommended by the attendees, meant aligning the chief components of education: goals, curricula, instruction, and tests.

Like the accountability of business executives, school accountability was thought to require simultaneous centralization and decentralization: centralization of standards at the state level and decentralization of operational responsibilities to the district or school level. State policy makers were to set goals and measure progress, but, unlike in the past, they were to leave local school districts and schools to develop and execute effective practices.

State officials could then set high targets for achievement and maintain more objectivity in evaluating the results, at least more than in the past when they tried to determine both goals and means. Without this division of labor, local districts might set easy-to-reach, unmeasurable, or obfuscated goals. Concern for achievement accountability was bipartisan, and surveys show that the public strongly supported and still supports objective testing, higher standards, and greater specificity about what students should learn.⁴

Another milestone was the congressional creation of the National Assessment Governing Board to develop achievement standards. The board’s reports showed that few American students could meet the Advanced standards level roughly comparable to that of Asian and European students. Dismayingly few could even reach the Proficient

level, roughly indicating the knowledge and skills required to proceed to the next grade level. The Governing Board's comparisons of states created even greater reform momentum, since state legislators could see how education systems for which they are responsible compare with others.

In 1994, the National Governors' Conference's resolve to develop measurable national education goals built further momentum. More and more states began testing programs and began setting standards. Though per-student costs of tests were small, they provided both school accountability information and useful data on the progress each student needed to make to attain standards. States that instituted more rigorous accountability programs, moreover, made better achievement progress.⁵

The Rise of State Accountability Systems

For most of the century, nearly all schools employed a variety of commercial tests that made accountability difficult since scores on different tests cannot be readily compared. *A Nation at Risk*, other reform reports, and rising concerns changed this substantially. The most important development is that most states have initiated accountability systems and have begun aligning their tests to state learning standards. Evaluation of the evidence in 2001 suggested that tests had been aligned at least in part to learning standards for the school subjects in the following numbers of states:

English	45 states
Mathematics	43 states
Science	29 states
History and social studies	23 states

Nearly all states employed multiple-choice tests, and about two-thirds used essay tests and short-answer questions. Only two employed portfolios of student work. To create school ratings, fourteen states em-

ployed only student test scores; the others employed dropout rates, attendance, and other indicators in addition to test scores.⁶

Lack of Achievement Progress

Despite such accountability milestones and state testing programs since *A Nation at Risk*, achievement progress has been disappointing. As shown in table 2, not even one of the eight Year 2000 National Educational Goals set in 1989 has been accomplished. Some results were even worse after a decade or so of effort.

As shown in table 3, few states rose to the standards of the National Assessment Governing Board, which Congress created to set forth national standards and measure their degree of attainment. Only six of twenty-two surveyed states claimed that more than half their students were proficient by their own state standards. Only one state, Connecticut, could show that more than one-third of its students met the standards of the National Assessment Governing Board.

Contrary to the long-standing goal of the federal legislation, moreover, vast amounts of extra spending for poor children failed to break the link between poverty and achievement. As shown in table 3, relatively poorer southern states such as Tennessee, Arkansas, and Louisiana ranked much lower than northern states such as Connecticut, Massachusetts, and Oregon.

Progress in Accountability Research

Even though the nation and individual states and schools have not reaped the full potential benefits of testing and accountability, considerable progress has been made in identifying effective and efficient practices of testing and accountability for improving students' learning. The most important and promising are discussed in this section.

Table 2. Achievement of the eight Year 2000 national education goals set in 1989

<i>Goal</i>	<i>Indicator</i>	<i>Baseline</i>	<i>Update</i>	<i>Achieved?</i>
All children in America will start school ready to learn.	Percentage of 3- and 5-year-olds whose parents read to them regularly	66%	69%	No
High school graduation rates will increase to at least 90 percent.	Percentage of 18- to 24-year-olds with a high school credential	86%	85%	No
All students will leave grades 4, 8, and 12 having demonstrated competency over challenging subject matter including English, mathematics, science, foreign languages, civics and government, economics, arts, history, and geography.	Percentage of students at National Assessment of Educational Progress proficient level	12% to 40%	16% to 40%	No
The nation's teaching force will have access to programs for continuing improvement of their professional skills.	Percentage of secondary school teachers who hold an undergraduate or graduate degree in their main teaching assignment	66%	63%	No
U.S. students will be first in the world in mathematics and science achievement.	U.S. rank of first on international assessments	No	No	No
Every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy.	Percentage of adults who score at the three highest literacy levels	52%	No update	No

Table 2. (continued)

<i>Goal</i>	<i>Indicator</i>	<i>Baseline</i>	<i>Update</i>	<i>Achieved?</i>
Every school will be freed of drugs, violence, and the unauthorized presence of firearms and alcohol.	Student reports	17% to 63%	16% to 63%	No
Every school will promote parental partnerships.	Parent reports that they participated in two or more school activities per year	63%	62%	No

Source: Adapted from *National Education Goals Panel, The National Goals Report: Building a Nation of Learners 1999*, vi, 17–21.

Examinations for Accountability

Frequent testing with essay and short-answer questions and multiple-choice tests leads to higher achievement. Students prepare more regularly, and frequent tests provide more information to both teachers and students about their strengths and weaknesses. 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 upon it. Such assessments may be termed "teacher-aligned" or integrated with instruction because they correspond to the content of the lessons being taught.⁷

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, and states 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

Table 3. Percentages of students meeting state and national proficiency standards for 8th-grade mathematics

	<i>State</i>	<i>National</i>	<i>Difference</i>
Connecticut	55	34	21
Massachusetts	34	32	-2
Oregon	49	32	17
Vermont	32	32	0
Indiana	64	31	33
North Carolina	81	30	51
Maryland	50	29	21
Idaho	15	27	-12
Illinois	47	27	20
New York	40	26	14
Virginia	61	26	35
Wyoming	32	25	7
Rhode Island	20	24	-4
Texas	26	24	2
Missouri	14	22	-8
Kentucky	25	21	4
Georgia	54	19	35
Oklahoma	71	19	52
South Carolina	20	18	2
Tennessee	40	17	23
Arkansas	16	14	2
Louisiana	8	12	-4

Source: *Education Week*, February 20, 2002; Internet www.edweek.org.ew/newstory.cfm?slug=23profchrt2.h21.

identical for all students. Standardized tests can widely sample many aspects of 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 a few percentage points. Sample surveys provide information quickly, efficiently, and cheaply. So, too, can thirty to sixty multiple-choice questions tap what a student knows about a broad subject constituted by thousands of facts and ideas.

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 as much time as would be required to answer a single essay question. Multiple-choice questions afford a much larger sample of students' knowledge and skills than do long essay questions. They are also fairer to students, since their scores do not depend heavily and arbitrarily on whether they happened to have concentrated on only one narrow aspect of the subject, which happens to be on an essay examination.

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 value to the information provided by students' scores on objective tests. So, the large extra cost of essay examinations is usually unwarranted by the marginal information they may provide (except possibly, as pointed out above, when educators want to encourage and measure essay writing as separate from knowledge and skills in a subject such as history, literature, or science).⁸

Cost of Tests, Standard Setting, and Accountability

Given the positive effects of testing on learning and their uses in accountability, they are one of the most cost-effective means of improving education. Though some educators have protested the costs of accountability systems, as Caroline Hoxby pointed out, their costs are surprisingly small, representing a miniscule percentage of school budgets.⁹ The payment to commercial firms for standardized testing, standard setting, and accountability in 2000 was \$234 million, which was less than 0.1 of 1 percent of K–12 school costs; it amounts to \$5.81 per American student. For the twenty-five states with available information, these per-student testing costs run between \$1.79 and \$34.00.

These costs, moreover, will undoubtedly decline in the longer run since they were estimated as states were developing accountability systems; after development and initial revision, much of the activity can be routinized at lower costs. Few states require tests more than

once a year, but, given their positive effects and small costs, there is good reason to administer them more frequently to measure the progress of teachers, schools, districts, and states. In addition, value-added measures or gains over time (discussed in a subsequent section) are more reliable when based on more than two test administrations.

Curriculum-Based External Examinations

Tests contribute to greater achievement when they are geared toward learning standards to be mastered. Curriculum-based external examinations have the common elements of being externally composed and geared toward agreed-upon subject matter that students within a certain nation, state, or province are to learn. Usually given at the end of related courses, they have substantial positive effects on learning.¹⁰ Cornell economist John Bishop has intensively studied the effects of curriculum-based external examination on learning. He analyzed the examination effects on learning by studying surveys of the U.S. Advanced Placement program, the New York State Regents, and U.S. state and Canadian provincial systems. He also analyzed examination effects on learning in the United States in comparison with effects in Asian and European nations. He has consistently found their effects on learning to be positive.

When made publicly available, curriculum-based external evaluations allow policy makers, educators, parents, and students to assess and compare achievement standings and progress. Schools that perform poorly are pressured to make progress. When students can choose schools to attend, failing schools risk losing students and even closing.

Division of Labor and Competition

Such external examinations may benefit from the ways that other nations organize their school systems. The largest and most sophisticated international comparative analysis of national achievement, per-

formed by Ludger Woessmann of the Kiel Institute of World Economics, corroborates Bishop's and related findings.¹¹ Using data from thirty-nine countries that participated in the Third International Mathematics and Science Study, his analysis showed that the following four factors consistently promote learning:¹²

1. External, curriculum-based examinations and close, outside monitoring of achievement progress
2. School autonomy over personnel and operations
3. Teacher discretion over teaching methods
4. Competition from privately governed schools

How and why should these factors yield striking national effects? Despite variations in design, the examinations cover uniform subject matter in humanities and sciences. Since the exams are graded by educators other than the students' own teachers, students have little incentive to challenge their teachers about difficult course content and standards. Instead, students and teachers work together toward the common goal of meeting examination standards. Because of examination and course uniformity, moreover, teachers can concentrate not on what to teach but how to teach, and the students' subsequent teachers can depend on what students have been taught.

Accountability Effects

A decade ago, few states specified what students should know and be able to do, but forty-nine states now do so at least for some subjects and grade levels, and the number of states with adequate academic standards has increased. The more sustained and comprehensive the accountability system, moreover, the better states' learning progress appears. A study commissioned by the National Educational Goals Panel revealed these reasons for North Carolina and Texas making the largest gains on the National Assessment of Educational Progress:

- Grade-by-grade standards with aligned curricula and textbooks
- Expectations that all students would meet the standards
- Statewide assessments linked to the standards
- Accountability for results, with rewards and sanctions for performance
- Deregulation and increased flexibility in ways the standards can be met
- Computerized feedback systems and achievement data for continuous improvement¹³

Employing standard economic principles, legislators are designing increasingly refined accountability systems and tying incentives to test results.¹⁴ For example, states increasingly “disaggregate” test scores to be sure that various groups are well served. Texas, for example, reports separate results for boys and girls, and for whites, blacks, and Hispanics. This disaggregation is the precedent, discussed below, for the federal No Child Left Behind act. Similarly, the National Assessment of Educational Progress reports percentages of students that meet Advanced, Proficient, Basic, and Below Basic standards, which encourages improvement at all levels rather than on only a single standard that is too easy for some students, schools, and districts and too challenging for others.

Improving Accountability

Despite such promising scholarly and policy breakthroughs, achievement has yet to improve. Has accountability failed in principle or in practice? Chester Finn and Marci Kanstoroom suggest that state practice is flawed. Designated as “irresponsible” in table 4, twenty-one states have weak accountability and either weak or no standards. Only five states—Alabama, California, North Carolina, South Carolina, and

Texas—make the table 4 honor roll in having both solid standards and strong accountability.

Strong accountability means the state employs report cards and ratings of schools, rewards successful schools, has authority to reconstitute or make major changes to failing schools, and exercises such authority. Solid standards are clear, measurable, comprehensive, and rigorous. The twenty-one irresponsible states and other twenty-four states have much work to do to catch up with the five on the table 4 honor roll.

Exemplary Standards

What do good standards look like? The Massachusetts History and Social Science Curriculum Framework¹⁵ is particularly notable, not only in exemplifying the Finn-Kanstoroom criteria but in representing principled content. The framework begins with three convictions: that democracy is the worthiest form of government, that its spread cannot be taken for granted, and that its survival depends on each new generation acquiring loyalty to the American founders' vision.

In seven Guiding Principles, the framework sets forth requirements for all K–12 history and social science content: (1) emphasis on the development of political principles and institutions of Western civilization, (2) recognition of each person as an individual while developing a common American civic identity, (3) need for understanding the world outside the United States, (4) learning of social science through current events and public policy, (5) continuous study of history and social science from prekindergarten through high school, (6) integration of content, concepts, and skills in a coherent course of study, and (7) drawing upon non–social science disciplines such as fine arts, literature, and mathematics.

These principles are exemplified in the framework's ten Concepts and Skills for Grade 2. They are shown on page 320.

Table 4. States classified by quality of standards and accountability

<i>Account-ability/standards</i>	<i>Solid standards A or B</i>	<i>Mediocre standards C</i>	<i>Inferior standards D or F</i>
Strong account-ability	<i>Honor roll</i> Alabama, California, North Carolina, South Carolina, Texas	<i>Shaky foundations</i> Florida, Illinois, Indiana, Kansas, Maryland, Nevada, New York, Oklahoma, Virginia, West Virginia	<i>Trouble ahead</i> Kentucky, New Mexico
Weak account-ability	<i>Unrealized potential</i> Arizona, Massachusetts, South Dakota	<i>Going through the motions</i> Delaware, Georgia, Louisiana, Mississippi, Nebraska, New Hampshire, Ohio, Utah, Wisconsin	<i>Irresponsible states</i> Alaska, Arkansas, Colorado, Connecticut, Hawaii, Idaho, Iowa, Maine, Michigan, Minnesota, Missouri, Montana, New Jersey, North Dakota, Oregon, Pennsylvania, Rhode Island, Tennessee, Vermont, Washington, Wyoming

Source: Chester E. Finn and Marci Kanstoroom, "State Academic Standards," in *Brookings Papers on Education Policy, 2001*, ed. Diane Ravitch (Washington, D.C.: Brookings Institution Press, 2001), 51, 131–80.

Accountability Principles

Table 5 explains a dozen accountability principles that, when implemented, are likely to improve achievement. They may be applied not only by states but by local districts to school administrators and, to some extent, by school administrators to teachers and teachers to students. Several of these overlap the Finn-Kanstoroom criteria. Others emphasize the importance of timeliness, fairness, balance, the expression and disaggregation of test scores, and the need to design accountability reports for the convenience and ready understanding

History and Geography

1. Use a calendar to identify days, weeks, months, years, and seasons.
2. Use correctly words and phrases related to time (now, in the past, in the future), changing historical periods (other times, other places), and causation (because, reasons).
3. Explain the information that historical timelines convey, then put in chronological order events in the student's life (such as the year he or she was born, started school, or moved to a new neighborhood) or in the history of countries studied.
4. Describe how maps and globes depict geographical information in different ways.
5. Read globes and maps and follow narrative accounts on them.
6. Identify cardinal directions (north, east, south, and west) and apply them to maps, locations in the classroom, school, playground, and community.

Civics and Government

7. Define and give examples of some of the rights and responsibilities that students as citizens have in the school (for example, students have the right to vote in a class election but have the responsibility to follow school rules).
8. Give examples of fictional characters or real people in the school or community who were good leaders and good citizens, and explain the qualities that made them admirable (for example, honesty, dependability, modesty, trustworthiness, courage).

Economics

9. Give examples of people in the school and community who are both producers and consumers.
10. Explain what buyers and sellers are and give examples of goods and services that are bought and sold in their community.

of those held accountable as well as of those who hold them accountable.

Few states, districts, and schools follow the twelve principles. But some do,¹⁶ which suggests that others can.

Value-Added Accountability

One of the principles in table 5 deserves further discussion: value-added achievement scores, which can also be called gain or progress

Table 5. Twelve accountability principles

Independence	To avoid bias and conflicts of interest, information should be sought from sources other than staff and institutions being evaluated.
Results-focused	Though indexes of inputs and processes may be useful, the chief focus of accountability should be on the attainment of intended and measured results.
Comprehensible	Written and oral accountability reports should be readily understandable.
Timeliness	Other things being equal, accountability value is proportional to how quickly it can be reported.
Incentive-driven	Consequences, preferably prespecified, should follow good and bad results.
Objective	Quantifiable information such as examination results should be preferred over anecdotes, public hearings, and the like.
Fair	Prespecified goals and content and curriculum-based external examinations should be favored.
Value-added	Accountability should include progress or gains in learning as well as end results.
Balanced	The scoring of multiple indicators should reflect the intended range and priority of subjects, topics, and skills.
Expressive	Success and failure should be displayed in ways that are readily comprehended by those responsible for policies and decisions.
Disaggregated	Results should be reported for girls and boys, poor and not, and various ethnic and language groups of concern.
Consumer-informed	What citizens, parents, and others are concerned about and their opinions about the quality of provided services should be included in accountability indicators.

Source: Herbert J. Walberg, "Principles for Accountability Designs," in *School Accountability*, ed. Williamson Evers and Herbert J. Walberg (Stanford, Calif.: Hoover Institution Press, 2002), 155–183.

scores, in contrast to scores that reflect a student's status on only one occasion. In their simplest form, value-added scores are the gains from one test administration to another, say, over the span of one year.

Value-added scores are important for several reasons. Status scores may be deceptive, since accountability consumers may attribute them

only to recent education experiences (say, the past year) when they are also determined by earlier schooling and extramural experiences including those in early childhood and peer groups. Thus, schools serving children in poverty may make excellent value-added progress even though their status scores are poor. Similarly, status scores in rich neighborhoods may misleadingly suggest that schools are effective even though the scores may have been largely determined by social advantages.

Despite such obvious appeal, value-added scores are not without controversy. Scholars do not completely agree on how they should be calculated and employed. They may provide a somewhat unreliable indication of a school's or a teacher's progress. Even so, the perfect should not be the enemy of the good. In science and practical affairs, an approximate answer to the right question is better than a precise answer to the wrong question. In holding institutions and people accountable, it is useful to know both their status and their progress, even though neither indication is precise.¹⁷

No Child Left Behind: Will the Federal Act Work?

Table 6 shows eight provisions of the act that directly concern or relate to accountability. All the provisions are mandated, and states that do not comply risk losing federal funds.

Will the mandated requirements be well implemented? Will the requirements be implemented at all? As the last column in the table shows, many states would have to expend considerable effort to be in nominal compliance. As pointed out above, in 2001 only nineteen states had complied with 1994 federal education requirements. Such previous noncompliance may suggest that the state departments of education may not move forward quickly. Indeed, as of September 2002, at least five states had the mistaken impression that they need not meet a key requirement of the No Child Left Behind Act for

Table 6. No Child Left Behind: State provisions, requirements, and recent state status

<i>Provision</i>	<i>Man- dated</i>	<i>Conse- quences</i>	<i>Recent (2000–2002) state status</i>
State academic standards and student achievement standards in reading, mathematics, and science	Yes	Yes	All states but Iowa have reading and mathematics standards; most states have science standards.
Adequate yearly progress	Yes	Yes	At least 22 states have analytic methods; the other states were adopting methods.
Annual student testing of grades 3 through 8	Yes	Yes	Between 7 and 24 states have at least one or various combinations of reading, mathematics, and science assessments for one or more grades.
Participation in biennial NAEP in grades 4 and 8 reading and mathematics	Yes	Yes	Between 36 and 40 states have recently participated in each examination.
State report cards	Yes	Yes	Many states do not report at the state, district, and school level; 32 states report graduation rates and 8 report the number or percentage of certified teachers; the number varies of states reporting separate results by ethnicity, gender, economic disadvantage, English-language learners, disability, and migrant status.
Consequences for low-performing schools/school improvement	Yes	Yes	Between 5 and 25 states sanctioned schools, districts, or both, including required improvement plans, dis-accreditation, funding withdrawal, imposed staff dismissals or reorganization, and takeovers
School support	Yes	Yes	Between 3 and 13 states provide support to schools, districts, or both.
School recognition	Yes	Yes	Nine states reward districts for performance; 20 reward schools.

Source: Adapted from Education Commission of the States; Internet posted February 2002 as <http://www.ecs.org/ecsmain.asp?page=/search/default.asp>.

school year 2002–03: to provide tutoring and other supplemental services for students in failing schools.¹⁸

It remains to be seen, of course, whether the federal government will withhold education funds from congressional districts. And will the provisions actually raise achievement?

Whatever the fifty state departments discover about No Child Left Behind in the many complex pages of federal legislation, then promulgate as official policy to local districts and schools, teachers may continue as they please. Despite the policy crescendo of state standards, tests, and accountability since *A Nation at Risk*, a gulf remains between what teachers teach and what is called for in standards-based reform such as No Child Left Behind. In the preface to a national survey of what fourth- and eighth-grade teachers actually teach, Chester Finn calls attention to the key findings:

First, a majority of teachers in both fourth and eighth grade opt for “student-directed learning” rather than “teacher-directed learning.” . . . Second, three-quarters of teachers have embraced the college-of-education dogma that the purpose of schooling is to help youngsters “learn how to learn” rather than to acquire specific information and skills. . . . Third, not even two out of five teachers in fourth grade base their students’ grades primarily on a “single, classwide standard,” while the majority place heavier emphasis on individual children’s abilities. . . . Fourth, teachers do not seem to have terribly high expectations for their pupils when it comes to how much and how well they will end up learning. . . . Finally, and most bluntly, one-third of fourth-grade teachers and 30 percent of eighth-grade teachers do not agree that “a teacher’s role is primarily to help students learn the things that your state or community has decided students should know.”¹⁹

Finally, aside from nominal compliance, quality is at issue: As indicated in table 4, only five states recently had strong accountability and solid standards. Can the other forty-five states be expected to improve their own work while simultaneously coordinating their pol-

icies with the No Child Left Behind act? Prediction is difficult especially if it involves the future, but it is hard to be sanguine about the prospects of the new federal legislation unless powerful sanctions and incentives are employed, such as closing failing schools and introducing market competition so that parents can choose their children's schools.

Conclusion

A Nation at Risk and subsequent reform reports led to the current primacy of interest in accountability for school outcomes, particularly achievement as measured on objective examinations. The public, parents, and legislators increasingly recognized the importance of knowledge and skills for individual and national welfare, but accountability proceeded only slowly and fitfully. In the meantime, scholarship and policy analysis showed that the application of accountability policies, principles, and standards can help promote educational outcomes effectively and efficiently.

Even though all states have made serious efforts, only five or so have solid standards and strong accountability systems. Perhaps none is well prepared to meet the accountability challenges of the federal No Child Left Behind act, which in principle threatens to withhold federal funds from noncomplying states, may require extramural tutoring for students in failing schools, and may force closure of failing schools.

The Public Agenda and Business Roundtable national surveys show that the public, parents, teachers, and students strongly support accountability and see a pressing need for more rigorous standards.²⁰ President Bush and Congress agreed on strong accountability provisions in the No Child Left Behind act, and legislators in fifty states have passed bills requiring new standards and testing. And we now have much better evidence that accountability works and how it best

works. In a nation still at risk, it behooves us to design the best accountability systems we can and implement them as well as we can.

Notes

1. On spending and other resource trends, see Caroline M. Hoxby's chapter in this volume and Erik Hanushek's and my chapters in Terry Moe's, edited, *A Primer on America's Schools*, edited by Terry M. Moe (Stanford, Calif.: Hoover Institution Press, 2001), 43–68 and 69–88.
2. On outcome trends, see Paul E. Peterson's chapter in the present volume, my chapter in the Koret Task Force report, *A Primer on America's Schools*, edited by Terry M. Moe (Stanford, Calif.: Hoover Institution Press, 2001), and my *Spending More While Learning Less* (Washington, D.C.: Thomas B. Fordham Foundation, July 1998).
3. Milton Goldberg and Susan L. Traiman, "Why Business Backs Education Standards," in *Brookings Papers on Education Policy 2001*, ed. Diane Ravitch (Washington, D.C.: Brookings Institution Press, 2001), 81–90.
4. Diane Ravitch, ed., Introduction to *Brookings Papers on Education Policy 2001* (Washington, D.C.: Brookings Institution Press, 2001), 1–8.
5. On testing costs, see Caroline Hoxby's chapter in *School Accountability*; on constructive effects of accountability, see the chapters by Julian Betts and Robert Costrell, Herbert Walberg, Milton Goldberg and Susan Traiman, Chester Finn and Marci Kanstoroom, David Grissmer and Ann Flanagan, and John Bishop in the Ravitch-edited Brookings 2001 volume.
6. The brief account in this section draws on Eric A. Hanushek and Margaret E. Raymond, "Sorting Out Accountability Systems" in Williamson M. Evers and Herbert J. Walberg, editors, *School Accountability* (Stanford, Calif.: Hoover Institution Press, 2002), 75–104.
7. See Herbert J. Walberg and Jin-Shei Lai, "Meta-Analytic Effects for Policy" in Gregory J. Cizek, *Handbook of Educational Policy*. (San Diego, Calif.: Academic Press), 419–52.
8. See Williamson M. Evers and Herbert J. Walberg, editors, *Testing Student Learning, Evaluating Teaching Effectiveness* (Stanford, Calif.: Hoover Institution Press, 2003).
9. Caroline M. Hoxby, "The Cost of Accountability," in *School Accountability*, ed. Williamson M. Evers and Herbert J. Walberg (Stanford, Calif.: Hoover Institution Press, 2002), 47–74.

10. For a summary, see John H. Bishop, "The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning," *International Journal of Educational Research*, 1996, 653–752.
11. Ludger Woessmann, "Why Students in Some Countries Do Better," *Education Next*, Summer 2001, 65–74.
12. The study also indicated that the influence of teacher unions on curriculum had negative effects and that spending variations made no difference in achievement.
13. The authors also attributed the gains in the two states to the intensity and stability of business support for the reforms but not to per-pupil spending, pupil/teacher ratios, proportion of teachers with advanced degrees, and average of teacher experience. See David Grissmer and Ann Flanagan, "Searching for Indirect Evidence for the Effects of Statewide Reforms" in *Brookings Papers on Education Policy 2001*, ed. Diane Ravitch (Washington, D.C.: Brookings Institution Press, 2001), 181–229.
14. Julian R. Betts and Robert M. Costrell, "Incentives and Equity Under Standards-Based Reform," in *Brookings Papers on Education Policy 2001*, ed. Diane Ravitch (Washington, D.C.: Brookings Institution Press, 2001), 9–74.
15. Massachusetts Department of Education, *History and Social Science Curriculum Framework, Final Draft* (Malden, Mass.: Author, 2002). Internet version: http://www.doe.mass.edu/frameworks/history/hss_draft_0926.pdf. Diane Ravitch kindly called this framework to my attention.
16. See my chapter "Principles for Accountability Designs" in the Evers-Walberg volume for further explanation of these principles and examples from districts and states.
17. Averaging value-added scores over two or more years, over several school subjects, and over large numbers of students yields more trustworthy accountability indexes. Four useful papers on value-added calculations and uses appeared in *Education Next* (Summer 2002), vol. 2, no. 2, 9–23.
18. Erik W. Robelen, "States Suffer Halting Start on Tutoring," *Education Week* (online), September 25, 2002. <http://www.edweek.org/ew/ews-story.cfm?slug=04supplement.h22>.
19. Christopher Barnes, *What Do Teachers Teach? A Survey of America's Fourth- and Eighth-Grade Teachers* (New York: Center for Civic Inno-

vation at the Manhattan Institute, September 2002). See online version at http://www.manhattan-institute.org/cr_28.pdf.

20. Summarized in my "Design Principles for Accountability" in Williamson M. Evers and Herbert J. Walberg, editors, *School Accountability* (Stanford, Calif.: Hoover Institution Press, 2002).