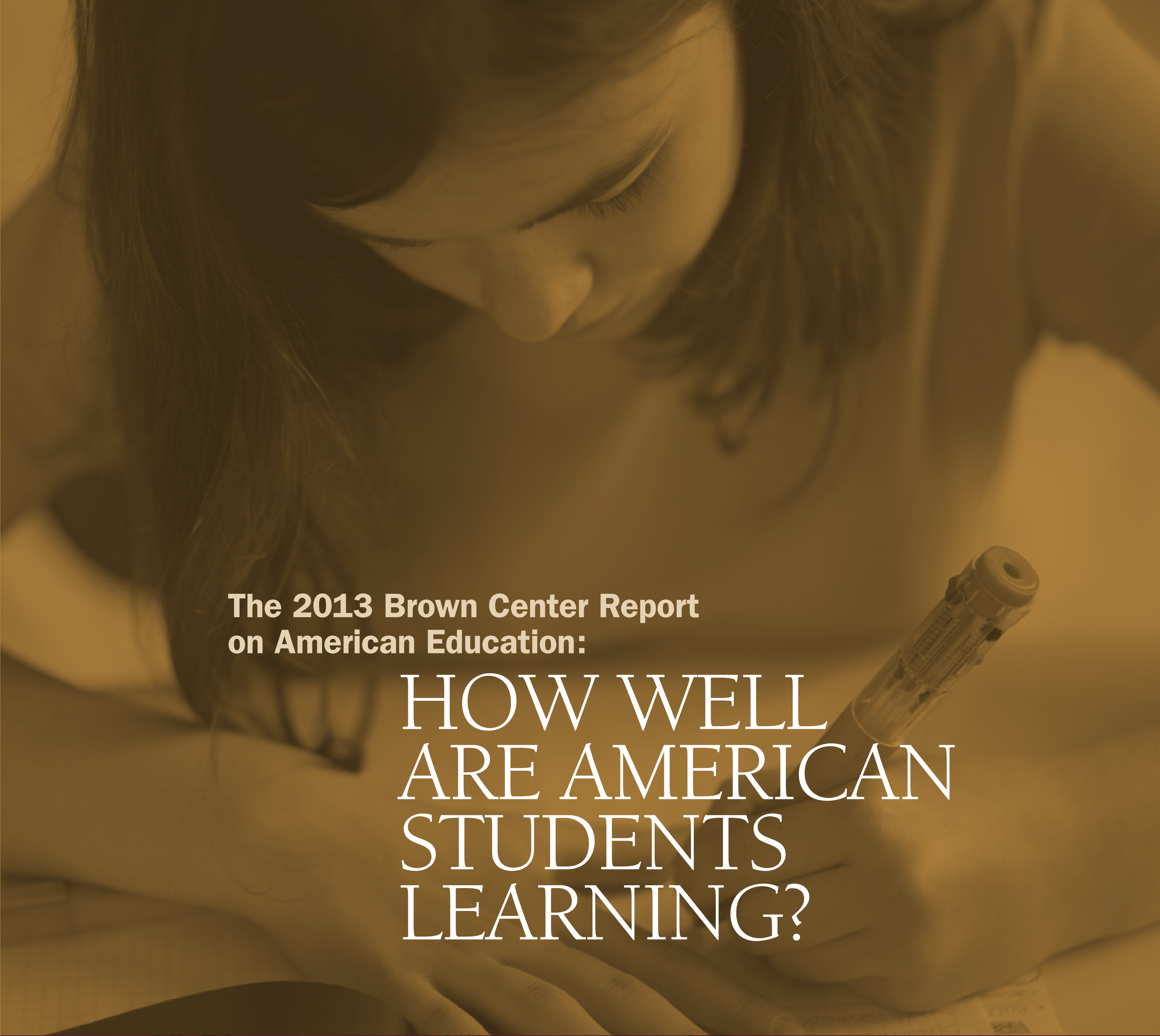


The 2013 Brown Center Report on American Education:

HOW WELL ARE AMERICAN STUDENTS LEARNING?

With sections on the latest international tests, tracking and ability grouping, and advanced math in 8th grade

BROWN CENTER on
Education Policy
at BROOKINGS

ABOUT BROOKINGS

The Brookings Institution is a private nonprofit organization devoted to independent research and innovative policy solutions. For more than 90 years, Brookings has analyzed current and emerging issues and produced new ideas that matter—for the nation and the world.

ABOUT THE BROWN CENTER ON EDUCATION POLICY

Raising the quality of education in the United States for more people is imperative for society's well-being. With that goal in mind, the purpose of the Brown Center on Education Policy at Brookings is to examine the problems of the American education system and to help delineate practical solutions. For more information, see our website, www.brookings.edu/about/centers/brown.

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

The 2013 Brown Center Report on American Education:

HOW WELL ARE AMERICAN STUDENTS LEARNING?

With sections on the latest international tests, tracking and ability grouping, and advanced math in 8th grade.

March 2013
Volume 3, Number 2

by:
TOM LOVELESS
Senior Fellow, The Brown Center on Education Policy

TABLE OF CONTENTS

Research assistance by:

Mike Gallaher

THE 2013 BROWN CENTER REPORT ON AMERICAN EDUCATION

This is the twelfth edition of the Brown Center Report. The structure of the report remains the same from year to year. Part I examines the latest data from state, national, or international assessments. This year the focus is on the latest results from the Progress in International Reading Literacy Study (PIRLS) and Trends in International Math and Science Study (TIMSS) released in December, 2012. The U.S. did relatively well, posting gains in reading, math, and science. Finland made headlines by registering declines from the last time it took the TIMSS math tests. At both fourth and eighth grades, the scores of Finland and the U.S. are now statistically indistinguishable in math. Part I also looks at the so-called "A+ countries," named that because they were the top nations on the first TIMSS, given in 1995. Part I offers "A Progress Report on the A+ Countries," and finds that, surprisingly, three of the six have registered statistically significant declines since 1995. Despite that, most of the A+ countries still score among the world's leaders. The exception is the Czech Republic, which scored at approximately the international average the last time it took TIMSS in 2007.

Part II explores a perennial theme in education studies—the topics that never seem to go away in terms of research and debate. This year it's on the controversial topics of tracking and ability grouping. An analysis of data from the National Assessment of Educational Progress (NAEP) documents a resurgence of ability grouping in fourth grade reading and mathematics. Tracking remains persistent in eighth-grade math, with about three-fourths of students in tracked classes. As readers are surely aware, both practices have been attacked for decades as inequitable, and many school analysts thought their use had diminished. Ability grouping was dominant for a long time in the elementary grades. Reading groups were the norm through most of the twentieth century and then declined dramatically in the 1990s. They are now coming back—and back strongly.

Part III is on a prominent policy or program. This year's analysis is on the national push for eighth graders to take algebra and other high school math courses. Algebra is now the single most popular math course in eighth grade. The study in Part III uses state variation in enrollment rates to ask the question: what has happened to the NAEP scores of states that boosted their eighth-grade advanced-math enrollments? The study uncovers no relationship between change in state NAEP scores and change in enrollments. States boosting advanced math taking are no more likely to show NAEP gains than other states.

A second analysis uncovers some evidence consistent with the idea that advanced math courses are being "watered down," that the mean achievement levels of advanced courses fall as enrollments go up. Again, change in NAEP score is the outcome of interest. The study shows that states that are more selective in math placements—not aggressively accelerating eighth graders into advanced courses—are more likely to show achievement gains in those courses.

There is one intriguing divergence from this finding: eighth-grade geometry classes. Geometry sits at the peak of the hierarchy of eighth-grade math courses, enrolling the nation's best math students (about 5%). Presumably, these are students who took algebra in seventh grade. Increases in eighth-grade geometry enrollments evidence no association with changes in mean achievement for the course, not what one would expect if unprepared students were being accelerated into the course. This suggests that schools are implementing two different types of acceleration, one based on the age or grade of students, the other based on students' preparation and readiness for advanced work. The analyses in the study are only correlational and cannot confirm or reject causality. Part III concludes with a discussion of hypotheses for future study to improve both strategies.

Part I

THE LATEST TIMSS AND PIRLS SCORES

IN DECEMBER 2012, THE LATEST INTERNATIONAL TEST SCORES were released. The Trends in International Math and Science Study (TIMSS) is given every four years, and the Progress in International Reading Literacy Study (PIRLS) is given every five years. The latest results came from the 2011 administration of both tests, a unique event. Because of their asynchronous schedules, the two tests share the same year only once every twenty years. Forty-nine nations and nine benchmarking participants took part in PIRLS, which is given at fourth grade, and 63 nations and 14 benchmarking participants took part in TIMSS, which is given at both fourth and eighth grades.[1]

U.S. National Achievement

The U.S. did reasonably well in all three subjects—reading, math, and science. In reading, the U.S. scored 556 on the international scale. All of the tests discussed in this section have a mean of 500 and a standard deviation of 100. Only four countries scored statistically significantly higher on the reading test. (In the discussion below, the term "significant" is used as shorthand for statistical significance at $p<.05$). Hong Kong led the world at 571, followed by the Russian Federation (568), Finland (568), and Singapore (567). The U.S. score for 2011 represented a 14-point gain since 2001 (significant).

In math, U.S. fourth graders scored 541, near the middle of second-tier countries on TIMSS. The top-tier countries were five Asian nations: Singapore (606), Korea (605), Hong Kong (602), Chinese Taipei (591), and Japan (585). The U.S. fourth-graders' score represents a 23-point gain since 1995 (significant). Eighth graders in the U.S. scored 509, which is significantly higher than the 500 international average—but just barely. The 509 score is a 17-point improvement over the 1995 U.S. score (a significant gain).

In science, U.S. fourth graders scored 544, with six countries scoring at significantly higher levels. The fourth-grade gain of 2 points since 1995 is not statistically significant. Eighth graders scored 525, significantly above the international average and significantly below students from eight other nations. The 12-point gain since 1995 is statistically significant.

To sum up, the latest international scores are mostly positive for the U.S. American students scored above the international average on all five assessments of grade-subject pairings. For four out of the five tests, the gains since 1995 are statistically significant. Despite these encouraging results, there is much room for improvement. Over the past decade, countries joining TIMSS have been economically developing nations or, in the case of the Middle East, nations possessing abundant national wealth but lacking a tradition of public schooling. Such compositional changes can make international averages easier to surpass. Leading the world in reading, math, or science remains a challenge for the U.S.

State Achievement on TIMSS

Nine states took part in the TIMSS assessment (see Table 1-1). Let's focus on eighth-grade mathematics as that is the only test on which all nine participated. As points of reference, be reminded that the international average for the test was 500, the U.S. national score was 509, and the top scoring nation was Korea at 613.

Massachusetts led the pack with a 561 score, followed by Minnesota (545) and North Carolina (537). Five of the states had taken TIMSS before, and three registered statistically significant gains from the first time they participated. As indicated in Table 1-2, the TIMSS scores map reasonably well onto NAEP scores. Because NAEP was also given in 2011, the National Center for Education Statistics was able to conduct a NAEP-TIMSS linking study.[2] Items from TIMSS and NAEP were embedded in the same booklets so that items from both tests were taken by the same student at the same time. Results of the study will be released later in 2013. The hope is that future analysts will be able to calculate, with reasonable precision, projected state TIMSS scores based on NAEP scores, allowing local leaders to place state performance in an international context.[3]

American students scored above the international average on all five assessments of grade-subject pairings.

State Scores on 2011 TIMSS, 8th-grade Math (Ranked by TIMSS 2011 Score)

Table **1-1**

State	TIMSS 2011 Score	First TIMSS Score (Year)	Change
Massachusetts	561	513 (1999)	+47*
Minnesota	545	518 (1995)	+26*
North Carolina	537	495 (1999)	+42*
Indiana	522	515 (1999)	+7
Colorado	518	–	–
Connecticut	518	512 (1999)	+5
Florida	513	–	–
California	493	–	–
Alabama	466	–	–

* $p < .05$

Note: Change statistic may not match scores due to rounding.

Source: Ina V.S. Mullis, Michael O. Martin, Pierre Foy, and Alka Arora, *TIMSS 2011 International Results in Mathematics* (Boston: TIMSS & PIRLS International Study Center, 2011).

State Scores on 2011 NAEP, 8th-grade Math (Ranked by NAEP Score)

Table **1-2**

State	NAEP Score
Massachusetts	299
Minnesota	295
Colorado	292
Connecticut	287
North Carolina	286
Indiana	285
Florida	278
California	273
Alabama	269

Note: The U.S. average score for 8th grade math was 284.

Source: "NAEP Data Explorer," National Center for Education Statistics, http://nces.ed.gov/nationsreportcard/naepdata/.

The "A+ countries" are six nations that scored at the top of the 1995 TIMSS.

Finland

Finland generated headlines from TIMSS. The "Finnish Miracle" story was called into question. In recent years, the popular press has been filled with stories about Finland's wonderful education system. Educational tourism took many observers to Finland to see schools firsthand. Tales abounded of no homework, no high stakes tests, no tardy bells, a short school day, and the national belief that requiring children to start school before age seven violates "children's right to be children."[4] Visitors marveled at the relaxed, home-like atmosphere—fireplaces in lounges, kids going shoeless, teachers called by their first names.[5] The current worldwide angst (especially evident in the U.S. and Great Britain) over achievement, productivity, and rising test scores pursued through reforms such as school choice and accountability furnishes such a stark contrast that it has even drawn a derogatory acronym—GERM—from a Finnish scholar. That stands for Global Educational Reform Movement.[6]

One problem. Finland's reputation is based largely on its performance on PISA, a very different test from TIMSS. The gap between the U.S. and Finland on PISA is statistically significant in mathematics literacy. On the 2011 TIMSS, however, Finland and the U.S. had statistically indistinguishable scores in both fourth and eighth-grade mathematics.

Look again at Table 1-1. Finland's score of 514 in eighth-grade mathematics places it near the middle of the list of states. The scores of Alabama and California are the only two states scoring statistically significantly below Finland; the scores for Colorado, Connecticut, and Florida are about the same as Finland; and four states—Massachusetts, Minnesota, North Carolina, and Indiana—scored significantly higher than Finland. If Finland had been a U.S. state in 2011, it probably would have scored in the middle of the pack on NAEP. More troubling for the Finns, their TIMSS scores have declined significantly. Finland's seventh graders took the test in 1999, scoring 520, and again in 2011, scoring 482. The 38 point decline is one of the largest recorded by a TIMSS participant.

A Progress Report on the A+ Countries

Cross-sectional data must be interpreted cautiously, and great care must be exercised when using them for predictive purposes. As Finland illustrates, a simple rule to remember is that sometimes things change.

Here is another example of that lesson, this time provided by a group of nations. The "A+ countries" are six nations that scored at the top of the 1995 TIMSS rankings in eighth-grade math. They are Belgium (Flemish community),[7] Czech Republic, Hong Kong, Japan, Korea, and Singapore. Much hoopla was made about them when the 1995 TIMSS scores were released. In 2008, they were referenced as exemplars in the Final Report of the National Mathematics Advisory Panel. William H. Schmidt, Richard T. Houang, and colleagues have published a number of studies featuring a rubric based on the A+ countries' math curriculums.[8] The idea is that other countries should be more like the A+ countries. A 2012 study by Schmidt and Houang declared the Common Core mathematics standards comparable to the A+ countries' curriculums in both focus and coherence. Moreover, they found that states with 2007 math standards similar to those of the A+ countries—again, using the same rubric from 1995—did very well on the 2007 NAEP. The findings were presented as implying that the Common Core will make the U.S. more like the A+ countries.[9]

Table 1-3 offers an update on the A+ countries. How are they doing? Let's examine the table from the bottom-up. The Czech Republic left the TIMSS study after 2007, a year that saw its TIMSS score fall by 42 points from its performance twelve years earlier. Belgium (Flemish) has not participated in TIMSS since 2003. Its performance on TIMSS declined by 13 points before it left the study. The other four countries all took TIMSS in 2011. Hong Kong (+17) and Korea (+32) registered significant gains, Japan a significant decline (-11), and Singapore showed no significant change (+2). Of the six nations, then, two had statistically significant gains, three had statistically significant losses, and one scored about the same. The average score change for the six nations is -2.5 points, approximately equal to the average change for the 20 nations that participated in both 1995 and 2011. Put another way, the average A+ country made no more progress in math achievement than any other country in TIMSS.

Giving letter grades to entire nations may seem silly to many people but since the A+ designations have attained such widespread acceptance, readers are asked for their tolerance. It's clear that A+ is no longer the appropriate grade for all of these countries.[10] Korea and Hong Kong added to their outstanding 1995 scores and still deserve an A+. Singapore, too, although not making significant gains, surely preserves its A+ status by being one of only three nations with a 600+ scale score. Then things get dicey. Flemish Belgium was slipping when it left TIMSS in 2003. Its fourth graders did participate in 2011, however, and did well, scoring 549. That's significantly higher than the U.S. at 541 and about the same as Florida at 545. But it represents no progress from the Belgian fourth graders' previous TIMSS scores. Call Flemish Belgium a question mark—maybe an A- or B+, but definitely not an A+. We don't know for sure without more recent eighth-grade data.

Table 1-3

TIMSS 8th-grade Math Scores for "A+" Countries, 1995–2011 (Ranked by TIMSS 1995 Score)

Nation	TIMSS 1995	Last TIMSS (Year)	Change
Singapore	609	611 (2011)	+2
Japan	581	570 (2011)	-11*
Korea	581	613 (2011)	+32*
Hong Kong	569	586 (2011)	+17*
Belgium (Flemish)	550	537 (2003)	-13*
Czech Republic	546	504 (2007)	-42*

* $p < .05$

Note: The U.S. had a score of 492 in 1995 and 509 in 2011, a change of +17*

Source: Mulles et al., *TIMSS 2011 International Results in Mathematics.*

Japan's score of 570 warrants an A, not an A+, and the downward trend is notable. Compare Korea with Japan. They both scored 581 in 1995. In 2011, Korea scored 43 points higher. The decline in the Czech Republic's scores is the most dramatic, 42 points. The 2007 score of 504 is statistically indistinguishable from the international average of 500. Like Flemish Belgium, the Czech Republic fourth graders did participate in TIMSS 2011, scoring 511, a 30-point decline from 1995. The Czech Republic gets a C+ or B-.

Conclusion

What should we make of this? In 1995, six high achieving nations were described as "A+" to spur the U.S. towards greater math achievement. Their math curriculums were held up as ideals. And yet, since 1995, the U.S. gain of 17 points in eighth-grade mathematics is only exceeded by one A+ nation, Korea, and matched by another, Hong Kong. The other four A+ countries made less progress than the U.S. So in terms of gains, the U.S. should not look to the A+

The average A+ country made no more progress in math achievement than any other country in TIMSS.

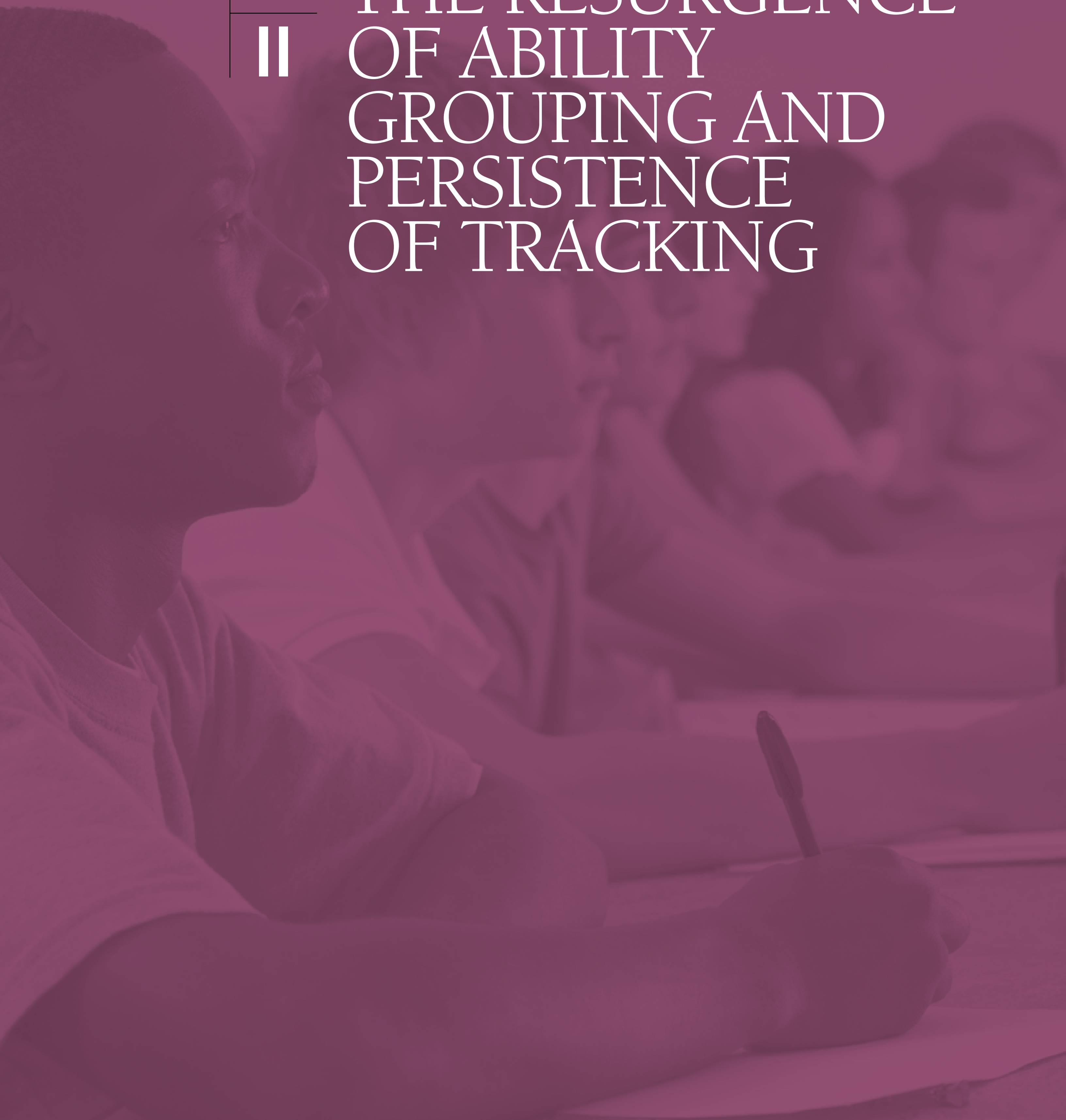
In terms of gains, the U.S. should not look to the A+ countries for guidance.

countries for guidance. That said, five of the six A+ countries continue to lead the world in eighth-grade math achievement, and they continue to score significantly higher than the U.S.

The divergence of gain scores and status scores illustrates a problem that will be addressed in both remaining parts of this report. The tendency is for observers, when test scores are released, to zero in on the top performers, to ask what it is that the leading nations are doing, and then to urge the rest of the world to do those things. That response is understandable—but it is also potentially misleading. Causality is difficult to determine from cross-sectional data. Curriculum undoubtedly plays a role, but much more work needs to be done identifying potential curriculum effects in international data and testing well-formulated hypotheses with longitudinal models. Ideally, randomized trials would be conducted on the best curriculum programs, to tease out unobserved influences on learning. Those influences include a culture that places great value on academic success, parenting practices that promote achievement, and peers who award status based on working hard at school. They surely play a part in why some nations are "A+" while others only aspire to be.

Part II

THE RESURGENCE OF ABILITY GROUPING AND PERSISTENCE OF TRACKING



THIS STUDY EXAMINES THE USE OF ABILITY GROUPING AND tracking in America's schools. Recent NAEP data reveal a resurgence of ability grouping in fourth grade and the persistent popularity of tracking in eighth-grade mathematics. These trends are surprising considering the vehement opposition of powerful organizations to both practices. Although the current study will not delve into the debate—it is interested in what schools are doing, not why or whether they should do it—discussion is offered at the end of the article on implications of the findings for the controversy surrounding the topic.

Ability grouping and tracking are often confused. They both attempt to match students with curriculum based on students' ability or prior performance, but the two practices differ in several respects. Tracking takes place between classes, ability grouping within classes. Tracking primarily occurs in high school and sometimes in middle school. In tracked academic subjects, students are assigned to different classrooms, receive instruction from different teachers, and study a different curriculum. The names of high school courses signal curricular differences. Advanced math students in tenth grade, for example, may take Algebra II while others take Geometry, Algebra I, or Pre-Algebra. Advanced tenth graders in English language arts (ELA) may attend a class called "Honors English" while other students attend "English 10" or "Reading 10." Excellent science students may take "AP Chemistry" while others take a course simply called "Chemistry" or "General Science." History may also be tracked, as when Advanced Placement courses are offered in U.S. or European history that not all students take. Some middle and high schools do not track at all, creating instead classes that are heterogeneous in ability. Students of all abilities study the same material.

What Tracking is Not

Perhaps the best way to clarify what tracking is, because of widespread misconceptions, is by describing what it is not. Tracking is decided subject by subject. Students are not assigned to college preparatory or vocational tracks that then dictate coursework all through high school; that practice died out in the U.S. in the late 1960s and early 1970s.[11,12] European and Asian school systems still practice a form of this type of tracking (they call it "streaming"), typically in the final two or three years of secondary schooling.[13] Students take placement exams and based on the scores are selected into separate schools with markedly different post-secondary destinations rather than attending different classes at the same school.[14] Exam-based selection into high schools was common in the U.S. in the 19th century and the early part of the 20th century, but fell to the wayside. The comprehensive high school—with all students of a particular community attending the same school and then divided into distinct tracks within the school—came to be enshrined as the American model.

Ability Grouping

Ability grouping typically is an elementary school practice. Most elementary classes feature a single teacher with a classroom of students who are heterogeneous in ability. To create more homogeneity, teachers may divide students into small instructional groups reflecting different levels of ability, most often for reading in the primary grades (K–3) and perhaps for reading or math in later grades (4–6).[15] While the teacher provides instruction to one group, the other students work independently—engaged in cooperative group activities or computer instruction or completing worksheets to reinforce skills. The teacher rotates among the groups so that each student receives a dose of teacher-led instruction in these small settings.

Researchers from Johns Hopkins conducted a comprehensive survey of ability grouping and tracking in 1986. The study analyzed national data augmented by an in depth survey of Pennsylvania schools. Several interesting patterns were uncovered that still hold true today. Disaggregating the data by grade level revealed that ability grouping is most prominent in first grade and then slowly recedes over subsequent grades. Ability grouping and tracking are inversely related; the school system's strategies for creating groups that are as homogeneous as possible shift over the K-12 grade span. Tracking is rare in the elementary grades and, after increasing dramatically in middle school (in mathematics, in particular) peaks towards the end of high school. It is rare for students, once grouped between classes by tracking, to be grouped again within classes by ability grouping.[16]

Because the groupings are within-class (and often decided by a single teacher), ability grouping is more flexible than tracking. Groups may be reshuffled periodically to reflect changes in student performance. Ability groups might study from different levels of the same textbook series or use the same book and move at a different pace (with enrichment activities for the faster groups until the others catch up). Instead of the formality of transcript designations for high school courses, ability groups often take the names of animals—redbirds, bluebirds, sharks, dolphins, and the like—or the names of the books in the reading series that the students are using.

The most popular alternatives to ability-grouped instruction are whole class instruction, in which all students in the same classroom receive the same instruction, and the creation of small heterogeneous groups. Sometimes cooperative learning strategies are employed with heterogeneous groups, but cooperative learning can be used with any small group regardless of the crite-

To create more homogeneity, teachers may divide students into small instructional groups reflecting different levels of ability.

Critics argued that tracking and ability grouping do not separate students into socioeconomic status-related groups by accident.

rion by which it is formed. Success for All, for example, is a popular program combining cooperative learning with small ability groups that are frequently reorganized to reflect student progress.[17]

Controversy

In the 1970s and 1980s, a barrage of studies criticized tracking and ability grouping. Race and class figured prominently in the debate. Grouping students by ability, no matter how it is done, will inevitably separate students by characteristics that are correlated statistically with measures of ability, including race, ethnicity, native language, and class. Critics argued that tracking and ability grouping do not separate students into socioeconomic status-related groups by accident. Ray C. Rist's "Self-Fullfilling Prophecy in Ghetto Education" (1970) followed a group of kindergarten students through the first few years of school and noted how the composition of reading groups rarely changed, consistently reflecting students' socioeconomic status (SES).[18] The SES differences are hardened, Rist argued, as teachers develop different expectations for groups of low and high performing students, even if those groups are given innocuous sounding names to mask their status.[19] James Rosenbaum's *Making Inequality* (1976) described working class youth at a New England high school who were channeled into vocational and remedial tracks that were nothing more than boring, academic dead ends.[20]

In 1985, Jeanie Oakes' classic book, *Keeping Track*, was published. Oakes drew on data from several junior and senior high schools. Building on the social reproductionist theories of Samuel Bowles and Herbert Gintis's *Schooling in Capitalist America*, Oakes argued that although tracking is typically justified by educators as a strategic response to student heterogeneity, the practice is undergirded by normative beliefs regarding race and class—and politically defended by white, middle-class parents to protect privilege. Black, Hispanic and poor children populate remedial classes; middle-class white children populate honors courses. Tracking and ability grouping are not mere bystanders to social injustice, Oakes and other critics charged. Such practices don't just mirror the inequalities of the broader society. They reproduce and perpetuate inequality.[21]

This critique had a profound effect on policy and practice. In the 1990s, several prominent political organizations passed resolutions condemning tracking, including the National Governors Association, the American Civil Liberties Union, the Children's Defense Fund, and the NAACP Legal Defense Fund. Some states urged schools to reduce tracking and ability grouping, most notably California and Massachusetts. A surprising implementation story ensued. Although the call to detrack was not accompanied by conventional incentives—the big budgets, regulatory regimes, and rewards and sanctions that draw the attention of policy analysts—detracking was, in a field famous for ignored or subverted policies, adopted by a large number of schools.[22]

Surveys of Ability Grouping

How much did ability grouping decline? A 1961 national survey revealed that about 80% of elementary schools grouped students by ability for reading instruction.[23] A three-group format was the dominant approach, with students organized into high, middle, and low performing groups. Although subsequent national surveys of ability grouping are scarce until the John Hopkins study in the mid-1980s (mentioned above), carefully crafted studies of local practice reported similar frequencies. Eighty percent or more of elementary schools used within-class ability groups.[24]

Then things changed. A mid-1990's survey of a random sample of pre-K through fifth grade teachers reported startlingly different results. When allowed multiple responses, only 27% of teachers reported using ability grouping for reading instruction. Another 56% of teachers indicated that they used flexible grouping. Some of the teachers with flexible grouping may have utilized ability as a criterion for grouping.[25] Whole class instruction was by far the most popular organizing strategy, with 68% of teachers reporting its use. Removing the overlapping responses makes it clear that ability grouping served a subordinate role as a method of organizing students. When teachers were held to one response and asked to identify their *primary* organizational approach, the order was: whole-class instruction (52%), flexible grouping (25%), and ability grouping (16%).

A more recent survey suggests ability grouping has regained favor among teachers. Barbara Fink Chorzempa and Steve Graham (2006) surveyed a national random sample of first through third grade teachers. Their questionnaire asked questions similar to the Baumann et al. survey of the 1990s, but also included questions about why teachers ability group. Three times as many teachers (63%) said they use ability grouping as the earlier survey. The authors explain that the discrepant findings may stem from the different grade levels of teachers in the two surveys. Pre-K and fourth- and fifth-grade teachers, who are included in the earlier survey but not in the latter, may be less likely to employ ability grouping than first through third-grade teachers, the target population of the latter survey. Interestingly, the top reason teachers gave for using ability grouping was "that it helps them meet students' needs;" however, respondents also expressed concern about the quality of instruction in low ability groups.[26] About 20% of teachers did not ability group at all because the practice was banned by district or school policy.

Is ability grouping in decline or on the rise again? How about tracking? Let's turn to NAEP data to shed light on these questions.

Table 2-1

Reading Grade 4
Basis for Reading Instructional Groups
(percent of students)

Year	Ability	Other	Not Created
2009	71	21	8
2007	64	20	15
2005	59	22	19
2003	47	25	27
2000	39	29	32
1998	28	33	39

Note: The category "other" is a collapsed variable comprising responses for "interest," "diversity," and "other." For 1998-2000 and 2009, NAEP question T068301; For 2003-2007, NAEP question T068351.

Source: "NAEP Data Explorer," National Center for Education Statistics, http://nces.ed.gov/nationsreportcard/naepdata/.

NAEP Data on Ability Grouping

Table 2-1 displays NAEP data on ability grouping in fourth grade reading. Teachers were asked on what basis they create instructional groups (ability, interest, diversity, and other) with "not created" also an option. Bear in mind that asking fourth-grade teachers about ability grouping, as compared to sampling teachers of several elementary grades, has both an upside and a downside in elucidating trends. The upside is that grade level is held constant over several surveys. This is important because we know ability grouping varies by grade level. The downside is that fourth grade isn't where the action is on ability grouping—that's first grade, where unfortunately NAEP

…the top reason teachers gave for using ability grouping was "that it helps them meet students' needs"…

Math Grade 4
Create Groups for Math Based on Ability?
(percent of students)

Table **2-2**

Year	Yes	No
2011	61	39
2009	54	46
2007	46	54
2006	49	51
2003	42	58
2000	41	59
1996	40	60
1992	48	52

Note: All data from NAEP question T044201.
Source: "NAEP Data Explorer," National Center for Education Statistics.

Tracking in 8th Grade
(percent of students)

Table **2-3**

Year	Mathematics	English Language Arts	Science	History
2011	76	–	–	–
2009	77	–	–	–
2007	75	–	–	–
2005	73	–	–	–
2003	73	43	–	–
2000	73	–	26	–
1998	–	32	–	15
1996	71	35	21	–
1994[1]	72	37	19	17
1992[1]	73	50	–	–
1990[1]	75	60	29	29

Source: "NAEP Data Explorer," National Center for Education Statistics.

The percentage of students placed into ability groups for reading instruction skyrocketed from 1998 to 2009, from 28% to 71%.

does not collect data. Fourth grade is well after ability grouping's apogee and somewhere near the midpoint of its diminishing use by elementary teachers.

Table 2-1 is revealing. The percentage of students placed into ability groups for reading instruction skyrocketed from 1998 to 2009, from 28% to 71%. And the percentage of students whose teachers did not create ability groups fell from 39% in 1998 to 8% in 2009. In other words, the odds of a fourth grader being ability grouped in reading were less than 50-50 in 1998 but by 2009 had increased to about 9 to 1. The question was not asked prior to 1998.

Table 2-2 shows the frequency of ability grouping in fourth-grade mathematics. Teachers were asked if they create math groups based on ability. This question was asked twice before 1998 and in 2011, so it gives a deeper historical perspective than the question on reading. Math ability grouping dips from 1992 to 1996 (48% to 40%), stays about the same until 2003 (42%), and then accelerates from 2003 to 2011 (reaching 61% in 2011).

The NAEP data support the general finding of a drop in ability grouping in the 1990s and a resurgence in the 2000s. The rebound is more subdued in math than in reading. It is apparent by 2000 in reading (it may have begun even before then; the data start in 1998) but does not begin in math until after 2003. In the years for which data are available for both reading and math (2000, 2003, 2007, 2009), the two subjects have comparable frequencies in 2000 (39% in reading and 41% in math), but reading is more often grouped in subsequent years. In the last year with data on both subjects, 2009, 71% of fourth grade students were ability grouped for reading and 54% for math.

NAEP Data on Tracking

Table 2-3 displays NAEP data on tracking in 8th grade. Note that unlike ability grouping, which is a classroom level practice and consequently a topic for teacher surveys, tracking is a school level practice and a topic for surveys of school principals. Although the wording of the survey item varies slightly from year to year, NAEP asks principals whether students are assigned to classes based on ability so as to create some classes that are higher in average ability or achievement than others. The question is asked sporadically and about different subjects in different years.

Math has the most data, surveyed ten times from 1990–2011. Tracking in math shows a slight dip in the 1990s and an increase in the 2000s, but most of the fluctuations are too small to consider significant. The trend is essentially flat, with about three-fourths of students attending tracked math classes over the past two decades. Typically, this means schools offer an algebra class for some eighth graders and a pre-algebra class for those who are not yet ready for formal algebra (see table 3-2 for enrollment statistics). Sometimes a third class is offered, perhaps geometry for students who took algebra in seventh grade or a basic math class for students several years behind.

Data on the other subjects are spotty. They exhibit much less tracking than math and greater variation over time. In 1990, principals reported that 60% of students were in tracked ELA classes, a statistic that declined over the next several years, hitting a low of 32% in 1998. The 43% frequency of tracking reported in 2003 is an increase from 1998; however, because it was the last time the question was asked in that subject, it is impossible to tell whether an enduring rebound in ELA tracking had begun. Science and history have even less data, with both subjects registering their highest figures in 1990 and then indicating diminished tracking after that. Science seems to show a rebound from 1994–2000. For all four subjects, the least amount of tracking occurred between 1994 and 1998, when the detracking movement was in full bloom.

The national pattern is consistent with previous studies of California and Massachusetts. In those two states, detracking was most intense in the early to mid-1990s, but differences among the subjects emerged. Mathematics resisted detracking while heterogeneously grouped classes became the norm in ELA, science, and history. In a 2009 survey of Massachusetts schools with eighth grades, for example, in math only 15.6% of schools offered heterogeneously-grouped classes; 49.2% offered classes with two ability levels; and 35.2% offered three levels. In other subjects, tracking had almost disappeared—72.7% offered only heterogeneously-grouped classes in ELA, 89.8% in history, and 86.7% in science.[27]

Discussion

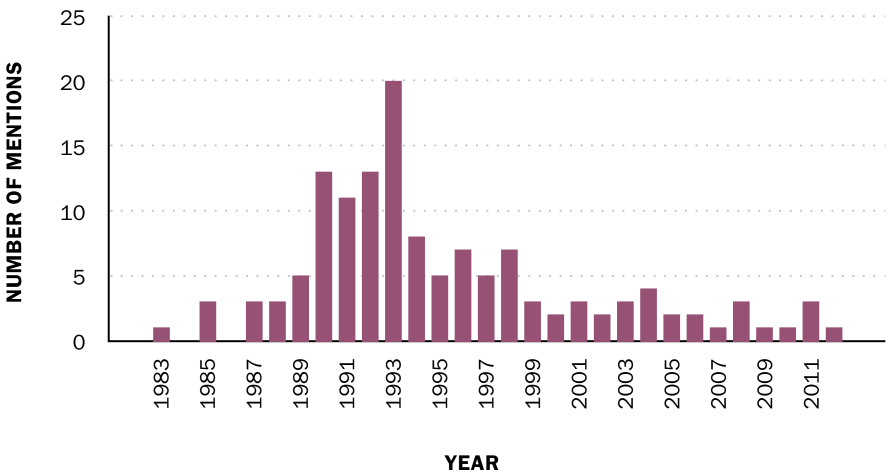
This study has explored trends in the use of ability grouping and tracking by American schools. It used NAEP data to examine the frequency that fourth graders are assigned to groups and eighth graders assigned to classes based on ability or prior achievement. The investigation focused on what schools are doing, not on whether tracking or ability grouping is a good idea.

NAEP data from 1990 to 2011 were examined. Ability grouping in fourth grade decreased in the 1990s and then increased markedly in the 2000's, with the rebound apparent in both reading and math. In reading, ability grouping has attained a popularity unseen since the 1980s, used with over 70% of students. As for tracking, it has remained commonplace in eighth-grade mathematics for the past two decades, with about three-quarters of students enrolled in distinct ability-level math classes. Tracking in ELA declined sharply from 1990 to 1998, and although there was a rebound in 2003, NAEP has not surveyed schools on tracking in ELA since then. And NAEP data are too sparse in other subjects to determine trends.

Ability grouping in fourth grade decreased in the 1990s and then increased markedly in the 2000's, with the rebound apparent in both reading and math.

Mentions of "Ability Grouping" in *Education Week* (1983–2012)

Fig. 2-1



Source: "Education Week Advanced Search," *Education Week*, http://www.edweek.org/search.html.

Do these trends matter? Why should anyone care about tracking and ability grouping? Although the debate today is more subdued than in the 1980s and 1990s, it does continue. A research review on the NEA website blasts both tracking and ability grouping as discriminatory.[28] Scholars continue to wrangle over the wisdom of both practices. Effectiveness and equity persist as the dominant themes of this literature. A 2010 meta-analysis of high quality studies calculated a positive effect size of 0.22, equal to about one-half year of learning, for within-class grouping in reading instruction.[29] A 2010 study of data from the Early Childhood Longitudinal Study (ECLS), on the other hand, found "students who are lower grouped for reading instruction learn substantially less, and higher-grouped students learn slightly more over the first few years of school, compared to students who are in classrooms that do not practice grouping."[30] That finding is especially relevant to closing achievement gaps between students who may populate high and low groups.

The controversy offers a very important lesson about how education policy gets implemented in schools. Schools are not merely the last step of a vast organizational ladder, not simply the education system's operational frontline, ready to put in place the policies that are passed down from above. Finley Peter Dunne famously observed that the U.S. Supreme Court "follows the election returns." Court decisions not only reflect the U.S. Constitution but public opinion as well. Our schools are another institution with an ear to the ground. Educators are aware of public debates and are influenced when particular school practices become controversial.

Figure 2-1 shows the number of times the term "ability grouping" appeared in *Education Week* from 1983 to December 2012. Consider this a proxy for media visibility over the past thirty years. The 135 appearances over these three decades represent an average of 4.5 mentions per year. The peak coverage occurred in 1993, with 20 mentions. The years immediately preceding 1993 show a gradual build up in coverage, with 5 mentions in 1989, 13 in 1990, 11 in 1991, and 13 in 1992. The years immediately after 1993 show a gradual decline—8 appearances in 1994, 5 in 1995, 7 in 1996, 5 in 1997, and 7 in 1998. The ten years from 1989–1998 are the only years with more than 5 annual mentions. Tracking and ability grouping were in the spotlight.

The data on media visibility are inversely related to the data on use. At the beginning of the 1990s, tracking and ability grouping were conventional practices but then declined —albeit with some lag time— when they were subjected to the most public scrutiny. The mentions in *Education Week* peaked in 1993. The use of ability grouping and tracking reached all time lows soon after that event. As the controversy died down in the 2000s, schools returned to both practices.

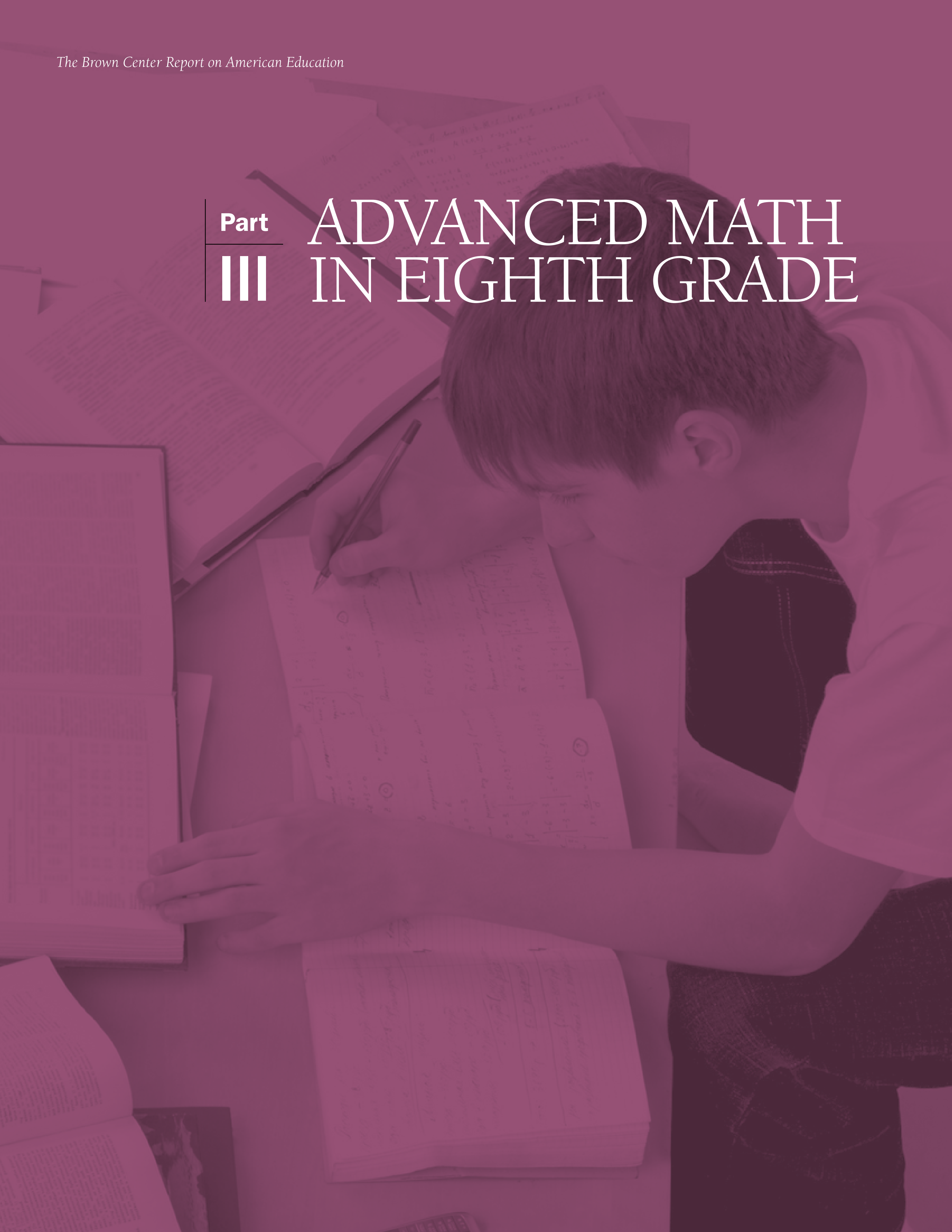
What else may have promoted the resurgence in the 2000s? Accountability systems, bolstered by the accountability provisions of No Child Left Behind, focus educators' attention on students below the threshold for "proficiency" on state tests. That provides a statutory justification for grouping students who are struggling. The increased use of computer instruction in elementary classrooms cannot help but make teachers more comfortable with students in the same classroom studying different materials and progressing at different rates through curriculum. The term "differential instruction," while ambiguous in practice, might make grouping students by prior achievement or skill level an acceptable strategy for educators who recoil from the term "ability grouping."

A substantial number of teachers believe that heterogeneous classes are difficult to teach. The 2008 *MetLife Survey of the American Teacher* asked teachers to react to the following statement: "My class/classes in my school have become so mixed in terms of students' learning ability that I/teachers can't teach them." Responses were: 14% "agree strongly," 29% "agree somewhat," 28% "disagree somewhat," and 27% "disagree strongly."[31] The percentages are surprising given the questionnaire's blunt assertion that heterogeneous classes are *impossible* to teach. Moreover, the 43 percent of respondents that either agree strongly or somewhat agree with the prompt is up from 39 percent on the same survey item in 1988. Teachers' beliefs about the impact of achievement heterogeneity on instruction undergird the use of ability grouping and tracking.

Let's look ahead. Will the uptrend in ability grouping continue? Not necessarily. The current period may be the lull before the storm. Theoretically, at least, the Common Core establishes a curriculum that most, if not all, students will study. It is unclear how students who have already mastered the Common Core standards before beginning a particular school grade will have their needs met under the new regime. The same goes for students who lag many years behind. Tracking and ability grouping have been common approaches to addressing such challenges. These two organizational strategies affect millions of students daily. Both practices shape aspects of schooling that we know to be important—the curriculum students study, the textbooks they learn from, the teachers who teach them, the peers with whom they interact. Despite decades of vehement criticism and mountains of documents urging schools to abandon their use, tracking and ability grouping persist—and for the past decade or so, have thrived.

The mentions in Education Week *peaked in 1993. The use of ability grouping and tracking reached all time lows soon after that event.*

Part III ADVANCED MATH IN EIGHTH GRADE



As recently as 1990, taking algebra in eighth grade was unique. That has changed dramatically in recent years, and now more eighth graders take algebra than any other math class. Enrollment in eighth-grade algebra—and in other advanced math classes—varies by state. This section of the Brown Center Report exploits that variation to study the relationship of states' enrollment in advanced math classes and scores on NAEP. The research question is whether a relationship exists between changes in advanced math enrollments and changes in 8th grade NAEP scores. Do states that boost advanced enrollments experience a concurrent increase in achievement? A second analysis uses the same technique to look at the potential that advanced courses are being "watered down." Are rising enrollments associated with lower mean achievement in advanced classes?

Background

In 1982 Robert Moses was awarded a MacArthur Fellowship. He used the money to start The Algebra Project, a community-based effort to bring algebra to historically underserved middle school students—primarily, children from low income households and students of color. Moses called algebra "the new civil right," an invocation of equity that cast course taking in a new light.[32] The Clinton Administration tied the equity theme to international competitiveness and pushed for more students to take algebra before high school. "Around the world, middle students are learning algebra and geometry," President Clinton observed. "Here at home just a quarter of all students take algebra before high school."[33]

Algebra soon came to be known as a "gatekeeper" course, a class standing like a sentry at the gateway to college. Take it and pass it and your odds of attending college were good. Take it and fail it and at least you had been exposed to challenging mathematics. Don't take it at all and your chances of attending college were near zero. Algebra's place in the typical high school math sequence enhanced its importance. Assume that college-going students should get some calculus under their belts in the senior year. In most high schools, a student who takes

Algebra I in ninth grade has three remaining years to take Algebra II, Geometry, Pre-Calc/Trigonometry, and then Calculus. That's four courses. Something has to give. Many schools change the order of the courses, and some mix in statistics with one of the year's offerings, but the fact remains: if taking Calculus as a senior in high school is the goal, taking Algebra I in ninth grade means there are four courses to complete in three years. Taking algebra in eighth grade opens up an additional year for advanced math.

Equity, international competiveness, and practical concerns about course sequences converged in the mid 2000s to boost the campaign for eighth-grade algebra. An "algebra for all" movement emerged that pushed universal, mandatory eighth-grade algebra. Minnesota established a new high school graduation requirement that, beginning with the class of 2015, all students must complete an Algebra I credit by the end of eighth grade. California used its school accountability formula to promote eighth-grade algebra, offering a choice of two eighth-grade math assessments (algebra and general eighth-grade math) but then, in the formula for calculating Academic Performance Index (API), discounting the performance level of students taking the general math test (for example, downgrading to "basic" those students who took the test and scored "proficient"). That incentive motivated schools to dramatically increase eighth-grade algebra enrollments, and although the AYP rule was later tossed out by the courts, California ranks as the top state in the nation for eighth-grade algebra and advanced math enrollments.[34]

NAEP Data on Advanced Math Enrollment

Table 3-1 illustrates the steady increase of U.S. eighth-grade enrollment in advanced mathematics courses. The data are taken from the NAEP eighth-grade math assessment. Students are asked: "what mathematics class are you taking this year?" The category "advanced mathematics" combines several responses, including Algebra I, courses that stretch Algebra I content over two years (whether it's the first or second year of such a course), and courses that typically are more advanced than Algebra I, including Algebra II and Geometry. This amalgamated response is noisy and receives further discussion below.

In 1990, only 16% enrolled in an algebra course, compared to 20% in pre-algebra and 61% in 8th grade math. In this paper, the latter two courses are referred to as "basic." By 2011, nearly half (47%) of all eighth graders took algebra or a more advanced course. Only 48% were in a basic math course, down from 81% in 1990. The advanced math percentage may be understated in Table 3-1 for the years prior to 2000 as that was the first

8th Grade NAEP Enrollment in Advanced Mathematics Courses, 1990–2011 — Table 3-1

Year	Percent
2011	47%
2009	44%
2007	43%
2005	42%
2003	33%
2000	27%
1996	25%
1992	20%
1990	16%

Note: Advanced math includes Algebra I, Algebra II, Geometry, Algebra "stretched over 2 years."
Spaces indicate when response categories changed.

Source: "NAEP Data Explorer," National Center for Education Statistics, http://nces.ed.gov/nationsreportcard/naepdata/.

As recently as 1990, taking algebra in eighth grade was unique.

More and more students are taking advanced math classes earlier and earlier. Is this a good idea?

time geometry, advanced algebra, and algebra stretch classes were response categories in the NAEP questionnaire for eighth graders.[35] Moreover, some students—both then and now—may mistakenly believe they are in an algebra or geometry class when in fact they are not. Notwithstanding these data limitations, advanced math enrollments clearly rose substantially from 1990 to 2011.[36]

More and more students are taking advanced math classes earlier and earlier. Is this a good idea?

Research on the Efficacy of Eighth-grade Algebra

The National Educational Longitudinal Study of 1988 (NELS) offers researchers a trove of information collected from a randomized sample of students. Several studies have used NELS data to investigate what happens when students take advanced math early in an academic career, whether eighth or ninth grade.[37] Researchers found gains for students taking algebra earlier rather than later, including—and this is important for the equity goal—low performing students. A recent meta-analysis of research on the topic (by Mary K. Stein and colleagues) reaffirmed that positive finding, with the caveat that "achievement gains occurred in settings where policies were accompanied by strong supports for struggling students, particularly more time for algebra instruction. We do not have strong evidence that universal algebra policies lead to achievement gains minus those strong supports."[38]

More recent evaluations of policies expanding algebra enrollment have raised cautionary flags. Chicago mandated that all ninth graders take what had been regarded as college preparatory classes, including algebra. Evaluators followed students for several years and concluded, "Although more students completed ninth grade with credits in algebra and English I, failure rates increased, grades slightly declined, test scores did not improve, and students were no more likely to enter college."[39] Studies of California's algebra policies found a trade-off: rising enrollments but also a rising number of failures. In North Carolina, researchers from Duke uncovered negative results after studying a Charlotte-Mecklenburg initiative to expand algebra in eighth grade: lower scores on the Algebra I test and then lower pass rates in Geometry and Algebra II in subsequent years.

Why have the more recent studies produced bleaker findings than suggested by the earlier work? The Duke researchers believe selection bias skewed the earlier findings. Stronger math students take algebra in eighth grade, and although they indeed may benefit academically from the course, that does not mean that weaker students will also benefit from taking algebra earlier. "Once this selection bias is eliminated, the remaining causal effect of accelerating the conventional first course of algebra into earlier grades, in the absence of other changes in the math curriculum, is for most students decidedly harmful."[40]

The Stein et al. meta-analysis and the Duke team's policy recommendations, although different in emphasis, do share a small patch of common ground. Stein et al. say that without "strong supports" achievement gains cannot be expected. And the Duke researchers foresee harmful effects "in the absence of other changes in the math curriculum." One is contingently positive, the other contingently negative. The common ground that they share is in forecasting the potential for a neutral effect.

Let's return to NAEP and see what its data have to say about state efforts to encourage enrollment in advanced math courses in eighth grade.

Analytical Method

Are eighth-grade enrollments in advanced math related to states' math scores on NAEP? To answer this question, an obvious first step is simply to examine the list of states, their NAEP scores, and the percentage of each state's students taking algebra, geometry, and other advanced math courses in eighth grade. There is no clear relationship. In 2011, the correlation between states' advanced math enrollments and NAEP achievement is 0.07, indistinguishable from 0.00. States with more eighth graders taking advanced math classes are no more likely to register a higher NAEP score in math than states with lower enrollments in those classes.

This kind of cross-sectional analysis is a reasonable place to start, but it's limited to revealing correlations between variables at a single point in time. That can be misleading. A study in the 2007 Brown Center Report, for example, showed how the number of instructional minutes that nations devote to math instruction is unrelated, on a cross-sectional basis, to national math achievement. In 1995, the correlation was 0.05. In 2003, the correlation was -0.20. Neither statistic is significantly different from 0.00. But when nations are examined longitudinally, and data from the two cross-sections are modeled as change variables, the question under scrutiny is shifted to whether national changes in instructional minutes from 1995 to 2003 are related to changes in test scores over the same time period. The correlation for that relationship is 0.42, which is statistically significant. Countries that increased the amount of time devoted to math instruction tended to experience a rise in TIMSS math scores; those countries that decreased the time devoted to math instruction tended to see their scores fall.

Why is the analysis of change variables beneficial? Two reasons. The first is that the technique helps to control for bias introduced by omitted variables (including selection), a shortfall plaguing cross-sectional analyses of achievement. In the case of instructional minutes, for example, school systems might strategically decide to place low achieving students in longer classes to help them catch up. That would make it appear that more instruction is associated with lower achievement. Assuming that omitted variable bias is present at both the beginning and end points of the time interval under study—and the relationship to the dependent variable (the outcome of interest) remains consistent over the interval—such bias washes out in the calculation of change (see Gustaffson, 2007, for further explanation and applications to other educational questions).[41]

The second benefit of this approach is that it poses a question paramount to policy analysis. Considering whether to adopt policy X leads to the question: if we adopt policy X, what is the expected change in outcome Y? What will happen? The cross-sectional question is this: what is the relationship of policy X to outcome Y at one point in time? One often hears of cross-sectional analyses showing something along the lines of "a one-standard deviation change in X would result in the following change in Y," but the prediction is only inferred, there being no observations of change (or data from different time periods) in the data set.

Analysis of Change Using NAEP Scores

The relationship between change in policy and change in outcome is the subject of the analysis below. The time period examined is 2005 to 2011. Be aware, notwithstanding the improvement over cross-sectional analysis, that the analysis is still only correlational and thus confined to generating plausible hypotheses for more rigorous research

States with more eighth graders taking advanced math classes are no more likely to register a higher NAEP score in math than states with lower enrollments in those classes.

National Enrollment Changes in 8th-grade Math Classes, 2005–2011 (percent of students, by course)

Table 3-2

	Geometry	Algebra II	Algebra I	Stretch Alg (2 yr)	Pre-Alg	General Math	Integrated	Other
2005	4	3	30	5	27	25	1	5
2011	5	4	34	4	23	25	1	4
Change	+1	+1	+4	-1	-4	0	0	-1
Change	Advanced +5				Basic -4		Other -1	

Source: "NAEP Data Explorer," National Center for Education Statistics.

States with rising advanced math enrollments experienced shrinking enrollments in basic courses.

designs. No causality is asserted here.

Table 3-2 shows the tail end of the long term trend sketched in Table 3-1—enrollment gains in advanced math classes and declines in basic classes. The slow, steady national trend masks considerable variation among the states. In 2005-2011, the average state increase in advanced math enrollments (as a proportion of eighth graders) was 5.5%, with a standard deviation of 8.4%. The top four states that boosted advanced enrollments were Minnesota (35%), and Pennsylvania, Virginia, and Washington (all with 17%). In contrast, two states stand out for going against the national trend with shrinking advanced math enrollments: Nevada (-22%) and Georgia (-17%).

In terms of specific courses, forty-five states boosted enrollments in Algebra I, while only three states shrank enrollments and three stayed the same (in this discussion of NAEP scores, the District of Columbia is considered a state). Twenty-eight states decreased enrollments in general math, twenty increased, and three stayed the same. In general, course enrollments behave like a tube of toothpaste—squeeze on one end and the other end bulges. States with rising advanced math enrollments experienced shrinking enrollments in basic courses. And vice versa. The two states singled out for declining enrollments in advanced math courses illustrate the point. Their basic math enrollments rose. Nevada's pre-algebra enrollments jumped 27%. Georgia's percentage of students in general math rose 33%.

Is there a relationship between states' change in course enrollments and change in NAEP scores? Did states experience gains on NAEP concurrent with increases in eighth graders taking advanced math? A series of correlation coefficients were computed to investigate these questions (see Table 3-3). The first model examines the relationship of advanced math enrollments and NAEP composite scores. The correlation coefficient (r = -0.01) is statistically indistinguishable from 0.00.

The NAEP composite score may assess mathematics too broadly to pick up the effects of emphasizing advanced math, which primarily involves boosting algebra. Fortunately, NAEP reports scores on specific content areas assessed within the test (called "strands"), including algebra and geometry. So the second model uses the NAEP subscore for the algebra strand as the achievement variable, which should be more sensitive to increased knowledge of algebra. Again, no significant relationship is found.

The third and fourth models use change in Algebra I enrollments as the course variable, instead of advanced math, in case aggregating several courses into the "advanced" category has muddied the waters. The change in composite NAEP score serves as the achievement variable in the third model and the change in the algebra strand score as the achievement variable in the fourth model. Neither correlation attains statistical significance.

Models five and six repeat the same treatment with geometry. Change in geometry course taking in eighth grade is used as the course variable—and the models calculate whether it is correlated with change in the NAEP composite in model five and change in geometry score in model six. Neither correlation is statistically significant.

In addition to the correlations reported here, multivariate regressions were run with three covariates controlled (also variables representing change)—change in state rates of child poverty, English language learners, and black and Hispanic students—demographic characteristics that are known correlates of state NAEP scores. The Great Recession unfolded during the time period under study, and some states, for example, witnessed growing rates of child poverty more than other states. If states experienced demographic changes, that could skew the results. It turned out not to be the case. None of the regression models were statistically significant.

In sum, no evidence was found in NAEP scores of a relationship between states raising enrollment in advanced math courses and raising achievement. States that increased the percentage of students taking algebra or geometry in eighth grade were no more likely to post NAEP gains than states with decreased enrollments in those two courses.

Relationship of Change in 8th Grade State Course Enrollment and State NAEP Score, 2005–2011

Table **3-3**

Model	Course (% enrolled)	State NAEP Score	Correlation
1	Advanced Math	Math Composite	-0.01
2	Advanced Math	Algebra	0.14
3	Algebra I	Math Composite	-0.03
4	Algebra I	Algebra	0.07
5	Geometry	Math Composite	0.11
6	Geometry	Geometry	-0.03

Note: Models 1-5 (n=50). Model 6 (n=36) because 14 states had insufficient enrollments in geometry to generate a score. Data for Alaska were unavailable.

Source: Author's calculations, "NAEP Data Explorer," National Center for Education Statistics.

Do Rising Enrollments Water Down Advanced Math Courses?

Whether advanced math courses are watered down because of increasing enrollments is an important question. The notion is that filling advanced classes with academically weaker students than in the past could diminish the amount of learning that the courses are able to impart. That could help to explain the neutral correlations reported above. It could also help to explain the neutral—or even negative effects—revealed by recent evaluations of policies promoting universal algebra in eighth and ninth grades. NAEP data can only go so far in indicating whether watering down is taking place, but they do offer interesting insights into how course-shifting and achievement may be related.

Table 3-4 reports correlations between enrollment change and change in the mean achievement of students taking each course. Data from four courses are displayed. Again, the percentage of a state's eighth graders taking each course serves as the enrollment variable. The courses are arranged hierarchically. Geometry is typically offered for the most advanced students and general math for the weakest ones. Three correlations are statistically significant.

Is there evidence of watering down? Yes, but not in all advanced courses. Let's start with the results supporting the watering down hypothesis. Increases in Algebra I enrollments are negatively associated with

In sum, no evidence was found in NAEP scores of a relationship between states raising enrollment in advanced math courses and raising achievement.

The analysis of whether accelerating students into advanced classes is watering down achievement points to two different types of acceleration.

Relationship of Change in State Course Enrollment and Mean NAEP Score for Students in Each Course, 2005–2011

Table 3-4

Course	Correlation
Geometry	0.27
Algebra I	-0.34*
Pre-Algebra	-0.34*
General Math	0.47**

* $p < 0.05$

** $p < 0.01$

Source: Author's calculations, "NAEP Data Explorer," National Center for Education Statistics.

achievement gains ($r = -0.34$, $p < .05$). Let's be clear what that means. The average state registered a 5.6 NAEP scale score gain among its Algebra I students. The NAEP scores for students in Algebra I classes did not go up as much in states that raised enrollments in Algebra I (+5.2) as in states that either held enrollments constant or decreased them (+9.2). For Pre-Algebra, rising enrollments are also negatively associated with test scores ($r = -0.34$, $p < .05$). Both correlations are consistent with the watering down hypothesis if students who would otherwise be placed in lower courses are migrating upward to higher courses. We cannot tell whether that is happening using NAEP data. And, to issue an important warning once again, correlations do not prove causality.

The strongest correlation involves General Math ($r = 0.47$, $p < .01$). The positive association is also consistent with the watering down hypothesis. If the overall trend is to move students into upper-level courses—and schools are selective in the students they accelerate—General Math courses, as they shrink, should be increasingly dominated by the students who struggle the most at math. These courses presumably would have lost their best students. Falling enrollments would therefore be associated with falling scores. General Math classes that manage to keep the students who are being accelerated elsewhere would, comparatively, register higher scores.

Geometry complicates matters. Its correlation coefficient (0.27) is inconsistent with the watering down story. Geometry sits at the top of the course hierarchy. Any indiscriminate acceleration of students upward (an inextricable assumption of the watering down argument) should ultimately result in a negative association of enrollment gains and achievement scores in the course at the top. And yet, Geometry's correlation coefficient has a positive sign and approaches statistical significance. Although statistically indistinguishable from 0.00 ($p = .11$), that could be due in part to the reduced number of states with data. Only thirty-six states have sufficient numbers of eighth-grade geometry students to produce a NAEP score.

Another possibility involves the noisy NAEP course variables. Perhaps more "real" geometry students are included in the NAEP course category for geometry in 2011 than in 2005—in other words, a larger proportion who are actually in a geometry class and not mistaken about their math course. As indicated in Table 3-2 above, only 5% of eighth graders were enrolled in Geometry in 2011, an increase from 4% in 2005. The mean NAEP score for geometry students was 290 in 2005 and 308 in 2011, a sharp increase of 18 points. The one-percentage-point gain in students seems to have packed a punch in terms of NAEP scores. The "real" geometry students probably took Algebra I in seventh grade. Much like algebra for eighth graders three or four decades ago, geometry is reserved for today's very best math students.

Discussion

This study analyzed variation in state enrollment patterns to test whether rising enrollments in advanced eighth-grade math courses are correlated with achievement gains on NAEP. No evidence was found that they are. States with rising percentages of eighth graders taking Algebra I, Geometry, and other advanced math classes were no more likely to raise their NAEP scores from 2005-2011 than states with declining percentages of eighth graders in those courses.

A second analysis, again looking at changes in policy and test scores over time, investigated whether boosting the percentage of students in higher level courses is associated with decreases in the mean scores of those courses—suggesting a watering down effect. The evidence is consistent with watering down in all but one course. Negative correlations were found for Algebra I and Pre-Algebra. In those courses, mean achievement gains declined as enrollments increased. Achievement gains in general math courses were positively associated with enrollment changes. All three of these correlations are statistically significant and supportive of the watering down hypothesis.

Geometry diverges from the other courses. A positive association was found that, although statistically indistinguishable from 0.00, suggests at least a neutral relationship between rising enrollment and changes in NAEP scores. If schools were indiscriminately accelerating students into eighth-grade geometry, one would expect a negative correlation.

None of these findings can confirm or reject causality, but they are useful in generating hypotheses for future study. They also shed light on the findings from previous research. For example, a key finding from evaluations of California's algebra policy is that universal algebra produces trade-offs. Many students benefit from the extra challenge. Rates of algebra enrollment for historically under-enrolled populations (in particular, low SES students) have increased. The raw number of students passing end of course exams has also increased. But the downside is that the number of students failing algebra goes up as well; and the failing students, too, are disproportionately low SES students.[42] One study from California suggests that many of the failing students would have been better off spending an additional year preparing for algebra instead of taking it.[43] These kinds of trade-offs, when aggregated to the state level, could produce a neutral net effect.

The analysis of whether accelerating students into advanced classes is watering down achievement points to two different types of acceleration. One is selective and decided on an individual basis. Each student's math skills are evaluated and a determination is made whether a more advanced math course is appropriate or not. That kind of acceleration appears to be occurring in eighth-grade geometry—and presumably in seventh-grade algebra. Students who would benefit from a more rigorous course are promoted. Mean test scores for eighth-grade geometry rise, or at least stay the same, despite rising enrollments.

The second type of acceleration is non-selective and group based. Students are advanced based on a characteristic independent of prior achievement or preparedness (e.g., grade level or age). Future research should compare these two types of acceleration and investigate who, when it comes to selective acceleration, should be accelerated and when. With age- or grade-based acceleration, a set of early indicators is needed (the universal algebra approach) that would identify students needing support and the type of support most ben-

Future research should compare these two types of acceleration and investigate who, when it comes to selective acceleration, should be accelerated and when.

Taking algebra in eighth grade is the new normal.

eficial for them. If the trade-offs of group acceleration are indeed real, then the policy goal should be to minimize negative effects and maximize benefits.

A final note on the Common Core. No one knows how gifted students' needs will be met in the Common Core Era. Taking algebra in eighth grade is the new normal, and taking algebra in the seventh grade is rapidly becoming the new normal for gifted math students. In California, 8.1% of seventh graders (nearly 38,000 students) took the algebra end of course exam in 2012. If Common Core means the same curriculum for all, a time will surely come when exceptional math students need an uncommon curriculum that is appropriate for them.

NOTES

1 In this section, the following rule was applied to ease the reading of the text. Subnational units, such as Hong Kong, may be referred to as "nations" or "countries."

2 "2011 NAEP-TIMSS Linking Study," National Center for Education Statistics, http://nces.ed.gov/timss/naep-link.asp.

3 Linking NAEP and international tests has been attempted before. See Gary W. Phillips, *International Benchmarking: State Education Performance Standards* (Washington, DC: American Institutes for Research, October 2010) and "Global Report Card," Jay P. Greene and Josh B. McGee, http://globalreportcard.org/.

4 See "The Finnish Miracle," Hank Pellissier, http://www.greatschools.org/students/2453-finland-education.gs.

5 Jenny Anderson, "From Finland, an Intriguing School Reform Model," *New York Times*, December 12, 2011, http://www.nytimes.com/2011/12/13/education/from-finland-an-intriguing-school-reform-model.html?pagewanted=all&_r=0.

6 Pasi Sahlberg, *Finnish Lessons: What Can the World Learn from Educational Change in Finland?* (New York: Teachers College Press, 2011).

7 The Flemish, French, and German speaking communities operate separate school systems.

8 William H. Schmidt and Richard T. Houang, "Lack of Focus in the Mathematics Curriculum: Symptom or Cause?" in *Lessons Learned: What International Assessments Tell Us about Math Achievement*, ed. Tom Loveless (Washington: Brookings Institution Press, 2007).

9 William H. Schmidt and Richard T. Houang, "Curricular Coherence and the Common Core State Standards for Mathematics," *Educational Researcher* 41, no. 8 (2012): 294-308.

10 Regression to the mean is possible, but the variance of the A+ countries' gain scores suggests it's unlikely.

11 Tom Loveless, *The Tracking and Ability Grouping Debate* (Washington, DC: Thomas B. Fordham Institute, July 1, 1998).

12 Samuel R. Lucas, *Tracking Inequality: Stratification and Mobility in American High Schools* (New York: Teachers College Press, 1999).

13 Even Finland and Sweden, famous for egalitarian reforms, divide students for the final two years of secondary school. Germany begins tracking at age 11.

14 Alan Smithers and Pamela Robinson, *Choice and Selection in School Admissions: The Experience of Other Countries*, accessed March 4, 2013, http://suttontrust.com/research/choice-and-selection-in-admissions/1smithers-final-report.pdf.

15 Robert Dreeben and Rebecca Barr, "The Formation and Instruction of Ability Groups," *American Journal of Education* 97, no. 1 (1988): 34-64.

16 See p. 36, Figure 5: James M. McPartland, J. Robert Coldiron, and Jomills H. Braddock II, *School Structures and Classroom Practices in Elementary, Middle, and Secondary Schools*, Report No. 14 (Baltimore: The Johns Hopkins University, 1987).

17 "Success for All—Home", Success for All Foundation, http://www.successforall.org/.

18 Ray C. Rist, "Student Social Class and Teacher Expectations: The Self-fulfilling Prophecy in Ghetto Education," *Harvard Educational Review* 40, no. 3 (1970): 411-451.

19 Ability grouping is called "setting" in Great Britain. Recent reports have been sharply critical of the practice, see: "Setting Harms Education of Some Young Children, Report Warns," *The Independent*, May 16, 2008, http://www.independent.co.uk/news/education/education-news/setting-harms-education-of-some-young-children-report-warns-829312.html.

20 James E. Rosenbaum, *Making Inequality; the Hidden Curriculum of High School Tracking* (New York: John Wiley & Sons, 1976).

21 See: Jeannie Oakes, *Keeping Track: How Schools Structure Inequality* (New Haven: Yale University Press, 1985). Also see: Jeannie Oakes, Amy Stuart Well, and Associates, *Beyond the Technicalities of School Reform: Policy Lessons from Detracking School* (Los Angeles: UCLA Graduate School of Education & Information Studies, 1996).

22 The politics and policies of tracking reform are investigated in: Tom Loveless, *The Tracking Wars: State Reform Meets School Policy* (Washington: Brookings Institution Press, 1999).

23 Mary C. Austin and Coleman Morrison. *The Torch Lighters: Tomorrow's Teachers of Reading* (Cambridge: Harvard University Graduate School of Education, 1961).

24 Rebecca Barr and Robert Dreeben, *How Schools Work* (Chicago, University of Chicago Press, 1983).

25 ECLS asked kindergarten teachers in 1999 the frequency with which they used ability groups in reading. Five response categories, ranging from 0 (never) to 4 (daily). 30% reported never using ability grouping. The average for all teachers was 1.64, indicating about once a week (1 = less than once a week; 2 = once or twice weekly). When the ECLS sample was in 3rd grade, 2001–2002, 50% of teachers employed ability grouping in reading, consistent with the NAEP figure for 4th grade in 2003 (47%). See p. 301, note 6 in Christy Lleras, and Claudia Rangel, "Ability grouping practices in elementary school and African American/Hispanic achievement." *American Journal of Education* 115, no. 2 (2009): 279–304.

26 Barbara Fink Chorzempa and Steve Graham, "Primary-Grade Teachers' Use of Within-Class Ability Grouping in Reading," *Journal of Educational Psychology* 98, no. 3 (2006): 529-541.

27 Tom Loveless, *Tracking, Detracking: High Achievers in Massachusetts Middle School* (Washington, DC: Thomas B. Fordham Institute, 2009).

28 "Research Spotlight on Academic Ability Grouping," NEA, http://www.nea.org/tools/16899.htm.

29 Kelly Puzio and Glenn Colby, *The Effects of Within Class Grouping on Reading Achievement: A Meta-Analytic Synthesis* (Evanston: Society for Research on Educational Effectiveness, 2010).

30 Christy Lleras and Claudia Rangel, "Ability Grouping Practices in Elementary School and African American / Hispanic Achievement," *American Journal of Education* 115, no. 2 (2009): 279.

31 Dana Markow and Michelle Cooper, *The Metlife Survey of the American Teacher: Past, Present and Future* (New York: Metlife, 2008).

32 Background information on the Algebra Project is available at www.algebra.org.

33 Remarks by President Clinton, Education Roundtable, Springbrook High School, Silver Spring, MD, March 16, 1998. Available at http://www.gpo.gov/fdsys/pkg/WCPD-1998-03-23/pdf/WCPD-1998-03-23.pdf.

34 History of California's algebra policy can be found in: *Algebra Policy in California: Great Expectations and Serious Challenges* (Oakland: EdSource, May 2009). Also see Tom Loveless, *The Misplaced Math Student: Lost in Eighth-Grade Algebra* (Washington, DC: The Brookings Institution, 2008).

35 The category "other" received about a 3% response rate before 2000 so the number of students taking more advanced classes was probably very small.

36 Jill Walston and Jill Carlivati McCarroll, *Eighth-Grade Algebra: Findings from the Eighth-Grade Round of the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K)* (Washington, DC: National Center for Education Statistics, October 2010).

37 See David Stevenson, Kathryn S. Schiller, and Barbara Schneider, "Sequences of Opportunities for Learning," *Sociology of Education* 67, no. 3 (1994): 184-198; Adam Gamoran and Eileen C. Hannigan, "Algebra for Everyone? Benefits of College-Preparatory Mathematics for Students with Diverse Abilities in Early Secondary School," *Educational Evaluation and Policy Analysis* 22, no. 3 (2000): 241-254; Julia Smith, "Does an Extra Year Make Any Difference? The Impact of Early Access to Algebra on Longterm Gains in Mathematics Achievement," *Educational Evaluation and Policy Analysis* 18 (1996): 141-153.

38 See Mary Stein, Julia Kaufman, Milan Sherman, and Amy Hillen, "Algebra: A Challenge at the Crossroads of Policy and Practice," *Review of Educational Research* 81, no. 4 (2011): 453-492.

39 Elaine Allensworth, Takako Nomi, Nicholas Montgomery, and Valerie E. Lee, "College Preparatory Curriculum for All: Academic Consequences of Requiring Algebra and English I for Ninth Graders in Chicago," *Educational Evaluation and Policy Analysis* 31, no. 4 (2009): 367-391.

40 Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor, *The Aftermath of Accelerating Algebra Evidence from a District Policy Initiative* (Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, American Institutes for Research, 2012).

41 Jan-Eric Gustafsson, "Understanding Causal Influences on Educational Achievement through Analysis of Differences over Time within Countries," in *Lessons Learned: What International Assessments Tell Us about Math Achievement*, ed. Tom Loveless (Washington: Brookings Institution Press, 2007).

42 Trish Williams, Edward Haertel, and Michael W. Kirst, *Improving Middle Grades Math Performance: A Closer Look at District and School Policies and Practices, Course Placements, and Student Outcomes in California. Follow-Up Analysis* (Mountain View: EdSource, 2011).

43 Jian-Hua Liang, Paul Heckman, and Jamal Abedi, "What Do the California Standards Test Results Reveal About the Movement Towards Eighth-Grade Algebra for All?" *Educational Evaluation and Policy Analysis* 34, no. 3 (2012): 328-343.

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

BROWN CENTER on
Education Policy
at BROOKINGS

BROOKINGS
1775 Massachusetts Avenue, NW • Washington, D.C. 20036
Tel: 202–797–6000 • Fax: 202–797–6004
www.brookings.edu

The Brown Center on Education Policy
Tel: 202–797–6090 • Fax: 202–797–2480
www.brookings.edu/brown