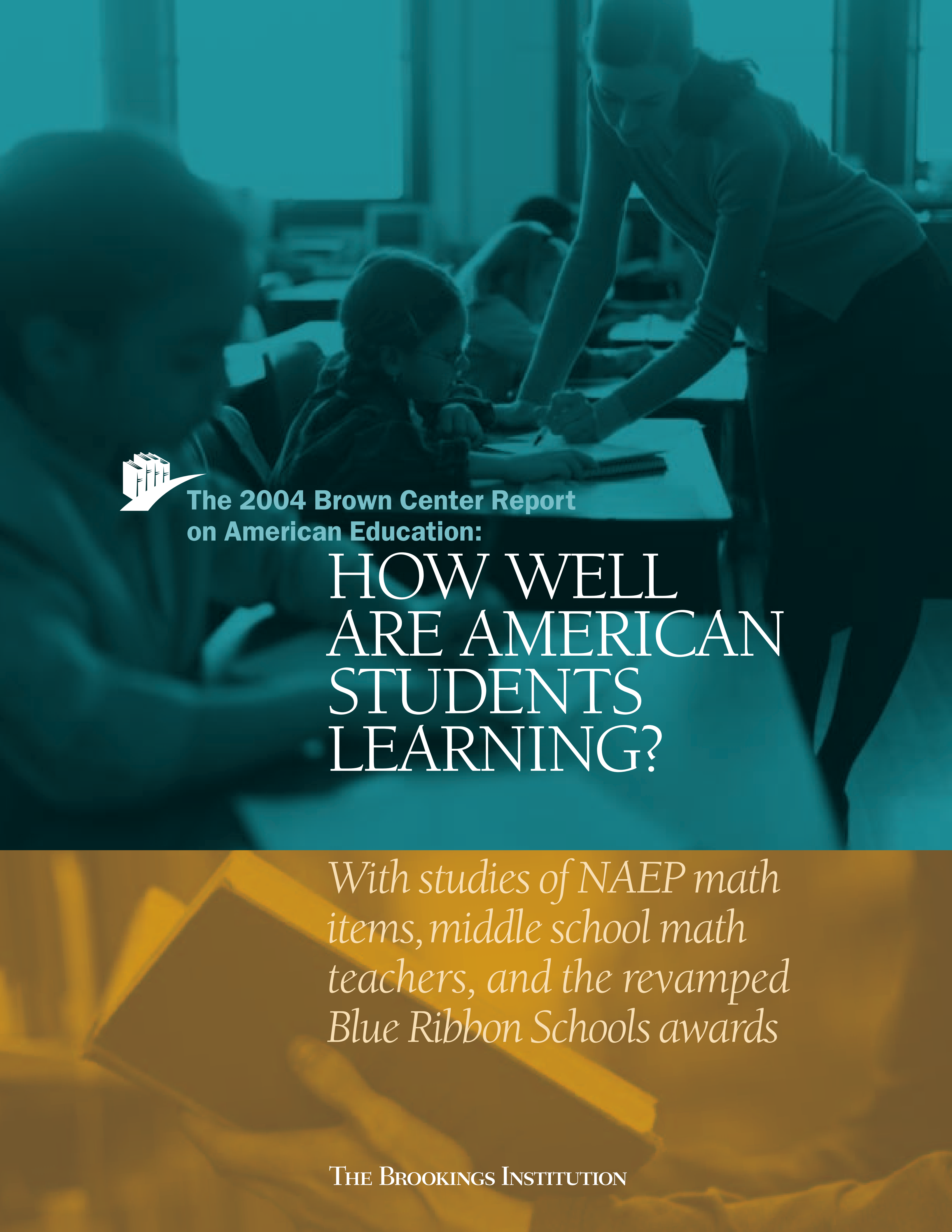


The 2004 Brown Center Report on American Education:

HOW WELL ARE AMERICAN STUDENTS LEARNING?

With studies of NAEP math items, middle school math teachers, and the revamped Blue Ribbon Schools awards

THE BROOKINGS INSTITUTION

ABOUT BROOKINGS

The Brookings Institution is a private nonprofit organization devoted to research, education, and publication on important issues of domestic and foreign policy. Its principal purpose is to bring knowledge to bear on current and emerging policy problems. The Institution maintains a position of neutrality on issues of public policy. Interpretations or conclusions in Brookings publications should be understood to be solely those of the authors.

BROWN CENTER ON EDUCATION POLICY

Established in 1992, the Brown Center on Education Policy conducts research on topics in American education, with a special focus on efforts to improve academic achievement in elementary and secondary schools. For more information, see our website, www.brookings.edu/browncenter.

To order copies of this report, please call 800-275-1447, fax 202-797-2960, e-mail BIBOOKS@brookings.edu, or visit online at www.brookings.edu.

This report was made possible by the generous financial support of The Brown Foundation, Inc., Houston.

The 2004 Brown Center Report on American Education:

HOW WELL ARE AMERICAN STUDENTS LEARNING?

With studies of NAEP math items, middle school math teachers, and the revamped Blue Ribbon Schools awards

November 2004
Volume I, Number 5

by:
TOM LOVELESS
Director, Brown Center on Education Policy

TABLE OF CONTENTS

Research assistance by:

PAUL DIPERNA
Brown Center on Education Policy

ALICE HENRIQUES
Brown Center on Education Policy

THE BROOKINGS INSTITUTION
1775 Massachusetts Avenue, NW
Washington, D.C. 20036
www.brookings.edu

THE 2004 BROWN CENTER REPORT ON AMERICAN EDUCATION

This is the fifth annual edition of the Brown Center Report on American Education. It analyzes the difficulty of items on the math portion of the National Assessment of Educational Progress (NAEP), examines the content training of middle school math teachers, and evaluates the Blue Ribbon Schools Program. The NAEP assesses student performance and periodically issues what is known as the *Nation's Report Card.* The No Child Left Behind Act requires that all teachers demonstrate mastery of the curriculum that they teach. The Blue Ribbon Schools Program singles out highly successful schools for national recognition. Measuring what students know, staffing classrooms with competent teachers, rewarding excellence—these are all activities of profound importance to the vitality of the nation's schools.

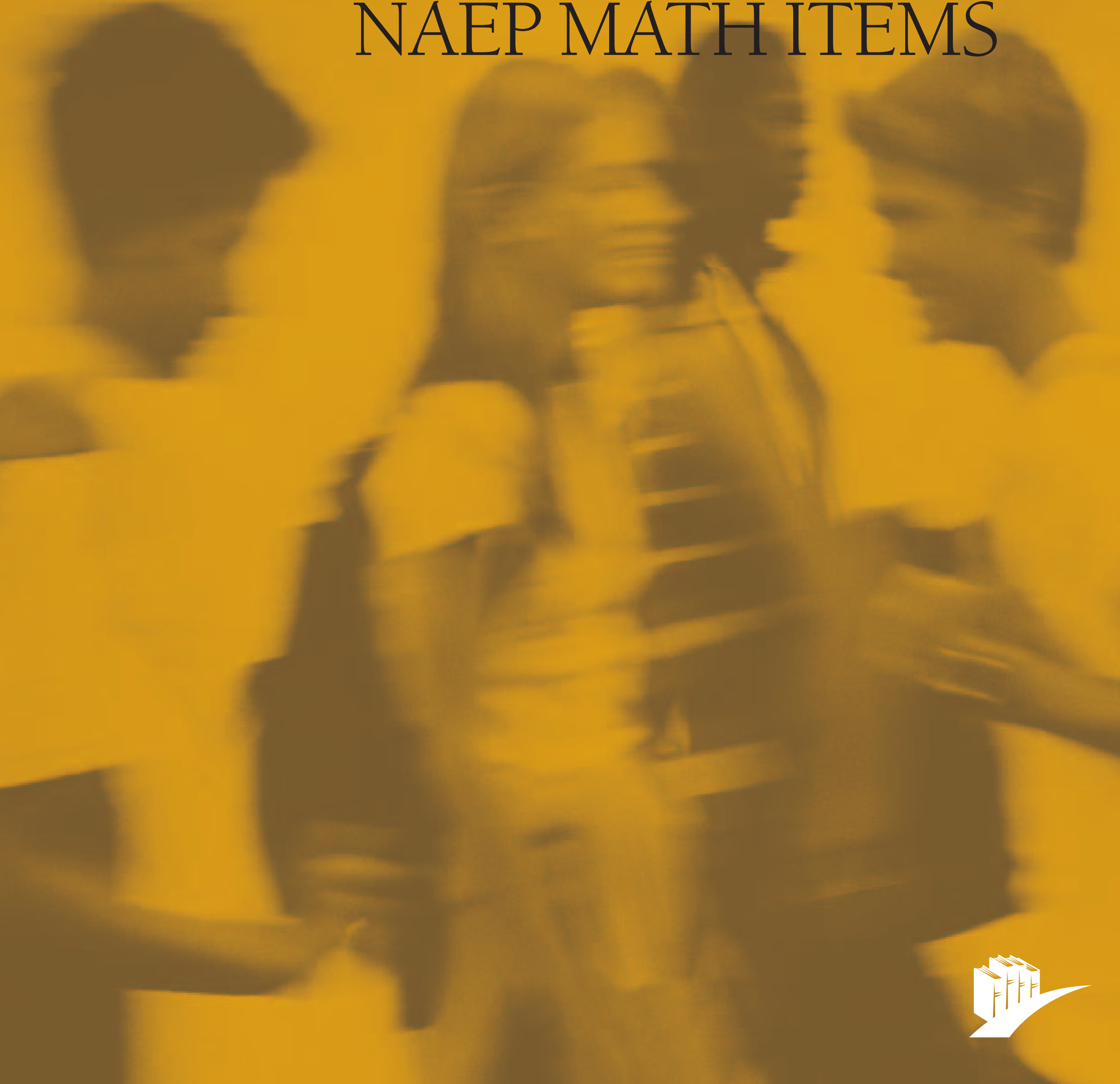
Most federal programs—including No Child Left Behind—honor federalism by leaving important powers in the hands of state and local officials. But final authority over NAEP and over Blue Ribbon Schools is held by officials in Washington, DC. They are truly federal government programs. And they are well-established. NAEP testing began in 1969. The Blue Ribbon Schools Program started in 1982. These programs have a track record.

The NAEP has publicly released more than 500 items from its mathematics tests. In the first section of this report, after reviewing test data released in 2004, we analyze a sample of NAEP items and discover that the mathematics required to solve many of the problems is extraordinarily easy. Most of the arithmetic one would need to know to solve the average item on the eighth grade NAEP is taught by the end of third grade. The second section of the report presents a survey of middle school math teachers, focusing on their educational background and professional development. A significant number of math teachers at this level lack formal undergraduate training in mathematics, and the professional development they are receiving appears to be inadequate to remedy the problem. The third section of the report examines the Blue Ribbon Schools Program. We replicate a study published in the first Brown Center Report, analyzing test data from schools that have won Blue Ribbons. In 2000, we found that at least one-fourth of the winning schools did not deserve to win the award, at least based on state test scores in reading and math. In this year's study, we find that the Blue Ribbon program today makes fewer errors. There is still room for improvement, but far fewer failing schools are receiving the awards.

A personal note. For eight years, Diane Ravitch has edited the Brookings Papers on Education Policy and organized the Brown Center's annual conference. Diane is now going to step down from those duties to spend more time on her own scholarship. She will maintain her non-resident senior fellow position at Brookings. The papers from the 2004 conference, featuring papers on eight great ideas in education, will be published early in 2005. Diane is our nation's leading education historian, and her penetrating analysis of contemporary educational problems is a national treasure. We at the Brown Center have been very fortunate to have worked with her—and look forward to working with her on future endeavors.

Part I

THE NATION'S ACHIEVEMENT/ NAEP MATH ITEMS



THIS SECTION OF THE REPORT EXAMINES THE LATEST DATA ON how well students are learning reading and mathematics. It also includes an analysis of the difficulty of mathematics items on the National Assessment of Educational Progress (NAEP).

Both federal and state test scores were released in the past year. In late 2003, the National Assessment of Educational Progress (NAEP) reported reading and math scores for fourth and eighth grades. The main NAEP has measured academic achievement since 1990. In the 2003 release, NAEP officials presented two trend lines—reflecting the progress of students with and without test accommodations. Students who are disabled or limited in English speaking ability may be offered extended test time or tested in small groups. Although the accommodations issue affects individual states' test scores, national trends appear similar under both testing conditions. As Figure 1-1 displays, reading scores remained essentially unchanged from 2002 to 2003. Scores dipped by a point at both fourth and eighth grades. Indeed, reading scores have changed very little since 1992.

Math scores are a different story (see Figure 1-2), rising strongly in 2003. Fourth graders' scores climbed 9 scale score points, from 226 to 235. Eighth graders' scores rose by 5 points. These gains continue a pattern of sharply rising math scores for more than a decade. The 2003 scores for eighth graders are 15 points higher than in 1990. Fourth graders' scores are up a whopping 22 points.

To appreciate the size of these gains in mathematics consider the 2003 scores in light of student performance in 1990. NAEP scores are calibrated so that scores from different grades can be placed on the same scale. In 1990, fourth graders scored 213 and eight graders scored 263. The midpoint of that range, 238, is a reasonable estimate of where sixth graders—the midpoint of the four grades separating fourth and eighth grades—would have scored if they had been tested in 1990. Now, the fact is sixth graders were not tested in 1990 and never have been

The 2003 scores for eighth graders are 15 points higher than in 1990. Fourth graders' scores are up a whopping 22 points.

Reading scores remain flat.

Fig
1-1

Scores at both grade levels have barely changed since 1992.

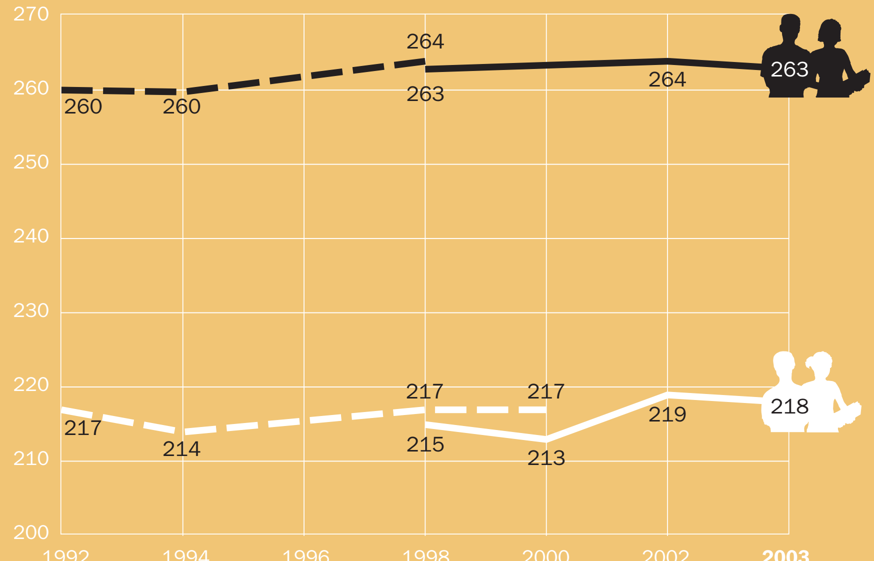
Note: Scores without accommodations were reported until 2000 for 4th grade and 1998 for 8th grade. After 2000, only scores with accommodations are reported. Accommodations are permitted for disabled and limited English-proficient students.

Source: *The Nation's Report Card*, National Center for Education Statistics, 2003.

8th Grade
4th Grade

Without Accomodations
With Accomodations

Main NAEP Reading Scores



But math scores are soaring.

Fig
1-2

From 1990 to 2003, fourth graders made an extraordinary gain of 22 points.

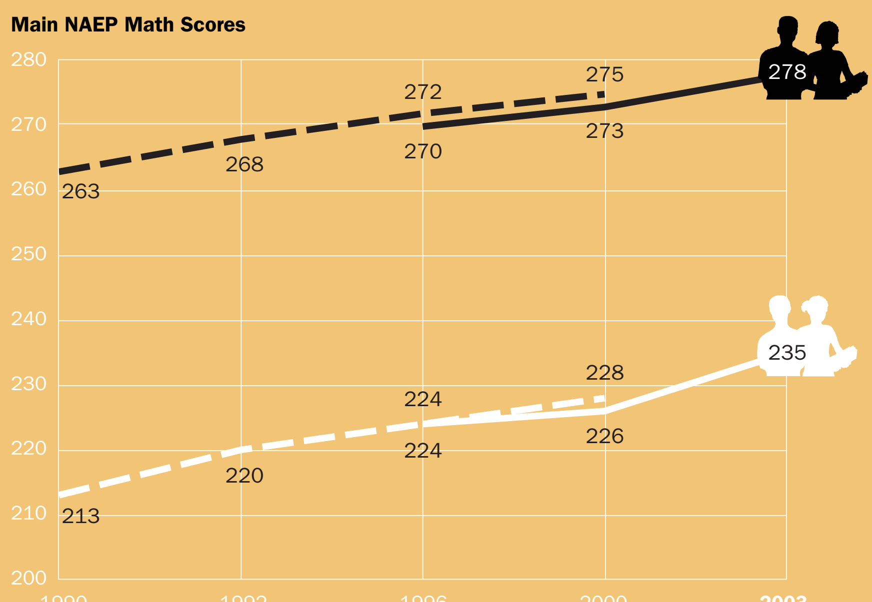
Note: Scores without accommodations were reported until 2000 for 4th grade and 1998 for 8th grade. After 2000, only scores with accommodations are reported. Accommodations are permitted for disabled and limited English-proficient students.

Source: *The Nation's Report Card*, National Center for Education Statistics, 2003.

8th Grade
4th Grade

Without Accomodations
With Accomodations

Main NAEP Math Scores

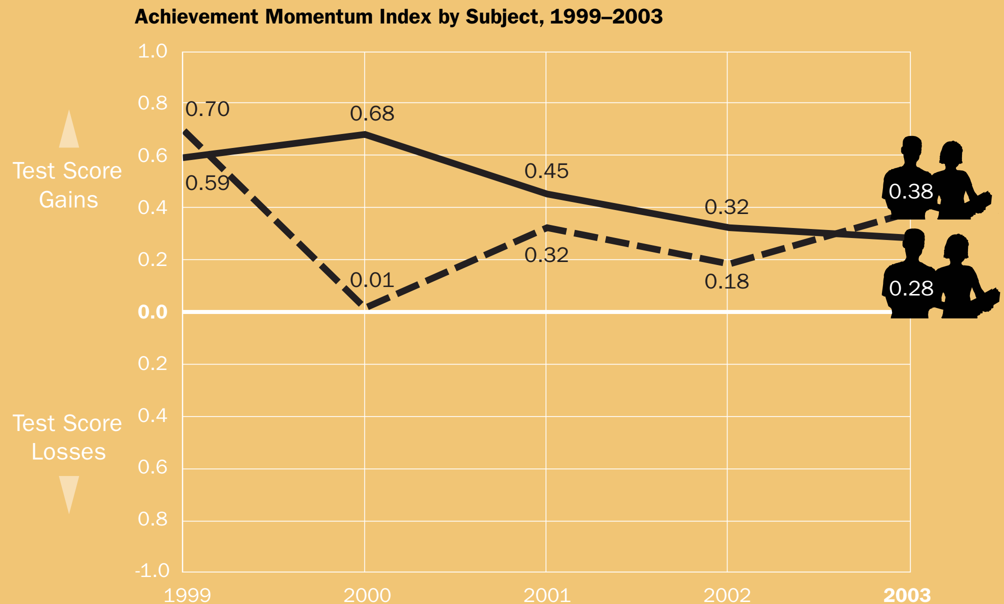


State tests show gains in reading and math.

Fig **1-3**

Reading scores ticked upward in 2003.

Source: Test scores from states that administered the same achievement test in consecutive years.

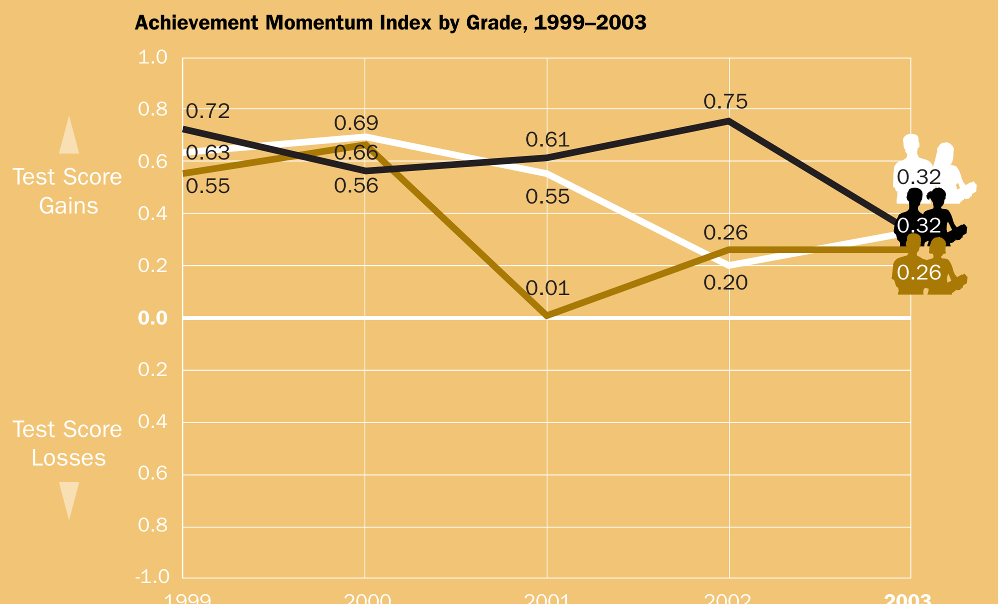


They also show gains at the 4th, 8th, and 10th grades.

Fig **1-4**

But there's a hint that 4th grade momentum may be slowing.

Source: Test scores from states that administered the same achievement test in consecutive years.



Today's fourth graders have almost caught up with 1990's sixth graders.

tested on NAEP. And it is possible that math learning is not evenly distributed in equal annual increments from fourth to eighth grades. Notwithstanding those caveats, a NAEP scale score of 238 is a good ballpark estimate of where the nation's sixth grade math performance stood in 1990.

Today's fourth graders, with a score of 235 in 2003, have almost caught up with 1990's sixth graders. That is astonishing—a two year gain. A gain that large is extraordinary on a national test, especially in a span of only fourteen years. Using first grade as the starting point for learning mathematics, it means kids today are learning in four years what it took six years to learn in 1990, a 50% gain in productivity. If similar gains are registered for the next two decades, fourth graders in 2024 will know as much mathematics as students who were beginning high school in 1990. Amazing.

Are the gains real? Almost certainly they are, but—and this is a crucial qualification—the gains probably only apply to the mathematics found on the NAEP test. What kinds of skills and knowledge are measured by NAEP? This report tackles that question below by examining the content of questions released from previous NAEP tests. The findings cast the skyrocketing NAEP gains in mathematics in a sobering light.

State tests are confirming that recent learning gains have occurred in math, but unlike NAEP, they are also showing gains in reading. The Achievement Momentum Index (AMI) combines state tests into a single national snapshot.[1] Each year, the AMI records whether individual state test scores have increased or decreased from the previous year. Scores are population-weighted and then aggregated into a national total. The weighting prevents California's test scores from being misconstrued as representing the same number of students as Alaska's. The statistic measures momentum—meaning the direction and magnitude of the overall trend—not the absolute level of learning. A statistic above 0.00 indicates scores are rising, below 0.00 indicates they are falling. A score of 1.00 indicates every state in the union is reporting score increases; a score of -1.00 indicates every state is reporting a decrease.

As displayed in Figure 1-3, states in 2003 reported gains in reading and mathematics, with both subjects coming in above 0.00 on the AMI. Momentum in mathematics appears to have slowed since 2000. State reading gains appear to have ticked up slightly in 2003, closing the gap between reading and math for the first time since 1999. Figure 1-4 breaks out the AMI by grade levels. State test scores in fourth, eighth, and tenth grades all show gains, but there is a hint that momentum in fourth grade scores may be slowing.

Let's sum up the latest data on student achievement. Federal NAEP data continue to sketch out two major trends that have been in place for several years. Students are making huge strides in mathematics. In reading, scores are basically flat. Fourth graders are registering test score gains that exceed those of students in later grades. State tests confirm the positive trend in mathematics and also indicate gains are occurring in reading.

Are NAEP's Math Gains Real? How Challenging Are NAEP Items?

The most dramatic story about NAEP in the past decade is the extraordinary gains students have made in mathematics. NAEP tests indicate that today's fourth graders know about two years more mathematics than fourth graders in 1990. The gain by eighth graders has been equally impressive. They have gained 15 scale score points, rising from 263 to 278, about half of the 31 point gap

The gain by eighth graders has been equally impressive.

An assumption of NAEP policy has been that if scores rise in the various strands that are reported on NAEP, then students are surely learning arithmetic skills as well.

that existed between eighth and twelfth graders in 1990. This suggests that the average eighth grader today knows about as much mathematics as the typical tenth grader in 1990—like the fourth graders, a two-year gain in knowledge.

To determine whether the gains are real, we evaluated the level of mathematics tested on NAEP. State tests have reported similar results as NAEP, but that may be because many states have modeled their own math assessments on the NAEP framework. In other words, the convergent trends could be due to federal and state assessments containing similar content.

How the Study Was Conducted

A total of 512 items from NAEP math tests have been publicly released and are available on the Web, along with data on how well students performed on each item.[2] The study focused on the arithmetic load of the fourth and eighth grade tests. We coded each item by the grade level in which the arithmetic skills and knowledge required by the item are taught. Items without arithmetic were not included. This is an important limitation to the study, and all conclusions concerning the difficulty of NAEP items should be interpreted accordingly. The arithmetic on the NAEP is extremely important to examine, however, because scores are not broken out and reported for how well students are learning arithmetic—even at the fourth grade. Since the advent of the main NAEP test in 1990, an assumption of NAEP policy has been that if scores rise in the various strands that are reported on NAEP, then students are surely learning arithmetic skills as well. Assessing the level of arithmetic skills demanded by the NAEP test will shed light on the validity of this assumption.

Multi-step problems were coded with the highest level of arithmetic required to answer the item. We limited the initial pool of items to the 197 classified by NAEP as involving problem solving. NAEP items are classified along three dimensions of mathematical processes—problem solving, concepts, and procedures. Analysts consider problem solving items to be among the most demanding on the NAEP test.

NAEP also classifies items by five content strands—algebra, data analysis, geometry, measurement, and number sense. The algebra content strand is also thought to be challenging. It certainly sounds impressive to hear that fourth graders are being tested on their knowledge of algebra. The number sense strand contains, but is not limited to, arithmetic skills. A second analysis was conducted examining a pool of items drawn from the algebra and number sense strands. In other words, in this comparison all arithmetic items were scrutinized, not just those emphasizing problem solving.

Sample rubric from Singapore math curriculum
(grade levels at which skills are taught)

Table **1-1**

Operations on Whole Numbers	1A (1.0)	1B (1.5)	2A (2.0)	2B (2.5)	3A (3.0)	3B (3.5)
Addition/subtraction of 2-digit numbers within 20	X					
Addition/subtraction of 2-digit numbers within 100		X				
Addition/subtraction of 3-digit numbers			X			
Addition/subtraction of 4-digit numbers					X	

Source: Primary Mathematics U.S. Edition Scope and Sequence for Singapore Math Curriculum, http://www.singaporemath.com/scope_and_sequence_USedition_PriMath.htm

When are key skills taught?
(comparing the study's rubric and state frameworks)

Table **1-2**

	Study's Rubric	California	North Carolina	Massachusetts	Florida
Whole Numbers					
Addition/subtraction of whole numbers less than 100	1	1-2	1	1-2	1-2
Addition/subtraction of 3- and 4-digit numbers	2-3	2-3	2-3	2-3	3
Multiplication times tables and corresponding division facts	2-3	2-3	3	3	4
Multiplication/division of 3-digit number by 1-digit number	3	3	3-4	3-4	4
Division of up to a 4-digit number by a 2-digit number	5	5	4	5	5
Fractions					
Halves and quarters	1	2	2	1-2	1
Comparing and ordering unlike fractions	3	4	3	3-4	3
Addition/subtraction of related fractions	4	4	4	3-4	4
Addition/subtraction of unrelated fractions	5	5	5	5	4
Product of fractions	5	5	6	5-6	5
Division of a fraction by a whole number	5	5-6	6	5-6	5
Percentages					
Concepts and computing (with) percentages	5-6	5-6	6	5-7	5-6

Source: Data compiled from respective state department of education websites; Primary Mathematics U.S. Edition Scope and Sequence for Singapore Math Curriculum, http://www.singaporemath.com/scope_and_sequence_USedition_PriMath.htm

An item's grade level was coded using a rubric based on the Singapore math textbook program, both because of the clarity of the program's scope and sequence chart and its international reputation for presenting a challenging math curriculum. Scope and sequence charts tell when topics are taught. The Singapore text divides topics between the first and second semesters at each grade, allowing two possible values per grade level. Topics taught in the first semester of the first grade, for example, were coded as 1.0; topics in the second semester as 1.5 (see Table 1-1 for sample topics). After examining a number of items, it became clear that using state frameworks as rubrics would have yielded similar results (see Table 1-2). In prominent textbook series, major arithmetic topics are typically taught within a range of one or two grade levels: addition and subtraction of whole numbers in first or second grade, multiplication and division of whole numbers in second or third grade, fractions introduced in third or fourth grade, and decimals and percentages covered in fifth or sixth grades. Even when a curriculum deviates from this sequence, topics tend to slide uniformly up or down the grade level continuum or invoke trade-offs—a topic taught earlier in one text offset by another topic taught later.

Major arithmetic topics are typically taught within a range of one or two grade levels.

The Results

Table 1-3 (see page 12) shows the grade level of arithmetic required by NAEP items and the percentage of students answering the average item in each grade cluster correctly. The first row shows that six problem solving items require knowledge of arithmetic skills taught in first grade; they comprise 15.4% of the fourth grade test items; and 49.7% of students correctly answered the average item at this level of difficulty. The percentage of students

What makes NAEP items so easy? The short answer is the dominance of whole numbers on the test.

answering correctly has a standard error of 8.2, with the statistic ranging from 25% to 78%. At the other end of the spectrum (in the final row of the table), three items assess topics typically taught in fifth grade. Only 30.7% of students were able to answer the average item at this level correctly. About one in five (20% and 21%) got the right answer on two of the problems and about half (51%) answered the remaining item correctly.

The note below the table provides summary data. The analysis was based on a total of thirty-nine items, the mean grade level of the arithmetic load was early third grade (3.1), and the percentage of students answering the average item correctly was 36.4%.

A couple of things stand out in the fourth grade portion of Table 1-3. First, the problem solving items on NAEP are not very challenging—at least not in the arithmetic required to answer them. Content taught in first and second grades is at least two years below grade level for fourth graders. But that is the level of difficulty of more than four out of ten (43.6%) problem solving items on NAEP. The second surprising finding is that even though the NAEP items are so easy, fourth graders do not do very well on them. The first and second grade items demand nothing more than being able to add and subtract whole numbers and knowing basic multiplication facts. Yet a majority of the nation's fourth graders miss the average item pitched at this level.

Even more dramatic findings are evident at eighth grade. The eighth grade items are only slightly more difficult than those on the fourth grade test (3.4 mean grade level). Almost four out of ten items (39.6%) address arithmetic skills taught in first and second grade—six years below the grade level of eighth graders taking the test. Indeed, more than three-fourths of the items (33/43) are at least four years below grade level, taught in the fourth grade or lower. Yet the percentage of eighth graders answering problem solving items correctly is an unimpressive 41.4%. Problem solving items on the eighth grade NAEP only require knowledge of very simple arithmetic. Despite this, eighth graders have trouble getting them right.

The Dominance of Whole Numbers

What makes NAEP items so easy? The short answer is the dominance of whole numbers

Problem solving items with arithmetic
(by grade level of item)

Table **1-3**

4th Grade

Grade Level	N	Students Answering Correctly (%)	Standard Error	Minimum	Maximum
1st	6 (15.4%)	49.7	8.2	25	78
2nd	11 (28.2%)	41.7	4.9	9	66
3rd	8 (20.5%)	31.6	3.4	20	47
4th	11 (28.2%)	28.8	4.4	8	58
5th	3 (7.7%)	30.7	10.2	20	51

Note: N= 39, Mean grade level: 3.1, Mean percent of students answering correct: 36.4

8th Grade

Grade Level	N	Students Answering Correctly (%)	Standard Error	Minimum	Maximum
1st	7 (16.3%)	54.0	8.4	25	92
2nd	10 (23.3%)	45.4	8.3	5	82
3rd	8 (18.6%)	41.4	7.5	4	69
4th	8 (18.6%)	32.6	8.7	6	73
5th	6 (13.9%)	38.5	6.6	17	58
6th	1 (2.3%)	NA	NA	NA	NA
7th	3 (7.0%)	27.7	9.3	13	45

Notes: N= 43, Mean grade level: 3.4, Mean percent of students answering correct: 41.4
The single sixth grade item is constructed response.

Source: Author's work and NAEP question tool http://nces.ed.gov/nationsreportcard/ITMRLS/pickone.asp

Summary of achievement and grade level of problem solving items (by content strands)

Table 1-4

4th Grade

Content Strand	N	Items with Whole Numbers Only	Average Grade Level of Items	Students Answering Correctly (%)
Algebra	5	5 (100%)	2.3	24.4
Data analysis	3	2 (66.7%)	2.5	45.3
Geometry	0	NA	NA	NA
Measurement	6	4 (66.7%)	2.8	37.3
Number sense	25	17 (48.6%)	3.4	37.5
Total	39	28 (71.8%)	3.1	36.4

8th Grade

Content Strand	N	Items with Whole Numbers Only	Average Grade Level of Items	Students Answering Correctly (%)
Algebra	8	7 (87.5%)	2.4	37.0
Data analysis	2	1 (50.0%)	4.5	33.5
Geometry	5	5 (100%)	3.2	22.6
Measurement	10	5 (50.0%)	3.7	37.4
Number sense	18	12 (66.7%)	3.6	49.4
Total	43	30 (69.8%)	3.4	40.4

Source: Analysis of NAEP items in public release.
Data on items available at http://nces.ed.gov/nationsreportcard/ITMRLS

Comparison of algebra and number sense items

Table 1-5

4th Grade

Content Strand	N	Items with Whole Numbers Only	Average Grade Level of Items	Students Answering Correctly (%)
Algebra	15	15 (100%)	2.3	40.5
Number sense	66	47 (71.2%)	3.4	48.1

8th Grade

Content Strand	N	Items with Whole Numbers Only	Average Grade Level of Items	Students Answering Correctly (%)
Algebra	25	23 (92.0%)	2.6	45.6
Number sense	54	31 (57.4%)	4.1	55.3

Source: Analysis of NAEP items in public release.
Data on items available at http://nces.ed.gov/nationsreportcard/ITMRLS

on the test. Table 1-4 examines items by content strands and shows that a majority of items do not involve arithmetic any more challenging than operations with whole numbers. This severely weakens the rigor of NAEP. Fourth graders rarely must solve problems requiring the use of fractions, decimals, or percentages. Whole numbers appear on 71.8% of items (28 out of 39). The five algebra items at fourth grade feature nothing more sophisticated than working with whole numbers. The eighth grade NAEP is barely more challenging, with 69.8% of items confined to whole number arithmetic (30 out of 43).[3] Algebra is the least challenging content strand at both grades. At the eighth grade level—where about one-fourth of the nation's students take a formal algebra course—seven out of eight algebra items only require skill with whole numbers.

How can items assessing problem solving ability in algebra demand so little in the way of arithmetic skills? It is hard to imagine a student who is competent in algebra without having some facility with fractions. To explore the algebra strand in greater depth, we broadened the pool and examined all algebra items containing arithmetic, meaning that we examined items measuring conceptual understanding and procedural knowledge as well as problem solving skills. We compared algebra items to items in the number sense strand (see Table 1-5). Algebra items are less demanding at both the fourth and eighth grades. At the eighth grade, the arithmetic demands of algebra items are pitched at only the mid-second grade level (2.6). That compares to a beginning fourth grade average for number sense items (4.1). Again, the reliance upon whole numbers is the major culprit. At the eighth grade, 92% of algebra items contain only whole number arithmetic.

At both fourth and eighth grades, calculators tend to be allowed with items that are more challenging.

False Rigor

Despite the simpler arithmetic on algebra items, fewer students answer the algebra questions correctly than the number sense questions, suggesting that some of what has been discovered here may be because of test design. It is clear that the algebra items are assessing something other than arithmetic. One assumes that the something else is algebraic—indeed it seems to be quite challenging to most eighth graders. Nevertheless, really knowing algebra means being able to solve equations that contain more sophisticated forms of numbers than whole numbers. Anything less challenging is appropriating the term "algebra" to convey a false sense of rigor to a pool of test items.

False rigor is exacerbated by the availability of calculators on the NAEP. Table 1-6 breaks out the calculator and non-calculator items in the algebra and number sense strands. At both fourth and eighth grades, calculators tend to be allowed with items that are more challenging. The eighth grade data tell the tale. The arithmetic on algebra items without calculators is pitched at about the second grade. A very low level. The items on which calculators are allowed contain arithmetic that is typically taught in the third grade, still very easy but about a year more difficult than the items solved only with pencil and paper.

False rigor, then, is incorporated into NAEP in three steps. First, a group of items is said to measure knowledge of algebra. That is quite impressive. Many people will assume that eighth graders who answer these items correctly have learned a mathematics that was once only studied by high school students preparing for college. Second, the arithmetic content of these items is restricted to the second or third grade level. Finally, students are allowed to use calculators on items involving anything more sophisticated than whole numbers.

Implications of the Study

Recall that NAEP scores offer a ballpark estimate that today's eighth graders know about as much mathematics as tenth graders in 1990. If this were true—and if the scores represent real gains in knowledge of mathematics—then a positive impact would be expected in enrollment figures for higher level math courses. That is, the courses that eighth graders are taking today should be relatively similar to the courses taken by tenth graders in 1990.

The two cohorts are not even close. There have been gains in Algebra I enrollments of eighth graders, from 16% in 1990 to 28% in 2003. Still, twice as many tenth graders had completed Algebra I in 1990 as there were eighth graders enrolled in the course in 2003. Almost half of all tenth graders (45%) had completed geometry in 1990, compared to a paltry 3% of eighth graders enrolled in geometry in 2003. Algebra II courses had been completed by 22% of tenth graders in 1990, ten times more than the 2% of eighth graders taking Algebra II in 2003.[4]

The gains of eighth graders on NAEP have not translated into the gains one would expect in course enrollments. Eighth graders today are enrolling in higher level math courses at rates far below tenth graders' completion rates in 1990. It is unlikely that educators are holding back students who are ready for Algebra II and geometry. A more reasonable explanation for the disparity—and this is admittedly speculative—can be found in the data we have just examined. Skyrocketing NAEP scores could be registering student gains in mathematical knowledge that is non-arithmetical and thus mostly irrelevant to preparation for advanced mathematics. If a solid grounding in arithmetic is prerequisite to learning advanced mathematics, a position argued by

Comparison of algebra and number sense items
(by calculator use)

Table 1-6

4th Grade

Content Strand	Calculator?	N	Items with Whole Numbers Only	Average Grade Level
Algebra	Y	8	100%	2.9
	N	7	100%	1.6
Number sense	Y	27	70.4%	3.8
	N	39	71.8%	3.1

8th Grade

Content Strand	Calculator?	N	Items with Whole Numbers Only	Average Grade Level
Algebra	Y	10	80.0%	3.4
	N	15	100%	2.1
Number sense	Y	21	42.9%	4.7
	N	33	66.7%	3.6

Source: Analysis of NAEP items in public release.
Data on items available at http://nces.ed.gov/nationsreportcard/ITMRLS

some of the world's most accomplished mathematicians, and if the arithmetic assessed on the eighth grade NAEP falls several years below grade level—a phenomenon documented here—then rising NAEP scores will not produce more students taking advanced mathematics courses.[5]

Discussion and Recommendations

This study casts the dramatic NAEP gains of the past decade in a new light. It appears that even if American students are making progress in mathematics, the arithmetic on which the gains are being registered is not very challenging, certainly no more demanding than that taught in third grade. Students know more math than in 1990, but the increase is probably far short of the two years worth of knowledge suggested by the NAEP scales. What can explain this? Are there aspects of the study that could skew the findings? Could the NAEP be more challenging than the data presented here suggest?

Students know more math than in 1990, but the increase is probably far short of the two years worth of knowledge suggested by the NAEP scales.

Yes. This study has limitations, and further study of NAEP items is sorely needed. Perhaps NAEP officials have released only the easiest items, so that the items used in the current analysis are not representative of the test as a whole. In addition, the disappointingly low levels of student performance on these items could be an artifact of test design. When standardized tests are constructed, hundreds of items can be field-tested. Items that differentiate among students—that is, those on which a significant number of students perform both well and poorly—are kept, while items that everyone aces are discarded. So the awful percentage correct figures may be exaggerating how little students know.

Other aspects of test design could skew the results. Consider that the problem solving items on NAEP are designed to isolate and measure problem solving skills. Algebra items are designed to measure algebra skills. The arithmetic load of these items may be purposely reduced, so that arithmetic does not get in the way of what the items are intended to assess. Following this line of reasoning, students who possess problem solving skills should get problem solving items right and students who have mastered algebra skills should get algebra problems right—whether they have mastered arithmetic or not.

That line of reasoning has holes in it. Not that measuring problem solving skills is a bad idea. Math reformers in the 1990s, including the National Council of Teachers of Mathematics, stressed that schools should teach mathematics as a problem solving activity.[6] The data presented here support that position by confirming widespread difficulty with problem solving. Students cannot solve problems employing even the simplest arithmetic.

Introduction to NAEP

The National Assessment of Educational Progress (NAEP) is commonly referred to as the *Nation's Report Card.* Since 1969, it has been the only nationally representative and continuing assessment of what America's students know and can do in academic subject areas.

There are two NAEP test types: (1) the main NAEP gauges national and state achievement while also reflecting current practices in curriculum and assessment, and (2) the long-term trend NAEP allows reliable measurement of change in national achievement over time. These assessments use distinct data collection procedures and separate samples of students.

Since 1990, the math test on the main NAEP has been governed by a framework reflecting recommendations of the National Council of Teachers of Mathematics (NCTM). Beginning with the 2002 assessments, the number of students selected for a NAEP national sample for any particular grade and subject has been 150,000 or more.

An exclusive focus on problem-solving, however, is shortsighted. The shortfall in students' current learning has two dimensions. One dimension is insufficient problem solving skills; the other is insufficient arithmetic skills. If students do not possess the tools to solve problems involving fractions, decimals, and percents—if students do not grasp forms of numbers other than whole numbers—then the only problems they will ever be able to solve will be mathematically trivial. The NAEP test measures problem solving skills and tracks student performance over time. It does not do so in arithmetic, and as indicated by the data presented here, assesses only the most rudimentary skills in that domain. Boosting students' competency in arithmetic and the ability to solve problems are not contradictory goals. Neither one need be denigrated in the pursuit of the other. Assessments designed to measure whether students are learning important mathematics should have items requiring students to demonstrate both a mastery of arithmetic and the ability to solve thoughtful problems.

Four Recommendations Flow from the Study:

1. Raise the level of arithmetic on NAEP problem solving items. Currently, the arithmetic skills needed for the average problem solving item on the eighth grade NAEP are taught in the third grade. Raise the level by decreasing the proportion of whole number items and increasing the proportion of items using fractions, decimals, and percentages.

2. Assess and report how students are doing on arithmetic skills. Previous studies in the Brown Center Report have documented a leveling off in students' ability to compute. This year's study makes the recommendation that we track arithmetic scores especially urgent. It is imperative

Boosting students' competency in arithmetic and the ability to solve problems are not contradictory goals. Neither one need be denigrated in the pursuit of the other.

The test can be raised to a more appropriate level by presenting fewer problems with whole number arithmetic and more with fractions, decimals, and percentages.

to measure arithmetic skills at fourth grade. Not knowing these skills places students at risk of falling far behind in mathematics. It is also important to assess arithmetic skills at eighth grade to see whether students are ready for algebra.

3. Eliminate the false rigor. Make sure algebra items assess real algebra. The following NAEP item is coded as assessing problem solving skills in algebra.[7]

Grade 8 Item Block: 1990-8M9 No. 8:

(2,5), (4,9), (6,13)

Which of the following describes what to do to the first number in each ordered pair shown above to obtain the corresponding second number?

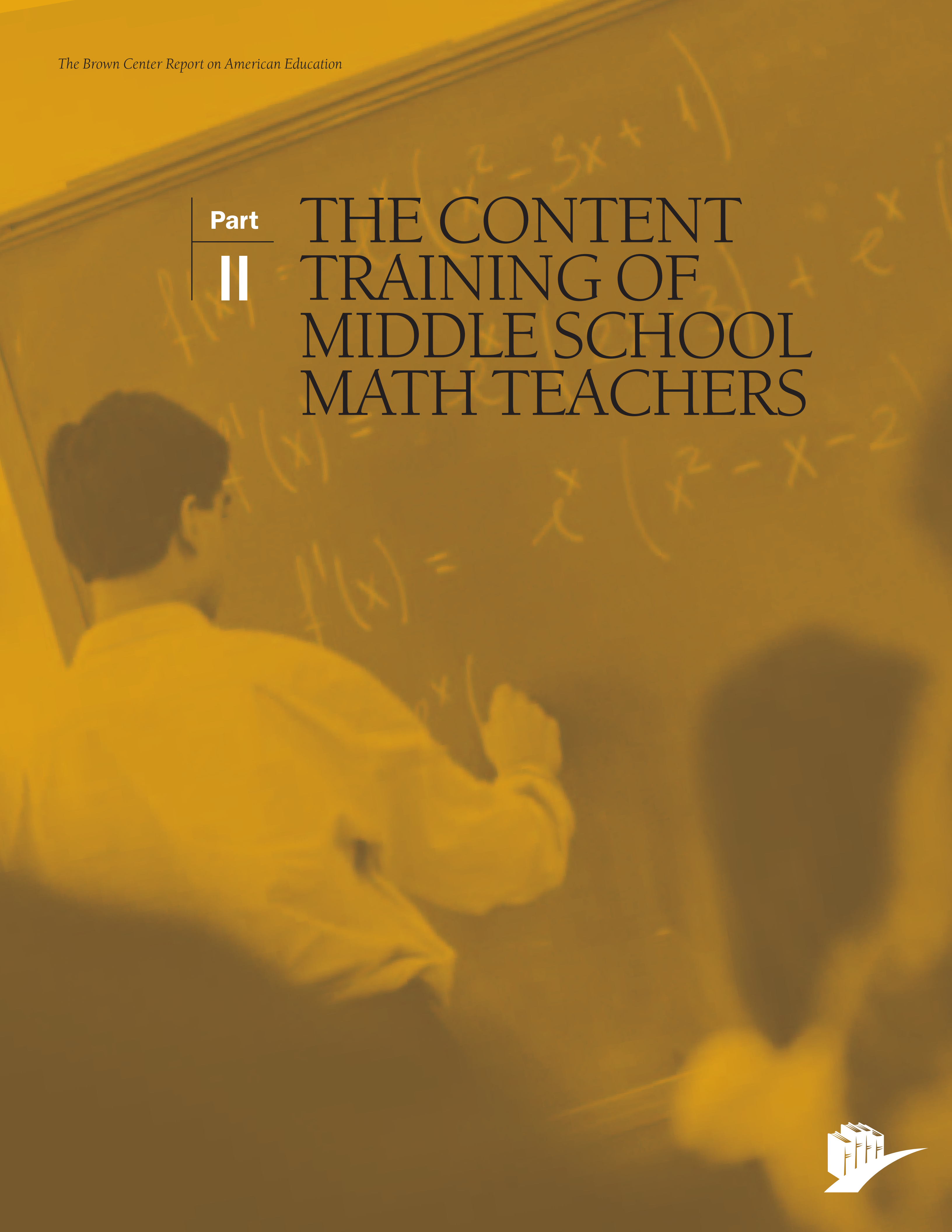
A) Add 3
B) Subtract 3
C) Multiply by 2
D) Multiply by 2 and subtract 1
E) Multiply by 2 and add 1

The correct answer is E. This item is absurdly easy and yields almost no information about a student's knowledge of algebra. The use of ordered pairs is mere window dressing, designed to make whole number arithmetic appear as algebra. Knowing what ordered pairs are or how they are used in algebra is irrelevant to arriving at a correct answer. The problem solving skills needed to answer the item are low level. A bright second grader can figure out the problem by taking the rules given in A-E, plugging in the first number of the ordered pairs, and seeing which rule produces all three of the second numbers. The arithmetic knowledge needed to solve the item is simply multiplying single-digit whole numbers by two and then adding one. To make matters worse, eighth graders are allowed to use a calculator on this item. A more appropriate algebra item would ask eighth graders to compute the slope of a line running through two or more points represented by ordered pairs. In fact, the three ordered pairs presented in this item lie on a line and could be used to pose such a real algebra problem.

4. Drop the use of calculators at fourth grade. Restrict calculator use at eighth grade. This recommendation goes hand in hand with the preceding ones. Allowing calculators interferes with determining whether students know arithmetic and can solve challenging problems. A recent Brown Center study compared the performance of nine year olds answering NAEP computation items correctly with calculators available to nine year olds answering the same items without access to calculators. The calculator group outscored the non-calculator group across the board—by as much as forty to fifty percentage points on multiplying whole numbers, a skill that all fourth graders should master. [8]

Conclusion

This study has analyzed the level of difficulty of NAEP mathematics items. In terms of arithmetic skills, they are not very challenging. Items that purport to assess problem solving skills or knowledge of algebra are particularly low level. The test can be raised to a more appropriate level by presenting fewer problems with whole number arithmetic and more with fractions, decimals, and percentages. NAEP scores have risen dramatically in mathematics, but it is unclear whether this is a significant accomplishment. The nation must have confidence that rising test scores mean that students know more real mathematics.



Part
II

THE CONTENT TRAINING OF MIDDLE SCHOOL MATH TEACHERS

THE NO CHILD LEFT BEHIND ACT (NCLB) REQUIRES STATES TO ensure that all classroom teachers are highly qualified by the end of the 2005-2006 school year. Content mastery is a key component of the requirement. Teachers in secondary schools must hold an undergraduate degree in the subject they teach or pass a rigorous exam on the subject's essential content. Most teachers in core academic subjects already meet the requirement. In mathematics, however, there is a serious problem.

According to the Secretary of Education's Second Annual Report on Teacher Quality, only 47% of the nation's secondary math teachers met the NCLB requirements in 1999-2000.[9]

The shortfall is particularly acute at the middle school level. A majority of middle school math teachers do not have a standard teaching certificate in mathematics. Nor do they have an undergraduate degree in the subject. The problem arises because middle schools hire a large number of teachers with multiple subject, elementary training and no specialized preparation in single subjects. In addition, many teachers who have received extensive training in mathematics are likely to have received it in schools of education rather than math departments. The Third International Mathematics and Science Study (TIMSS) data reveal that the U.S. is unique in this respect. Internationally, 71% of teachers of eighth grade mathematics hold bachelor's degrees in mathematics. In the U.S., only 41% hold math degrees. Approximately 54% of eighth grade math teachers earned their degrees in education.[10]

Deficiencies in content knowledge have significant implications for the professional development of math teachers. A recent study conducted by Horizon Research revealed that one-third of middle school math teachers had never taken a college algebra course and half had not taken a college geometry course.[11] To assess the current extent of this problem and to pinpoint math topics that should be the focus of professional development, we conducted a national survey of middle school math teachers in spring 2003.

Methods

A sample of 600 middle schools was randomly selected from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data, a file containing data on public schools in the United States. Middle schools were defined as all schools having grades seven and eight, with no grade beyond grade nine and no grade below grade five. The survey was administered by Direction Service, Inc., in Eugene, Oregon, and supported by a contract with the U.S. Department of Education.[12] A mail survey was prepared, addressed to the math department of each school, and mailed in three successive waves. To randomize teacher selection within schools having more than one math teacher, recipients were asked to forward the survey to the teacher whose last name is first alphabetically.

Completed surveys were received from 252 teachers, a response rate of 42%. In the analyses presented here, surveys with items left blank were excluded. Although the surveys were filled out by teachers, the findings should be considered as representative of schools, the initial unit of randomization. A random draw of all middle school teachers in the U.S. would produce a greater proportion of teachers in urban schools and in schools with large enrollments. It is not known how the findings presented here would differ under a different sampling scheme. The data can be weighted to more closely match the population of U.S. teachers. Although analyses of weighted data have not been conducted, a cursory examination suggests that any differences from the results reported here would be minor.

Like all mail surveys, this one is dependent on self-selected respondents. How selection factors may have affected the results are unknown. Teachers who are better trained, more confident, and more enthusiastic about teaching may have been more likely to return the survey than teachers who are not. The response rate of 42% is sufficient to allay fears of self-selection seriously undermining the results.

The educational backgrounds and current teaching assignments of teachers are presented in Table 2-1. The spotty formal training that most middle school teachers have received in mathematics stands out. Fewer than one-fourth (22%) of the teachers majored in math while in college (see Table 2-1), about half of the percentage reported in TIMSS.[13] The discrepancy is probably due to the fact that the TIMSS estimate applies to eighth grade math teachers, whereas the current study's estimate applies to math teachers teaching in any of the grades housed within middle schools. Schools assign their most qualified teachers to teach the most sophisticated courses, and in the case of middle schools, those courses are found in eighth grade. Less than one-half of the teachers in our survey, only 41%, hold a teaching certificate in mathematics. However, in terms of coursework, 77% of teachers report that they took four or more courses offered by mathematics departments, not schools of education, while attending college (see Table 2-2). A large majority report having received some form of pedagogical training. Only about one-fourth (24%) have not taken courses from an education school (see Table 2-3).

The data on how many teachers took four or more courses from mathematics departments offer a reasonable proxy for measuring content-related coursework. It probably is a high estimate. Many math courses sound rigorous but contain weak content, and math departments can offer courses especially for teacher

Table 2-1: Middle school math teacher backgrounds

Background	Percent
Undergraduate major in mathematics	22%
Undergraduate minor in mathematics	19%
Teaching credential or certificate in mathematics	41%

Table 2-2: Middle school math teachers were asked:

"In college, how many math courses did you complete that were offered by the mathematics department—not offered by the education school/department?"

Completed Courses	Percent
Four or more	77%
Three	7%
Two	7%
One	5%
None	4%

Table 2-3: Middle school math teachers were asked:

"In college, how many math courses did you complete that were offered by the education school/department—not offered by the mathematics department?"

Completed Courses	Percent
Four or more	17%
Three	9%
Two	26%
One	22%
None	24%

Percent of middle school mathematics teachers completing various college courses

Table 2-4

Completed Courses	Percent
General methods of teaching	92%
Methods of teaching mathematics	78%
Supervised student teaching in mathematics	47%
Mathematics for middle school teachers	45%
Instructional uses of computers/other technologies	43%
Geometry for elementary/middle school teachers	36%
College algebra/trigonometry/elementary functions	66%
Probability and statistics	56%
Geometry	47%
Computer programming or other computer science	45%
Calculus	43%
Linear algebra	28%
Number theory	27%
Applications of mathematics/problem solving	27%
Other upper division mathematics	25%
Abstract algebra	22%
Advanced calculus	21%
Differential equations	19%
History of mathematics	16%
Discrete mathematics	12%
Real analysis	11%
Engineering (any)	6%

Source: Dawayne Whittington, "Status of Middle School Mathematics Teaching," (Horizon Research, Inc., December 2002), p. 4.

Middle school math teachers were asked to describe their teaching responsibilities in a single day

Table 2-5

Responsibilities	Percent
All math courses	75%
Mostly math courses	17%
Some math courses	8%
No math courses	0%

Amount of professional development reported during the past two school years

Table 2-6

	No time	Less than 1 hour	1-2 hours	3-5 hours	More than 5 hours
Geometry	43%	15%	15%	9%	17%
Algebra	29%	10%	18%	10%	32%
Fractions and arithmetic	44%	14%	17%	9%	17%
Use of hands-on materials	23%	10%	23%	16%	28%
Integrating math topics	32%	13%	21%	12%	21%
Use of calculators	43%	12%	16%	11%	18%
Writing in math	33%	19%	23%	11%	13%
Math projects	47%	13%	20%	9%	11%
State standards	15%	9%	19%	13%	44%

candidates that emphasize pedagogical components. The Horizon Research study asked teachers to list the specific courses that they completed in college (see Table 2-4). The top two courses were "General methods of teaching" (92%) and "methods of teaching mathematics" (78%). Of courses covering formal content areas of mathematics (e.g., linear algebra), few were taken by more than 50% of teachers.[14]

Even if four math courses is the norm for middle school math teachers, it is clear that the courses do not cohere into a curriculum that systematically and comprehensively covers essential mathematical content. The courses appear to be fragmented among several topics, with some of them only tenuously related to mathematics (e.g., computer programming, history of mathematics). In addition, neither the Horizon survey nor the current study asked teachers to report how many of the math courses that they completed were remedial in nature, addressing mathematics meant to be mastered in high school.

Table 2-5 reports the teaching assignments of respondents. One-fourth of the teachers teach math courses for only part of the day. This is a little remarked-upon risk factor associated with deficiencies in content mastery. It arises from staffing practices governed by collective bargaining agreements, fluctuating enrollments at some schools, and the myriad constraints administrators face in putting a competent teacher in every classroom. Just as within schools the most qualified teachers are assigned to teach the most demanding courses, within school districts the most qualified teachers are offered full-time teaching positions. Schools with half-time positions struggle to find high-quality teachers to fill these slots. Part-time math faculty should be a source of concern in terms of ensuring that all students receive instruction from well-qualified teachers.

Professional Development Experiences

Table 2-6 shows the amount of time middle school math teachers have devoted to professional development in the past two years. The table reports the time spent by teachers who received training on nine topics, organized by whether they address content (dark-shaded rows), pedagogy (rows in white), or policy (lighter shading). A large number of teachers (44%) reported spending more than five hours learning about state standards. Of topics related to content, algebra is the most common subject of professional development, on which 32% of teachers reported more than five hours of training. Significantly fewer teachers

Teachers see their own needs paralleling the needs of colleagues.

spent that much professional development time on geometry (17%) or fractions and arithmetic (17%). The most prevalent pedagogical topics were teaching strategies related to contemporary math reform—the use of hands-on materials, integrating math topics, and the use of calculators. Instruction using math projects and writing in mathematics were less frequent topics.

These data suggest a lack of focus in professional development. Instead of concentrating on one or two key ideas, professional development covers an array of topics. Granted, some might be of interest or otherwise valuable to math teachers. Table 2-7 tallies the number of topics on which teachers reported receiving professional development in the previous two years. The average teacher received training on 5.9 topics. Only 3% reported no training on any of the nine topics. About one in six (16%) received professional development on one to three topics. A large number of teachers have received professional development on several topics—44% on four to seven topics and 36% on eight or more. The impression is that professional development strives to acquaint teachers with a broad list of current issues in middle school mathematics rather than focus on the core knowledge and skills that teachers must master.

We asked teachers to evaluate the helpfulness of professional development activities. Table 2-8 displays the results. Two pedagogical topics, one content area, and state standards stand out as the most helpful topics. Forty-one percent of teachers deemed the training they received on calculator use as very helpful. Forty percent described professional development on the use of hands-on materials in such positive terms. Professional development in algebra received similar ratings, with 41% of teachers calling it very helpful. State standards were also a favorite topic, receiving a "very helpful" rating from 39% of teachers. Professional development on math projects received the lowest marks. Twenty-nine percent of teachers described training on math projects as not helpful.

Table 2-9 displays data on teachers' assessment of their own professional development needs. Teachers express a need for training on a mix of content and pedagogical topics, a desire that underpins the smorgasbord nature of professional development today. The same pattern is evident when teachers are asked to evaluate their math colleagues' professional development needs (see Table 2-10). One reason for asking this question is to determine whether teachers assess their own needs differently from their colleagues' needs. Especially on topics such as fractions and arithmetic—basic material that teachers might be reluctant, even in a confidential survey, to admit they could use training on—the self-assessment and assessment of colleagues' needs may differ.

That is not the case here. Teachers see their own needs paralleling the needs of colleagues. The correlation of self- to colleague-need assessment is positive

Table 2-7

Number of topics on which teachers received professional development

Number of Topics	Percent of Teachers
None	3%
1-3 topics	16%
4-7 topics	44%
8 or more topics	36%

Table 2-8

Teachers were asked to evaluate professional development they've received during the past two years.

	Not helpful	Somewhat helpful	Very helpful
Geometry	14%	57%	29%
Algebra	5%	54%	41%
Fractions and arithmetic	11%	56%	33%
Use of hands-on materials	8%	52%	40%
Integrating math topics	21%	52%	27%
Use of calculators	15%	44%	41%
Writing in math	17%	54%	29%
Math projects	29%	53%	18%
State standards	12%	48%	39%

Responses from teachers who have participated in professional development

What are teachers' perceptions of their own professional development needs? How helpful to receive training on the following topics?

Table 2-9

	Not helpful	Somewhat helpful	Very helpful
Geometry	8%	48%	44%
Algebra	11%	40%	49%
Fractions and arithmetic	24%	41%	34%
Use of hands-on materials	11%	46%	43%
Integrating math topics	10%	43%	47%
Use of calculators	21%	39%	40%
Writing in math	12%	49%	39%
Math projects	12%	43%	45%
State standards	15%	47%	38%

Teachers' perceptions of their colleagues' professional development needs

Table 2-10

	No need	Some need	Great need
Geometry	13%	44%	43%
Algebra	6%	42%	52%
Fractions and arithmetic	13%	44%	43%
Use of hands-on materials	7%	38%	54%
Integrating math topics	5%	40%	55%
Use of calculators	14%	45%	41%
Writing in math	8%	39%	53%
Math projects	9%	43%	48%
State standards	8%	50%	41%

Education and teaching assignments of teachers with high, middle, and low levels of content training

Table 2-11

	Low (n=56)	Middle (n=133)	High (n=53)
4 or more education courses	23.6	18.3	9.3
Math credential	17.9	60.3	61.1
Full time	53.6	83.8	77.8

Algebra is viewed as a content area in great need of professional development by 52% of respondents.

and—statistically speaking—moderately high (r =.30 to .50). The percentage of teachers seeing no need for training in fractions and arithmetic drops from 24% when assessing one's own needs to 13% when assessing the needs of colleagues. But that's not the big story. The important finding here, again, is that every topic has an audience. As shown in Table 2-10, all nine topics are seen by at least 40% of teachers as areas in which colleagues greatly need training. Three pedagogical topics are endorsed as in great need by over half of the teachers (integrating math topics—55%; use of hands-on materials—54%; writing in math—53%). Algebra is viewed as a content area in great need of professional development by 52% of respondents.

Content Training of Teachers

The foregoing discussion examined the survey responses of all teachers. The report now turns to comparing how teachers with high, middle, and low content training view professional development. Here, the content training of teachers is assumed to be reflected in their formal mathematics education. We designated teachers who earned undergraduate degrees in mathematics as those with high content training, teachers who took at least four courses in math departments but did not major in the subject as those with middle content training, and teachers who took fewer than four math courses in college as those with low content training. The high and low groups each comprise about 25% of the teachers in the sample and the middle group about 50%. The low group represents approximately 16,000 middle school math teachers nationwide and should be the target of any serious professional development program designed to boost content knowledge.[15]

Table 2-11 displays statistics on the education and teaching assignments

Financial remuneration is seen as the best inducement for intensive professional development.

of the three groups. Teachers in the low group are far more likely to have taken education courses, especially when compared to the high group teachers. Almost one-fourth (23.6%) of the teachers with low content training took four or more education courses, compared to only 9.3% of the teachers with high content training. The high and middle groups look the same as to whether teachers currently hold math teaching credentials. But only 17.9% of teachers in the low group are certified to teach math. And only about half teach full-time (53.6%).

Table 2-12 shows the percentage of teachers who received more than five hours of professional development. All three groups received the most training on state standards. Teachers with low content training do not stand out as spending more time on the content related topics—geometry, algebra, and fractions/arithmetic. The group with high content training received significantly more training than the low group on one pedagogical topic—using calculators. The low content group received significantly more training than the high group on integrating math topics. Overall, the professional development that teachers received was only weakly related to their previous education in mathematics. Professional development activities on content and instructional topics seem to be provided to everyone, whether needed or not.

Table 2-13 displays the teachers' assessment of their own professional development needs. Teachers with low content training are more likely to see a need for training in content topics. Algebra is the number one topic on which these teachers believe professional development would be very helpful. Teachers with high content training list a pedagogical topic as the area of greatest need—math projects. Not shown here is that when teachers were asked to assess their colleagues' needs on these same topics, the responses were similar to those in Table 2-13.

Incentives

All teachers were asked to rate the effectiveness of incentives for attending summer institutes offering four-hour daily sessions for two to three weeks. Table 2-14 displays the results. Clearly, financial remuneration is seen as the best inducement for intensive professional development, either in the form of extra pay or salary schedule advancement. At least 80% of teachers described financial incentives as very effective. Responses did not vary by the education of teachers or the nature of their teaching assignments (e.g., part-time or full-time).

Conclusion

This study was conducted in response to heightened awareness of the importance of teacher quality in improving student achievement. Due in part to the teacher quality provisions of No Child Left Behind, the content knowledge of teachers is now receiving intense interest. The study surveyed middle school math teachers on their education, training, and professional development needs. Three findings are important to future professional development efforts in middle school mathematics.

1. Focus on core topics. Professional development currently lacks focus, spans several unrelated topics, and gives equal emphasis to pedagogy and content. If content mastery is the primary objective of professional development, offering training on two or three carefully defined content areas, organized into a coherent curriculum, would focus scarce resources most productively.

Two or three carefully defined content areas, organized into a coherent curriculum, would focus scarce resources.

Percentage of teachers reporting more than five hours of professional development on various topics during the past two years, by level of content knowledge

Table **2-12**

	Low	Medium	High
Geometry	18%	18%	11%
Algebra	33%	33%	30%
Fractions and arithmetic	18%	17%	9%
Use of hands-on materials	31%	27%	23%
Integrating math topics	25%	20%	17%
Use of calculators	13%	20%	19%
Writing in math	9%	14%	11%
Math projects	13%	7%	15%
State standards	42%	48%	38%

Percentage of teachers who would find professional development on the following topics "Very helpful," by level of content knowledge

Table **2-13**

	Low	Medium	High
Geometry	51%	40%	26%
Algebra	60%	47%	34%
Fractions and arithmetic	45%	33%	19%
Use of hands-on materials	47%	45%	30%
Integrating math topics	52%	46%	32%
Use of calculators	51%	38%	26%
Writing in math	44%	39%	25%
Math projects	45%	45%	36%
State standards	38%	41%	23%

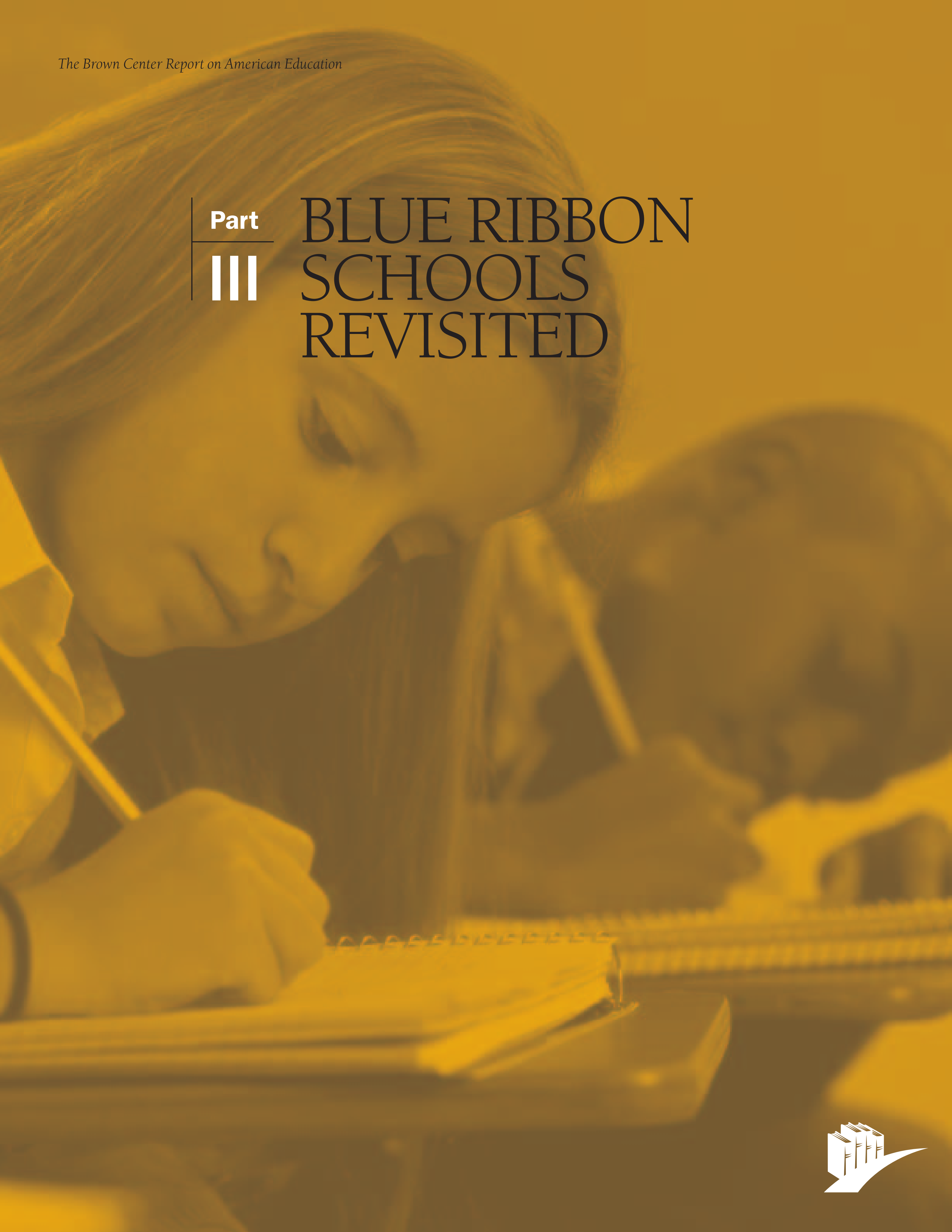
Middle school math teachers were asked: "Which of the following incentives do you think would be effective in persuading teachers to attend a summer institute?"

Table **2-14**

	Not effective	Somewhat effective	Very effective
Extra pay or stipend	0%	13%	87%
Salary schedule advancement	2%	18%	80%
Credits toward a degree	7%	36%	58%
Credit for leave or release time	22%	38%	40%

2. Target teachers in need. About one-fourth of middle school math teachers have completed fewer than four undergraduate math courses offered by a mathematics department. They are less likely than their peers to hold math teaching credentials (only 16%) or to teach full time (only about half do so). Numbering approximately 16,000 teachers, this should be the target group for intensive, content-oriented professional development. These teachers recognize their need for more extensive content training, with more than half saying professional development in algebra (60%) and geometry (51%) would be "very helpful."

3. Offer financial incentives. Four out of five teachers describe financial incentives, either stipends or salary schedule increases, as effective inducements for intensive professional development. The survey gave as an example of intensive professional development attending summer institutes of two to three weeks duration. Teachers are willing to attend such institutes as long as they are compensated for their time. Any serious, long-term effort to boost the content mastery of teachers should consider financial incentives for maximizing the commitment of participants.



Part III BLUE RIBBON SCHOOLS REVISITED

FOUR YEARS AGO, THE 2000 BROWN CENTER REPORT CONDUCTED an evaluation of a popular federal program, the Blue Ribbon Schools Program (BRSP). The program selects exemplary schools and honors them with a special flag as a symbol of national recognition. At the time of the study, nearly 4,000 schools had received the award since the program began in 1982. The study uncovered the surprising fact that several schools won the award despite mediocre scores on state tests of academic achievement.

In 2003, the federal government revamped the BRSP and re-oriented the award toward academic achievement. Now is a good time to replicate the earlier study to assess the impact of the reforms. Is academic achievement now the deciding factor in being selected as a Blue Ribbon School?

In the 2000 study, Brown Center researchers collected reading and math scores of elementary school Blue Ribbon winners and compared them to those of other schools in the same state. State test results were used. The scores were adjusted for socioeconomic status (students' family income) in order to place schools on an equal playing field. Only public schools were included in the analysis. Were the Blue Ribbon winners the highest achieving schools in their respective states? Not necessarily. Of the seventy Blue Ribbon schools in the study, seventeen scored below the 50th percentile. That means that about one-fourth of the Blue Ribbon winners had lower test scores than the average school with a comparable demographic profile.[16]

In 2001, a follow-up study was conducted, focusing on Blue Ribbon winners from California. The states nominate candidates for the award—and many have their own awards program—so examining individual state practices is important to understanding how the Blue Ribbon program works. Because California is the largest state, and each state is granted a number of nominees proportional to its population, California is allotted the most Blue Ribbon schools. On the Stanford Achievement Test (SAT-9), the test that California used at the time for monitoring school performance, the California Blue Ribbon schools looked unimpressive. In line with the findings from the larger, national

How can so many schools win a prestigious national award but not be academic stand-outs?

study, about one-fourth of California's Blue Ribbon winners scored below average. The California study also examined gains made by schools from 1998 to 2000. Perhaps the Blue Ribbon program identified schools that were only temporarily exhibiting mediocre test scores. Having adopted promising educational practices, these schools might later show dramatic gains in test scores. No, actually the average school in California made slightly greater gains than the Blue Ribbon winners.[17]

In sports and at county fairs, blue ribbons are given to recognize excellence. Why would blue ribbons be given to schools that, in terms of teaching the key subjects of reading and mathematics, appear no better than average? How can so many schools win a prestigious national award but not be academic stand-outs?

Two structural flaws plagued the old Blue Ribbon program. First, schools nominated themselves by submitting a lengthy application. This element of self-selection biased the award in favor of schools already plugged into state and federal awards networks and gave an additional edge to those with the time, energy, and know-how to toot their own horns in a forty-plus page application. Second, the selection process was geared toward emphasizing what schools were doing, not what they were accomplishing. The application presented a checklist of educational practices. The benefits of most of the activities (for example, reducing ability grouping or using hands-on instructional strategies) had never been verified by carefully conducted research. All of them were the kind of trendy, cutting-edge practices championed for decades by educational progressives. Granted, the application asked for documentation that children were learning, but academic achievement was only one of several criteria used to judge the nominees. These flaws—self-nomination and an application process favoring popular practices over academic accomplishments—proved to be fatal. They led to a large number of schools winning Blue Ribbon awards while posting mediocre results on state tests.[18]

The two studies raised a storm of criticism. Critics of tests weighed in, Alfie Kohn noting, "Genuine achievement is not reflected in standardized test scores."[19] The controversy was reawakened when reforms to the program were announced in 2002. The program was renamed the No Child Left Behind—Blue Ribbon Schools Program, tying the award to the Bush administration's major education initiative.

The first schools to win the award under the new rules were selected in 2003. The same rules continue today. Key elements of the old program remain. Schools continue to be initially screened by the states. The application form was streamlined to ten to twelve pages and narrowed to emphasize achievement. To qualify for the award, schools must score in the top 10% on state assessments or have at least 40% of their students from disadvantaged backgrounds and show dramatic improvement on state tests. Dramatic improvement is defined as schools registering the gains required for demonstrating "adequate yearly progress" (AYP) under the No Child Left Behind Act. Schools qualifying under the dramatic improvement clause must also score no lower than the 55th percentile on a state-mandated or nationally normed test.[20]

Former Blue Ribbon winners have criticized the changes. According to the *Tampa Tribune,* a school principal in Florida said that he valued his school's Blue Ribbons more than high rankings from the state's accountability system, which is based on

Achievement of 1999 Blue Ribbon Schools
(1999 test scores)

Table **3-1**

State	Number	99th-90th	89th-50th	49th-1st
		Percentile		
California	**39**	**12** (31%)	**18** (46%)	**9** (23%)
Illinois	**3**	**1** (33%)	**1** (33%)	**1** (33%)
Indiana	**4**	**0** (0%)	**2** (50%)	**2** (50%)
Michigan	**9**	**1** (11%)	**5** (56%)	**3** (33%)
Pennsylvania	**10**	**4** (40%)	**6** (60%)	**0** (0%)
New Mexico	**1**	**0** (0%)	**0** (0%)	**1** (100%)
Washington	**4**	**1** (25%)	**2** (50%)	**1** (25%)
Total	**70**	**19** (27%)	**34** (49%)	**17** (24%)

Note: Test scores from 1998-1999 school year, adjusted for socioeconomic status (SES). Public schools only. Blue Ribbon awards given in 1999.

Source: Tom Loveless. *The Brown Center Report on American Education: How Well Are American Students Learning?* (2000)

Achievement of 1999 Blue Ribbon Schools
(2003 test scores)

Table **3-2**

State	Number	99th-90th	89th-50th	49th-1st
		Percentile		
California	**38**	**2** (5%)	**30** (79%)	**6** (16%)
Illinois	**3**	**0** (0%)	**3** (100%)	**0** (0%)
Indiana	**4**	**0** (0%)	**3** (75%)	**1** (25%)
Michigan	**9**	**0** (0%)	**6** (67%)	**3** (33%)
Pennsylvania	**10**	**2** (20%)	**7** (70%)	**1** (10%)
Washington	**4**	**0** (0%)	**3** (75%)	**1** (25%)
Total	**68**	**4** (6%)	**52** (76%)	**12** (18%)

Note: Test scores from 2002-2003 school year, adjusted for socioeconomic status (SES). Public schools only. Blue Ribbon awards given in 1999.

Source: Data compiled from respective state department of education websites.

student performance on tests of academic achievement. "Although we had three [A grades] in a row, the Blue Ribbon meant a lot more to me because it was more a comprehensive assessment of the entire school," the principal commented. "I strongly believe you don't just look at test scores to see how a school is doing."[21]

Another step was added to the process when the first winners under the new rules were announced in September, 2003. South Carolina officials complained when at least one school was denied a Blue Ribbon because of special education students failing to meet AYP. Federal officials responded that schools would be allowed to appeal such disqualifications.[22]

The Current Study

This year's evaluation of Blue Ribbon schools has two aims. First, recent test scores of the 1999 winners will be examined to determine how they are doing today. Defenders of the old program claimed that simply participating in the awards process promoted school improvement. So an obvious question arises. Have the 1999 winners improved? Second, the 2003 winners' test scores will be examined to see how they compare to the 1999 winners. In other words, a comparison will be made of schools selected under the old and new systems. As with all previous evaluations, reading and math scores on state tests were collected and combined into a composite score, scores were adjusted to take into account the socioeconomic status (SES) of families served by the schools, and schools were ranked among public schools in their own states. The analysis is limited to public elementary schools.

Table 3-1 shows the test performance of 1999 Blue Ribbon winners as originally reported in the 2000 Brown Center Report. Between one-fourth and one-third (27%) were truly excellent, scoring in the top 10% of schools (90th to 99th percentiles). About half (49%) scored between the 50th and 89th percentiles—above average schools with good, but not outstanding, test scores. The remaining 24% of schools are instances when the previous program clearly misfired, giving awards to schools that had below average test scores.

How are those schools doing today? Table 3-2 displays the 2003 academic performance of the 1999 winners.[23] The percentage of outstanding schools, 90th percentile and above, has fallen to 6%. A net total of fifteen schools shifted downward into the 50th to 89th percentile category.[24] The misfires (18%) are also down from 1999, as five more schools shifted upward in ranking to the middle category, now representing 76% of the schools. Over all, the 1999 Blue Ribbon schools have not improved since the year they won the Blue Ribbons. Their test scores have fallen slightly. It is possible of course that the schools improved in ways that are not captured by achievement tests. Nevertheless, the notion that the old application process led to self-improvement is not supported by data measuring academic achievement in reading and math.

Table 3-3 presents the test scores of schools selected under the new award process. We dropped Michigan from the analysis because it had no Blue Ribbons in 2003, and New Mexico was dropped because of no available school level data. We added Missouri and Ohio because they had several Blue Ribbon schools. This batch of winners is higher achieving than the 1999 group. About one-third of the Blue Ribbon schools (31%) score in the top 10% of schools, four percentage points more than in 1999. The big difference is in the percentage of clearly undeserving schools, which have been reduced to only 9%. And all but one of them are from California. Two of the four California schools scoring below average moved up in that state's "similar schools index," a measure comparing schools with similar demographic characteristics. They were selected for having a large disadvantaged population and making significant academic progress. The other two schools score in the top 10% of schools in the state, but since they are in relatively wealthy neighborhoods, their scores appear less than sterling when compared with schools serving students of similar socioeconomic backgrounds.

Achievement of 2003 Blue Ribbon Schools
(2003 test scores)

Table **3-3**

State	Number	Percentile 99th-90th	Percentile 89th-50th	Percentile 49th-1st
California	24	**2** (8%)	**18** (75%)	**4** (17%)
Illinois	6	**1** (17%)	**4** (66%)	**1** (17%)
Indiana	3	**2** (67%)	**1** (33%)	**0** (0%)
Missouri	8	**5** (63%)	**3** (37%)	**0** (0%)
Pennsylvania	3	**1** (33%)	**2** (67%)	**0** (0%)
Ohio	10	**7** (70%)	**3** (30%)	**0** (0%)
Washington	3	**0** (0%)	**3** (100%)	**0** (0%)
Total	57	**18** (31%)	**34** (60%)	**5** (9%)

Note: Test scores from 2002-2003 school year, adjusted for socioeconomic status (SES). Public schools only. Blue Ribbon awards given in 2003.

Source: Data compiled from respective state department of education websites.

Discussion and Recommendations

The old system for selecting winners of the federal government's Blue Ribbon schools program did not work very well. The 1999 winners included a substantial number of schools with mediocre academic records. At approximately one-fourth of the Blue Ribbon schools, students test scores in math and reading were no better, and in many cases were substantially worse, than the scores of students attending schools with similar demographic profiles. The current study has found that those Blue Ribbon schools have not improved since winning the award in 1999, and in fact, their test scores have slightly declined.

A second finding is that the recent reforms of the Blue Ribbon program have succeeded in sharply diminishing the number of glaringly undeserving winners. Of the seven states examined in the 2003 cohort of winners, only California and Illinois continue to honor schools scoring below average on SES-adjusted test scores, but those schools were just barely below average with the statistical techniques we employed.

A third finding is that reform of the Blue Ribbon program has not gone far enough. Streamlining the selection process and placing a clear emphasis on academic

Blue Ribbon schools have not improved since winning the award in 1999, and in fact, their test scores have slightly declined.

Fairness dictates that all excellent schools have an equal chance of winning the award.

achievement have been beneficial for the program. However, more than half of the Blue Ribbon schools score between the 50th and 89th percentiles, a good but not great level of performance that large numbers of non-winning schools also attain. This problem may arise from states still relying on schools to nominate themselves for state award programs. The state awards often serve as the initial screen for the federal program.

California's Distinguished Schools Program offers a good example of how some of the flaws of the old program persist. Based on an initial screening of test scores, California officials invited 1881 schools to apply for its Distinguished Elementary Schools award in 2002. That is about one-third of California's public elementary schools. Applications were received from 912 schools, 247 won awards, 35 were nominated for Blue Ribbons, and 32 won Blue Ribbons.[25] At the state level, the applications were judged by a team of educational experts on a rubric embracing a number of trendy practices, including conflict resolution programs, real world experiences, collaborative community partnerships, curriculum integration, and incorporating students' developmental characteristics into school goals. These characteristics are unverified by research as benefiting student learning. Schools that know how to game this type of application—to present themselves as champions of cutting-edge practices—have a distinct advantage in California's award program. They are also more likely to self-select into the process than schools with traditional practices that, despite their old-fashioned ways, are doing an excellent job of teaching children how to read and to do mathematics.[26]

The Blue Ribbon program should stop soliciting nominees produced by state rubrics and expert panels and go out and find schools worthy of public recognition. When the program began more than two decades ago that task would have been unduly costly in both time and money. Technology makes it easier now. Achievement data are routinely collected that can be used to identify outstanding schools. Private sector organizations such as Just for the Kids, Standard and Poors, and School Match already do just that by collecting and analyzing a wealth of data on school performance. The federal government should be able to do the same thing.

Federal officials should nationally standardize the gains that qualify as "dramatic" improvement. A school that qualifies in Arizona should not be disqualified if it were in Alabama. If a common standard is not given, states should publicize the precise criteria they are using to determine dramatic improvement. All states should periodically review selection criteria to make sure that their awards programs are operating in a fair and efficient manner. Fairness dictates that all excellent schools have an equal chance of winning the award. Efficiency demands that schools that are truly exemplary—that are making extraordinary gains, teaching students more mathematics and reading than other schools—have a greater likelihood of receiving the award than less worthy schools.

The Blue Ribbon Schools Program is an opportunity for the nation to celebrate educational excellence, to honor the teachers and principals who make a positive difference in the lives of students. The program is doing a better job of recognizing exemplary schools today than it did five years ago. It is worth making it even better for the future.

ENDNOTES

1 Data for the AMI were based on the following: 28 states for 4th grade; 37 states for 8th grade; 17 states for 10th grade; 39 states for reading; 42 states for mathematics.

2 NAEP Question tool, http://nces.ed.gov/nationsreportcard/ITMRLS/pickone.asp

3 A 1997 analysis of content on the 8th grade NAEP found that 40% of items involved computation and only 13% of items involved decimals or fractions. Don McLaughlin, John Dossey, and Fran Stancavage, "Validation Studies of the Linkage Between NAEP and TIMSS Eighth Grade Mathematics Assessments," in U.S. Department of Education, National Center for Education Statistics, *Linking the National Assessment of Educational Progress (NAEP) and The Third International Mathematics and Science Study (TIMSS): A Technical Report* by Eugene Johnson and Eugene Owen, (Government Printing Office, September 1998) p. A-2.

4 NAEP Data tool, available at: http://nces.ed.gov/nationsreportcard/naepdata/getdata.asp Stephen Ingels, Barbara Schneider, Leslie Scott, Stephen Plank, and Shi-Chang W, *A Profile of the American High School Sophomore in 1990.* (U.S. Department of Education, December 1994).

5 Hung-Hsi Wu, "How to Prepare Students for Algebra." *American Educator* Vol. 25, No. 2. (Summer 2001), pp. 10-17.

6 *Principals and Standards for School Mathematics,* (National Council of Teachers of Mathematics, 2000. Available at http://standards.nctm.org/document); *Curriculum and Evaluation Standards for School Mathematics,* (National Council of Teachers of Mathematics, 1989). Available to NCTM members at: http://my.nctm.org/standards/previous/currevstds

7 NAEP Question Tool, available at http://nces.ed.gov/nationsreportcard/ITMRLS/pickone.asp

8 Tom Loveless, "Computation Skills, Calculators, and Achievement Gaps: An Analysis of NAEP Items," Presented at the 2004 American Educational Research Association conference, April 2004.

9 U.S. Department of Education, Office of Policy Planning and Innovation, "Meeting the Highly Qualified Teachers Challenge: The Secretary's Second Annual Report on Teacher Quality," (2003).

10 Ina Mulis, Michael Martin, Albert Beaton, Eugenio Gonzales, Dana Kelly, Teresa Smith, *Mathematics and Science Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Report* (The International Study Center at Boston College, 1998).

11 Dawayne Whittington, "Status of Middle School Mathematics Teaching," (Horizon Research, Inc., December 2002), p. 4.

12 The survey was conducted as part of the project "Improving Teachers' Content Knowledge in Mathematics," supported by the Fund for the Improvement of Education/Unsolicited Grants of the U.S. Department of Education.

13 Ina Mulis, Michael Martin, Albert Beaton, Eugenio Gonzales, Dana Kelly, Teresa Smith, *Mathematics and Science Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Report* (The International Study Center at Boston College, 1998).

14 Dawayne Whittington, "Status of Middle School Mathematics Teaching," (Horizon Research, Inc., December 2002), p. 4.

15 "The QED National Registry of Teacher Names," (Quality Education Data, 2003) estimates a national total of 65,113 middle school and junior high math teachers. Available at http://www.qeddata.com/teachername.asp

16 Tom Loveless, *The Brown Center Report on American Education.* (The Brookings Institution, 2000), pp. 26-31.

17 Tom Loveless and Paul DiPerna, "In Praise of Mediocrity." *Education Next,* (Summer 2001). Available at http://www.educationnext.org/20012/30.html

18 U.S. Department of Education, "1999-2000 Blue Ribbon Schools Program Middle School and Secondary School Nomination Requirements," (November 16, 1998).; U.S. Department of Education, "2000-2001 Blue Ribbon Schools Program Elementary School Nomination Requirements," (June 1, 1999).

19 *The Associated Press,* "Study: Most Blue Ribbon schools not academically excellent," September 5, 2000.

20 "2003-2004 No Child Left Behind - Blue Ribbon Schools Program," (U.S. Department of Education, September 2003).

21 Ellen Gedalius, "Emphasis Changes Award," *Tampa Tribune,* August 31, 2002, p. 1.

22 Cindy Landrum, "A Few Students Can Fail a School," *The Greenville News,* September 26, 2003, p. 112A.

23 One 1999 Blue Ribbon school in California and one in New Mexico did not report test scores in 2003.

24 In this discussion, the number of schools changing categories refers to the net change. Fifteen specific schools did not literally shift from one category to another.

25 "Eastin announces California's 2002 Distinguished Schools," available at: http://www.cde.ca.gov/nr/ne/yr02/yr02rel15.asp (April 24, 2002); "O'Connell Announces California's 2003 Distinguished Schools," available at: http://www.cde.ca.gov/nr/ne/yr03/yr03rel21.asp (April 16, 2003); "2004 Basis for Selection of Distinguished Schools," available at: http://www.cde.ca.gov/ta/sr/cs/basis04.asp; "California Department of Education Closes 2003 With a Top 10 List of Accomplishments," available at: http://www.cde.ca.gov/nr/ne/yr03/yr03rel84.asp (December 29, 2003).

26 California Department of Education, "2004 Basis for Selection of Distinguished Schools," available at: www.cde.ca.gov/ta/sr/cs/basis04.asp

Views expressed in this report are solely those of the author.

Designed by Lipman Hearne, Chicago/Washington

THE BROOKINGS INSTITUTION

1775 Massachusetts Avenue, NW • Washington, DC 20036
Tel: 202-797-6000 • Fax: 202-797-6004
www.brookings.edu

THE BROWN CENTER ON EDUCATION POLICY

Tel: 202-797-6469 • Fax: 202-797-2480
www.brookings.edu/browncenter