Achievement Growth:

International and U.S. State Trends in Student Performance

Eric A. Hanushek • Paul E. Peterson • Ludger Woessmann

Harvard's Program on Education Policy and Governance & Education Next
Taubman Center for State and Local Government
Harvard Kennedy School

Acknowledgments

We are grateful for the Kern Family Foundation and the Searle Liberty Trust's support for this project.
We are indebted to Kathryn Ciffolillo and Susan Pasternack for careful editorial work,
to Bruce Sanders and Robin Cheung for excellent design work,
to Ashley Inman for table and figure construction,
and to Maura Roche and Antonio M. Wendland for administrative assistance and technical support.
The views expressed here are our own and should not be attributed to any other party
or to the institutions with which we are affiliated.

Websites:
hks.harvard.edu/pepg
educationnext.org

Achievement Growth:

International and U.S. State Trends in Student Performance

by

Eric A. Hanushek

Paul E. Peterson

Ludger Woessmann

Harvard University

Program on Education Policy & Governance

Education Next

Harvard's Program on Education Policy and Governance & Education Next

Taubman Center for State and Local Government

Harvard Kennedy School

PEPG Report No.: 12-03

July 2012

Websites:

hks.harvard.edu/pepg

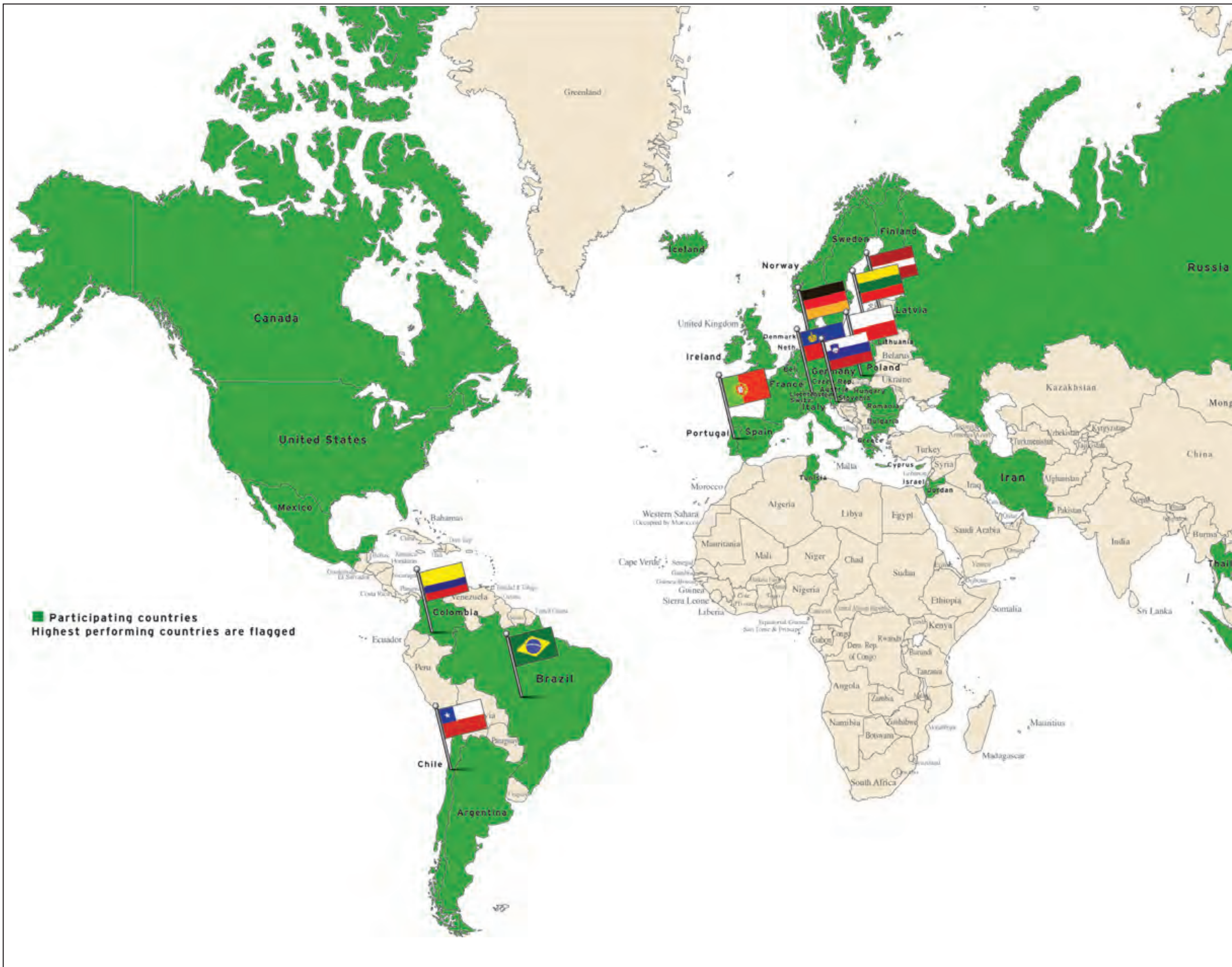
educationnext.org

Table of Contents

Photography:

Cover: Getty Images/Gregor Schuster: background, Anaheim Union High School District: photo
P. vii: Getty Images/KidStock
P. 1: Bettmann/Corbis/AP Images
P. 2: Bob Daugherty/AP Images
P. 3: Getty Images/Stephen Lovekin/Staff
P. 4: Marcy Nighswander/AP Images
P. 7: Colorbox
P. 9: Mark Humphrey/AP Images
P. 10: Getty Images/Jennifer Pottheiser/Contributor
P. 15: Getty Images/Chip Somodevilla/Staff
P. 17: Engel & Gielen
P. 19: Image Source
P. 23: Getty Images/KidStock

Tables and Figures


Greenland
Iceland
Canada
United States
Mexico
Colombia
Brazil
Chile
Argentina
Norway
Sweden
Finland
Russia
Latvia
Lithuania
Poland
Germany
France
Italy
Ireland
Portugal
Spain
Iran
Participating countries
Highest performing countries are flagged

Achievement Growth: International and U.S. State Trends in Student Performance

ERIC A. HANUSHEK
PAUL E. PETERSON
LUDGER WOESSMANN

Executive Summary

"The United States' failure to educate its students leaves them unprepared to compete and threatens the country's ability to thrive in a global economy." Such was the dire warning recently issued by a task force sponsored by the Council on Foreign Relations. Chaired by former New York City schools chancellor Joel I. Klein and former U.S. secretary of state Condoleezza Rice, the task force said that the country "will not be able to keep pace—much less lead—globally unless it moves to fix the problems it has allowed to fester for too long."[i]

The report's views are well supported by the available evidence. In a 2010 report, only 6 percent of U.S. students were found to be performing at the advanced level in mathematics, a percentage lower than those attained by 30 other countries.[ii] Nor is the problem limited to top-performing students. Only 32 percent of 8th-graders in the United States are proficient in mathematics, placing the United States 32nd when ranked among the participating international jurisdictions.[iii]

Although these facts are discouraging, the United States has made substantial additional financial commitments to K–12 education and introduced a variety of school reforms. Have these policies begun to help the United States close the international gap?

While 24 countries trail the U.S. rate of improvement, another 24 countries appear to be improving at a faster rate.

i. Independent Task Force, Council on Foreign Relations (2012).
ii. Hanushek, Peterson, and Woessmann (2010).
iii. Peterson, Woessmann, Hanushek, and Lastra-Anadón (2011).

International Assessment Data

To find out the extent of U.S. progress toward closure of the international education gap, we provide estimates of learning gains over the period between 1995 and 2009 for the United States and 48 other countries from much of the developed and some of the newly developing parts of the world. We also examine changes in student performance in 41 states within the United States between 1992 and 2011, allowing us to compare these states with each other.

Our findings come from assessments of performance in math, science, and reading of representative samples in particular political jurisdictions of students who at the time of testing were in 4th or 8th grade or were roughly ages 9–10 or 14–15.

The data come from one U.S. series of tests and three series of tests administered by international organizations. Using the equating method described in Appendix A, it is possible to link states' performance on the U.S. tests to countries' performance on the international tests, because representative samples of U.S. students have taken all four series of tests.[iv]

Overall Results

The gains within the United States have been middling, not stellar. While 24 countries trail the U.S. rate of improvement, another 24 countries appear to be improving at a faster rate. Nor is U.S. progress sufficiently rapid to allow it to catch up with the leaders of the industrialized world.

In the United States, test-score performance has improved annually at a rate of about 1.6 percent of a standard deviation (std. dev.). Over the 14 years, gains are estimated to be about 22 percent of a std. dev. or the equivalent of about a year's worth of learning. By comparison, students in three countries—Latvia, Chile, and Brazil—improved at an annual rate of 4 percent of a std. dev., and students in another eight countries—Portugal, Hong Kong, Germany, Poland, Liechtenstein, Slovenia, Colombia, and Lithuania—were making gains at twice the rate of students in the United States. Gains made by students in these 11 countries are estimated to be at least two years' worth of learning. Another 13 countries also appeared to be doing better than the U.S.

Student performance in nine countries declined over the same 14-year time period. Test-score declines were registered in Sweden, Bulgaria, Thailand, the Slovak and Czech Republics, Romania, Norway, Ireland, and France. The remaining 15 countries were showing rates of improvement that were somewhat lower than those of the United States.

iv. The four ongoing series are as follows:
1) National Assessment of Educational Progress (NAEP), administered by the U.S. Department of Education;
2) Programme for International Student Assessment (PISA), administered by the Organisation for Economic Co-operation and Development (OECD); 3) Trends in International Mathematics and Science Study (TIMSS), administered by the International Association for the Evaluation of Educational Achievement (IEA); and
4) Progress in International Reading Literacy Study (PIRLS), also administered by IEA.

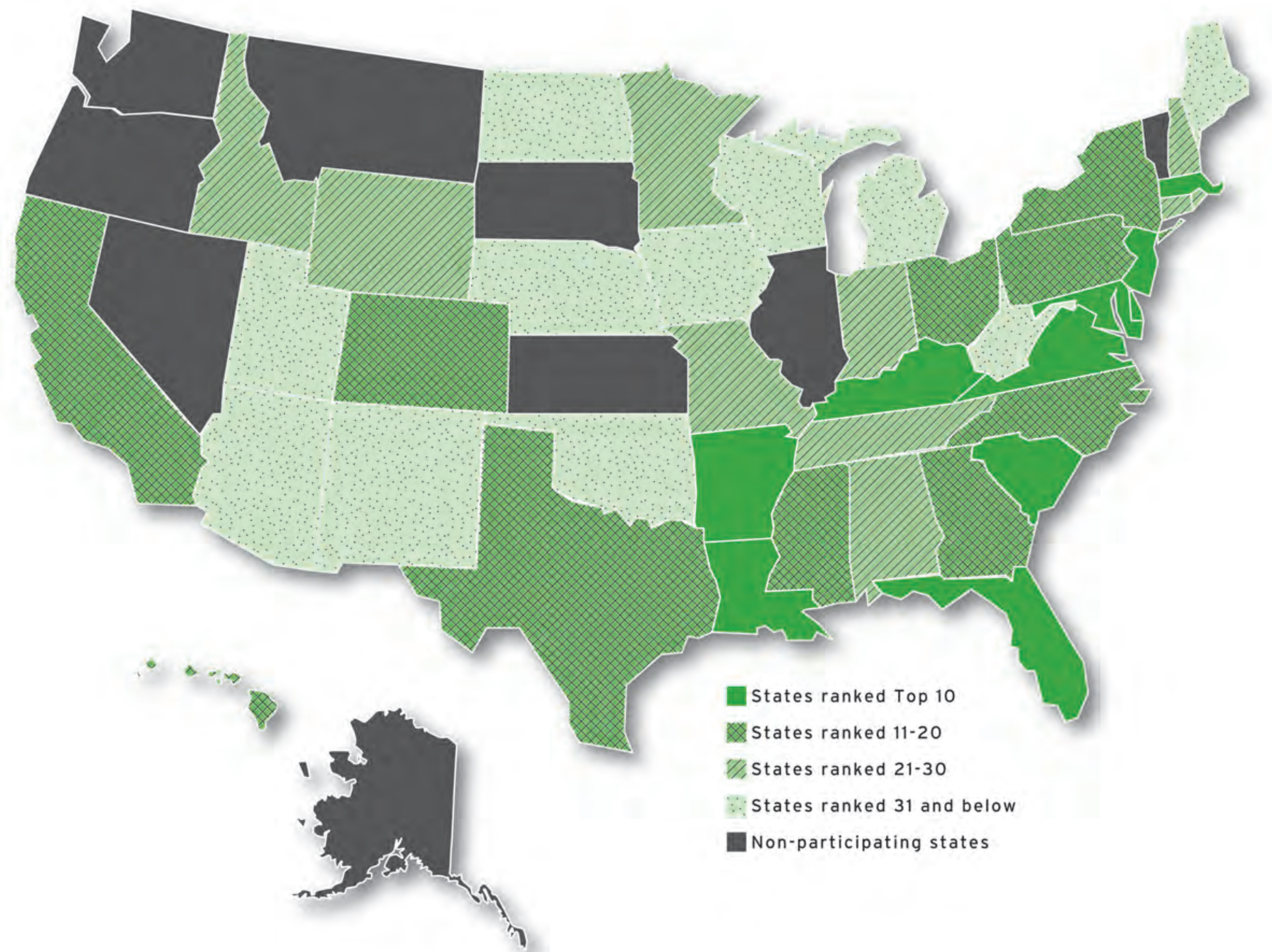
Progress was far from uniform across the United States, however. Indeed, the variation across states was about as large as the variation among the countries of the world. Maryland won the gold medal by having the steepest overall growth trend. Coming close behind, Florida won the silver medal and Delaware the bronze. The other seven states that rank among the top-10 improvers, all of which outpaced the United States as a whole, are Massachusetts, Louisiana, South Carolina, New Jersey, Kentucky, Arkansas, and Virginia.

Iowa shows the slowest rate of improvement. The other four states whose gains were clearly less than those of the United States as a whole, ranked from the bottom, are Maine, Oklahoma, Wisconsin, and Nebraska. Note, however, that because of nonparticipation in the early NAEP assessments, we cannot estimate an improvement trend for the 1992–2011 time period for nine states—Alaska, Illinois, Kansas, Montana, Nevada, Oregon, South Dakota, Vermont, and Washington.

The states making the largest gains are improving at a rate two to three times the rate in states with the smallest gains. States that were further behind in 1992 tend to make larger gains than initially higher-performing states. However, their initial level of performance explains only about a quarter of the variation among the states. Also, variation in state increases in per-pupil expenditure is not significantly correlated with the variation in learning gains.

States with the largest gains in average student performance also tend to see the greatest reduction in the percentage of students performing below the basic level. They also are the ones that experience the largest percent shift of nonproficient students to the level of proficiency set by NAEP. However, there are some exceptions to this overall pattern. At the 8th-grade level, the gains by educationally disadvantaged students in Texas were larger relative to other states, given the percentage of nonproficient students who attained NAEP proficiency. Conversely, nonproficient students in Utah, Nebraska, Pennsylvania, Maine, Wisconsin, and Minnesota were more likely (relative to other states) to cross the proficiency bar, given the gains being made by the most educationally disadvantaged students. Otherwise, an educational tide within a state that lifted an average boat lifted all boats fairly uniformly.

The gains within the United States have been middling, not stellar.


States ranked Top 10
States ranked 11-20
States ranked 21-30
States ranked 31 and below
Non-participating states

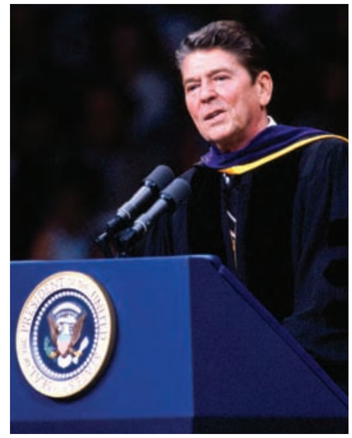
Achievement Growth: International and U.S. State Trends in Student Performance

ERIC A. HANUSHEK
PAUL E. PETERSON
LUDGER WOESSMANN

Introduction

Policymakers in the United States have long recognized that improved education was important to the nation's future.[1] Immediately after the Soviet Union launched the *Sputnik* satellite, the U.S. Congress passed the National Defense Education Act in 1958 to ensure the "security of the Nation" through the "fullest development of the mental resources and technical skills of its young men and women."[2] National security was no less on the mind of a 2012 task force that inquired into the extent to which U.S. schools were competitive with those in other countries. Sponsored by the Council on Foreign Relations and chaired by former New York City schools chancellor Joel I. Klein and former U.S. secretary of state Condoleezza Rice, the task force warned, "The United States' failure to educate its students leaves them unprepared to compete and threatens the country's ability to thrive in a global economy."[3]

In between the 1958 and 2012 proclamations has been a long series of exhortations to restore America's school system to a leading position in the world. Concerns about the quality of U.S. schools intensified in 1983, when a government task force submitted to the Ronald Reagan administration a widely heralded report carrying the title "A Nation at Risk."[4] In 1989, with the calls for improvement continuing, President George H. W. Bush, together with the governors of all 50 states, set goals that would bring U.S. education to the top of world rankings by the year 2000.[5] In his first year in office, in 1993, President Bill Clinton urged passage of the Goals 2000: Educate America Act "so that all Americans can reach internationally competitive standards."[6] Two years later the legislation was enacted into law by a wide, bipartisan congressional majority. When announcing his competitiveness initiative in 2006, President George W. Bush observed that "the bedrock of America's competitiveness is a well-educated and skilled workforce."[7]



"The educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future."

—*A Nation At Risk* Report issued to Ronald Reagan Administration by the National Commission on Educational Excellence, 1983

1. Authors listed alphabetically. Ludger Woessmann took primary responsibility for the analysis of the trends across nations, Eric A. Hanushek took primary responsibility for the analysis of the trends among the U.S. states, and Paul E. Peterson took primary responsibility for overall direction of the project and the preparation of the report.
2. Flattau et al. (2006).
3. Independent Task Force, Council on Foreign Relations (2012).
4. National Commission on Excellence in Education (1983).
5. Peterson (2010), p. 168.
6. President William Clinton, "Message to the Congress Transmitting the 'Goals 2000: Educate America Act,'" April 21, 1993, (http://www.gpo.gov/fdsys/pkg/PPP-1993-book1/pdf/PPP-1993-book1-doc-pg477.pdf). Accessed on June 6, 2012.
7. President George W. Bush, "President's Letter to the Nation Announcing 'American Competitiveness Initiative,'" February 2, 2006, (http://georgewbushwhitehouse.archives.gov/stateoftheunion/2006/aci/index.html). Accessed on June 6, 2012.

President George H. W. Bush set goals that would bring U.S. education to the top of world rankings by the year 2000.

Despite these proclamations, the position of the American school remains problematic when viewed from an international perspective. In a report issued in 2010, we found only 6 percent of U.S. students performing at the advanced level in mathematics, a percentage lower than those attained by 30 other countries.[8] And the problem is not limited to top-performing students. In 2011, we showed that 32 percent of 8th-graders in the United States were proficient in mathematics, placing the United States 32nd when ranked among the participating international jurisdictions.[9]

Nor is the public unaware of the situation. When a cross section of the American public was asked how well the United States was doing in math, as compared to other industrialized countries, the average estimate placed the United States at the 18th rank, only modestly better than its actual standing.[10] Americans do not find it difficult to agree with the summary words of the Council on Foreign Relations task force report: "Overall, U.S. educational outcomes are unacceptably low." [11]

In this report, we inquire as to whether there is evidence that the educational situation in the United States has improved. We ask the simple question: "Is the United States beginning to do better?"

American governments at every level have taken education-related actions that would seem to be highly promising. Federal, state, and local governments spent 35 percent more per pupil—in real-dollar terms—in 2009 than they had in 1990.[12] States began holding schools accountable for student performance in the 1990s, and the federal government developed its own nationwide school-accountability program in 2002.[13]

And, in fact, U.S. students in elementary school do seem to be performing considerably better than they were a couple of decades ago. Most notably, the performance of 4th-grade students on math tests rose steeply between the mid-1990s and 2011. Perhaps, then, after a half century of concern and efforts, the United States may finally be taking the steps needed to catch up.

To find out whether the United States is narrowing the international education gap, this report provides estimates of learning gains over the period between 1995 and 2009 for 49 countries from most of the developed and some of the newly developing parts of the world. We also examine changes in student performance in 41 states within the United States, allowing us to compare these states with each other.

Data and Analytic Approach

Data availability varies from one international jurisdiction to another, but for many countries enough information is available to provide estimates of

8. Hanushek, Peterson, and Woessmann (2010).
9. Peterson, Woessmann, Hanushek, and Lastra-Anadón (2011).
10. Howell, Peterson, and West (2009).
11. Independent Task Force, Council on Foreign Relations (2012), p. 19.
12. National Center for Educational Statistics (2011).
13. Peterson (2010), ch. 8.

change for the 14-year period between 1995 and 2009. For 41 U.S. states, one can estimate the improvement trend for a 19-year period—from 1992 to 2011. Those time frames are extensive enough to provide a reasonable estimate of the pace at which student test-score performance is improving in countries across the globe and within the United States.

Our findings come from assessments of performances in math, science, and reading of representative samples in particular political jurisdictions of students who at the time of testing were in 4th or 8th grade or were roughly ages 9–10 or 14–15. The political jurisdictions may be nations, states, or other subnational units. The data come from one U.S. series of tests and three series of tests administered by international organizations. Using the equating and estimation methods described in Appendix A, it is possible to link states' performance on the U.S. tests to countries' performance on the international tests, because representative samples of U.S. students have taken all four series of tests.[14]

Our international results are based on 28 administrations of comparable math, science, and reading tests between 1995 and 2009 to jurisdictionally representative samples of students in 49 countries and 4 subordinate jurisdictions. Our state-by-state results come from 36 administrations of math, reading, and science tests between 1992 and 2011 to representative samples of students in 41 U.S. states. These tests are part of four ongoing series: 1) National Assessment of Educational Progress (NAEP), administered by the U.S. Department of Education; 2) Programme for International Student Assessment (PISA), administered by the Organisation for Economic Co-operation and Development (OECD); 3) Trends in International Mathematics and Science Study (TIMSS), administered by the International Association for the Evaluation of Educational Achievement (IEA); and 4) Progress in International Reading Literacy Study (PIRLS), also administered by IEA.

Comparisons across Countries

Let us first consider in absolute terms the overall gains on NAEP that provide the benchmark against which every state and all foreign jurisdictions are compared. Americans will be pleased to learn that the performance of U.S. students in 4th and 8th grade in math, reading, and science improved noticeably between 1995 and 2009. Using information from all administrations of NAEP tests to students in all three subjects over this time period, we observe that student achievement in the United States is estimated to have increased by 1.6 percent of a standard deviation (std. dev.) per year, on average. Over the 14 years, these gains equate to 22 percent of a std. dev. When interpreted in years of schooling, these gains are notable. On

"Measured against global standards, far too many U.S. schools are failing to teach students the academic skills and knowledge they need to compete and succeed."

— Independent Task Force Report, p. 3, Joel Klein, co-chair, Council on Foreign Relations

14. Other, less comprehensive estimations of trends in student performance across nations include the following: OECD (2010); Martin, Mullis, and Foy (2008); Mullis, Martin, Kennedy, and Foy (2007); Mourshed, Chijioke, and Barber (2010); OECD (2011).

In his first year in office in 1993, President Bill Clinton urged passage of the Goals 2000: Educate America Act "so that all Americans can reach internationally competitive standards."

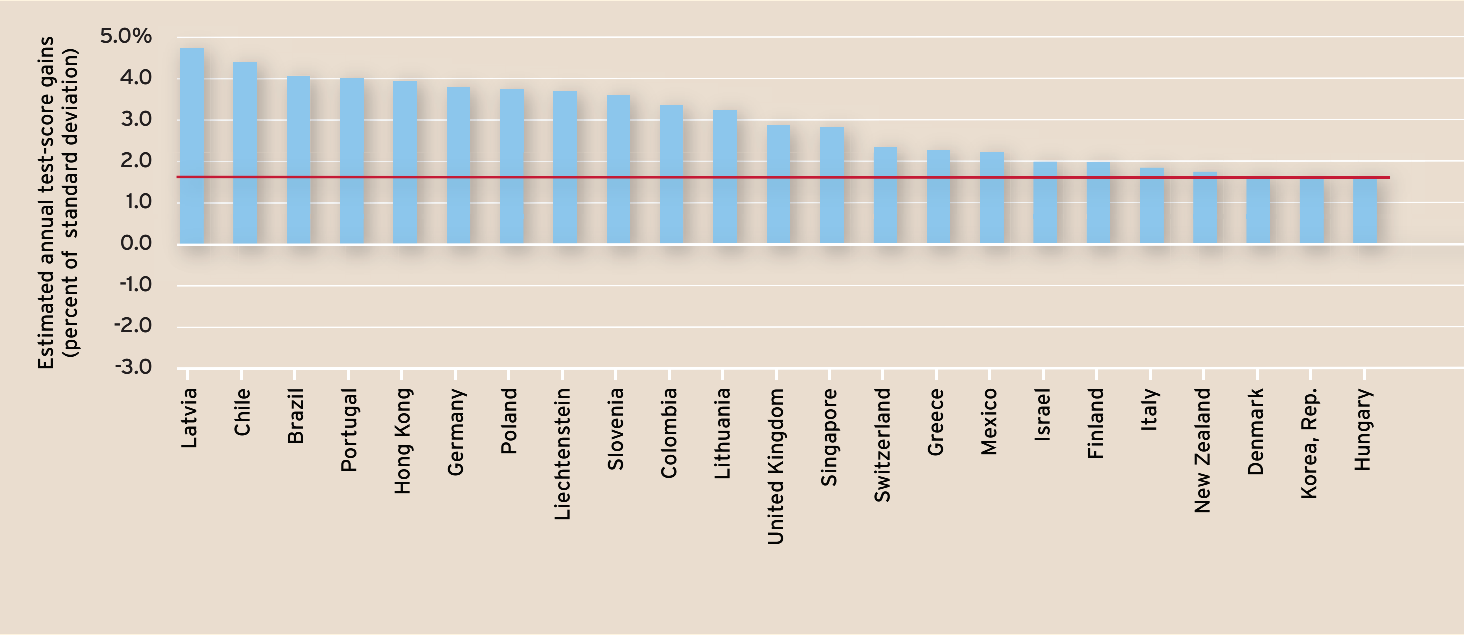
most measures of student performance, student growth is typically about 1 full std. dev. on standardized tests between 4th and 8th grade, or about 25 percent of a std. dev. from one grade to the next. Taking that as the benchmark, we can say that the rate of gain over the 14 years has been just short of the equivalent of one additional year's worth of learning among students in their middle years of schooling.

Yet when compared to gains made by students in other countries, the progress gains within the United States are shown to be middling, not stellar (see Figure 1 and Table B.1). While 24 countries trail the U.S. rate of improvement, another 24 countries appear to be improving at a faster rate. Nor is U.S. progress sufficiently rapid to allow it to catch up with the leaders of the industrialized world.

Students in three countries—Latvia, Chile, and Brazil—improved at an annual rate of 4 percent of a std. dev., and students in another eight countries—Portugal, Hong Kong, Germany, Poland, Liechtenstein, Slovenia, Colombia, and Lithuania—were making gains at twice the rate of students in the United States. By the previous rule of thumb, gains made by students in these 11 countries are estimated to be at least two years' worth of learning. Another 13 countries also appeared to be doing better than the U.S.

Student performance in nine countries declined over the same 14-year time period. Test-score declines were registered in Sweden, Bulgaria, Thailand, the Slovak and Czech Republics, Romania, Norway, Ireland, and France. The

Figure 1. Overall annual rate of growth in student achievement in math, reading, and science in



Note: See Table B.1 for numerical values remaining 15 countries were showing rates of improvement that were somewhat but not significantly lower than those of the United States.

In sum, the gains posted by the United States in recent years are hardly remarkable by world standards. Although the U.S. is not among the 9 countries that were losing ground over this period of time, 11 other countries were moving forward at better than twice the pace of the United States, and all the other participating countries were changing at a rate similar enough to the United States to be within a range too close to be identified as clearly different.

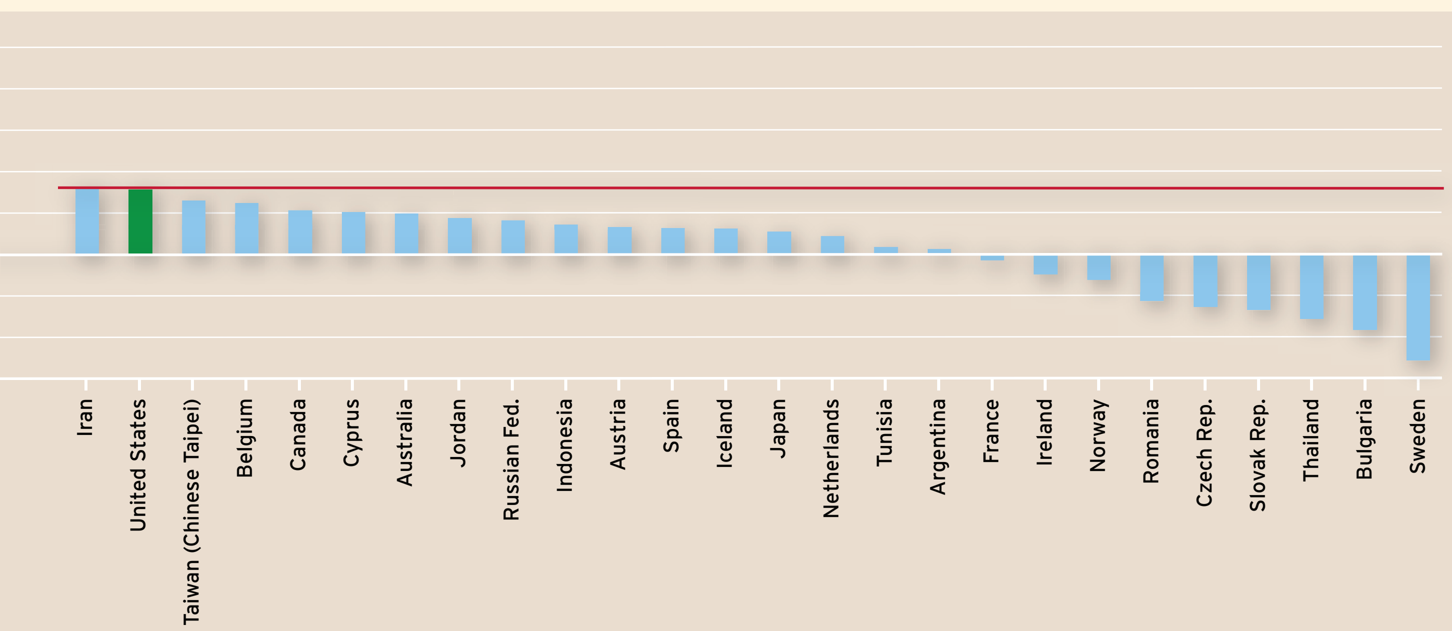
While 24 countries trail the U.S. rate of improvement, another 24 countries appear to be improving at a faster rate.

Comparisons among States

Progress was far from uniform across the United States. Indeed, the variation across states was about as large as the variation among the countries of the world. Maryland won the gold medal by having the steepest overall growth trend. Coming close behind, Florida won the silver medal and Delaware the bronze. The other seven states that rank among the top-10 improvers, all of which outpaced the United States as a whole, are Massachusetts, Louisiana, South Carolina, New Jersey, Kentucky, Arkansas, and Virginia. See Figure 2 for an ordering of the 41 states by rate of improvement.

Iowa shows the slowest rate of improvement. The other four states whose gains were clearly less than those of the United States as a whole, ranked from

49 countries, 1995–2009



The variation across states was about as large as the variation among the countries of the world.

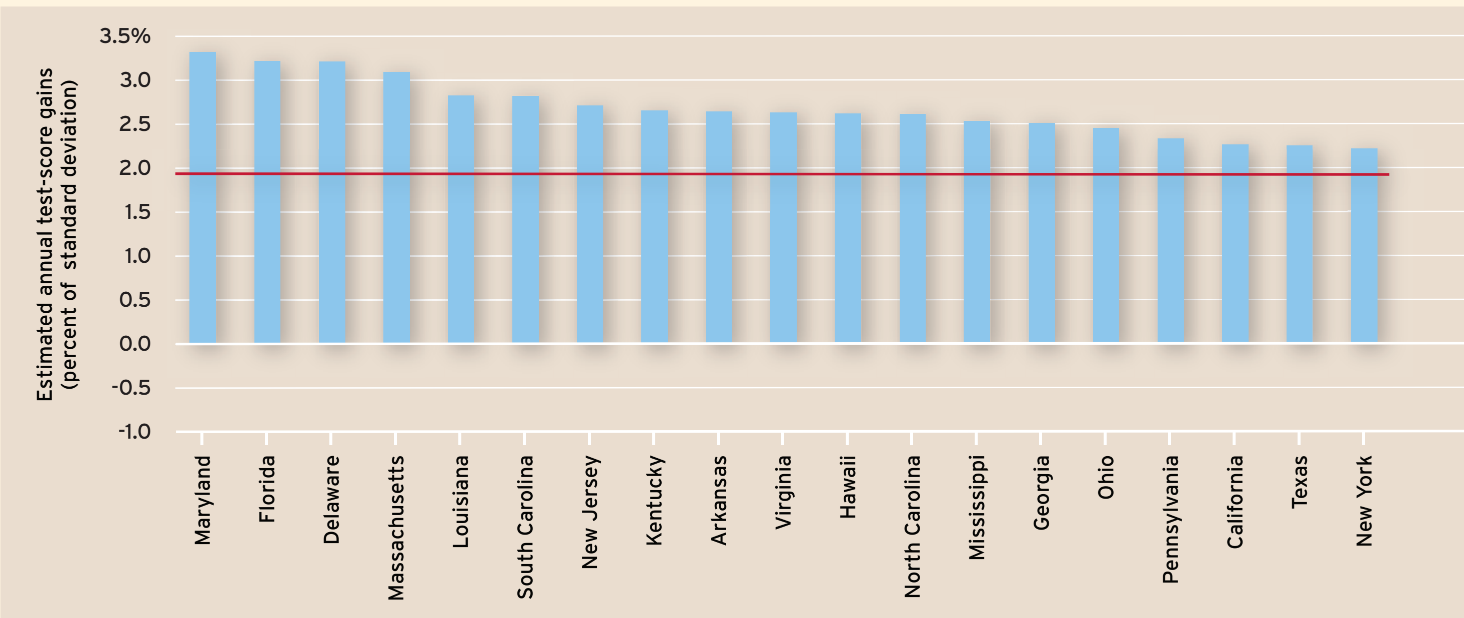
the bottom, are Maine, Oklahoma, Wisconsin, and Nebraska. Note, however, that because of nonparticipation in the early NAEP assessments, we cannot estimate an improvement trend for the 1992–2011 time period for nine states—Alaska, Illinois, Kansas, Montana, Nevada, Oregon, South Dakota, Vermont, and Washington.[15]

Cumulative growth rates vary widely. Average student gains over the 19-year period in Maryland, Florida, Delaware, and Massachusetts, with annual growth rates of 3.1 to 3.3 percent of a std. dev., yielded gains of some 59 percent to 63 percent of a std. dev. over the entire time period, or better than two years of additional learning. Meanwhile, annual gains in the states with the weakest growth rates—Iowa, Maine, Oklahoma, and Wisconsin—varied between 0.7 percent and 1.0 percent of a std. dev., which translate over the 19-year period into learning gains of one-half to three-quarters of a year. In other words, the states making the largest gains are improving at a rate two to three times the rate in states with the smallest gains.

Had all students throughout the country made the same average gains as those in the four leading states, the United States would have been making progress roughly comparable to the rate of improvement in Germany and the United Kingdom, bringing the United States reasonably close to the top-performing countries in the world.

15. After the 2002 federal law, No Child Left Behind, mandated NAEP testing in every state, these nine states participated in NAEP. Between 2003 and 2011, the annual gains in std. dev. were as follows: Nevada, 2.94; Montana, 2.06; Vermont, 1.93; Illinois, 1.92; Kansas, 1.43; Washington, 1.30; Alaska, 0.83; South Dakota, 0.81; and Oregon, 0.32. Five of the nine states performed below the national gains during this period, which was 1.85 std. dev.

Figure 2. Annual rate of growth in student achievement in math, reading, and science in 41 U.S.



Note: Outcomes are available for only four subnational jurisdictions outside the United States—England and Scotland in the United Kingdom and Ontario and Quebec in Canada

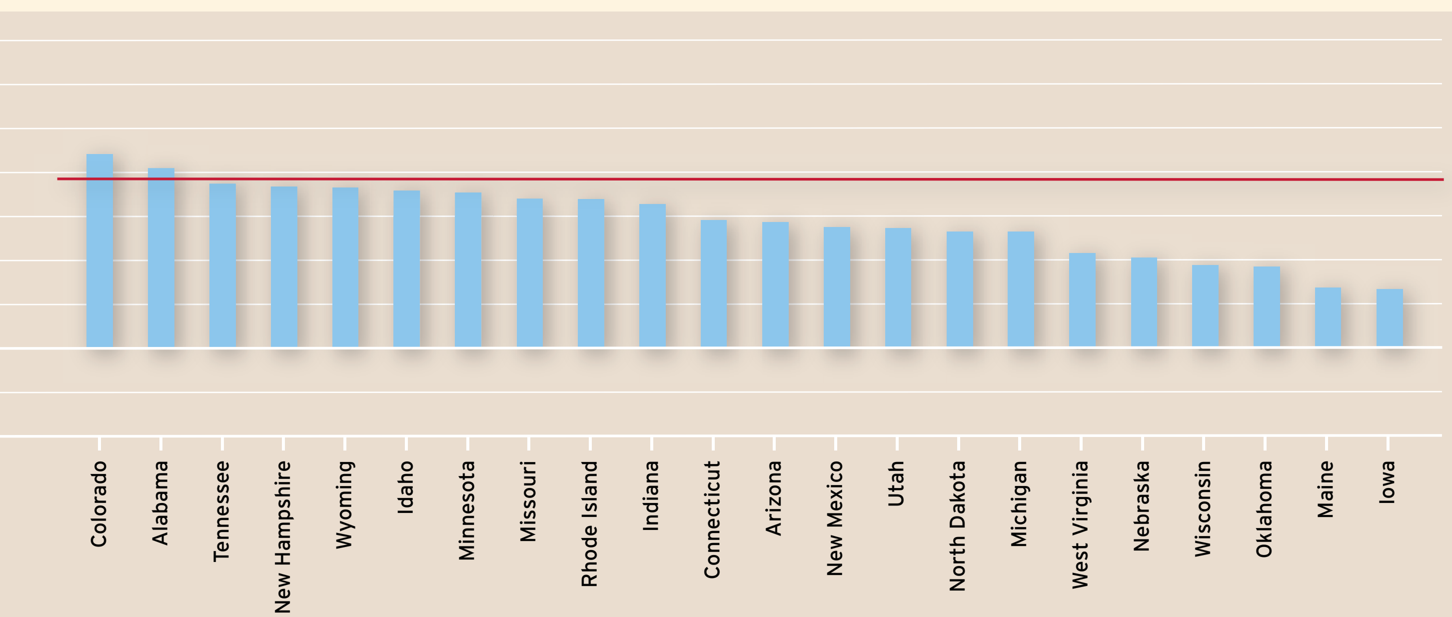
Gains by Low-Performing and High-Performing Students

NAEP has set three benchmarks for student performance—advanced, proficient, and basic. According to these standards, very few U.S. students are performing at the advanced level and a clear majority of students score at a level below that which the NAEP governing board deems is necessary to demonstrate math proficiency. However, a substantial majority of students have what NAEP regards as basic mathematics knowledge. (See sidebar for NAEP definitions of the 8th-grade basic level and 8th-grade proficiency and examples of the kinds of questions 8th-grade students are expected to be able to answer.) Among 4th-graders, 7 percent performed at or above the advanced level, 40 percent at or above the proficiency level, and 82 percent at or above the basic level. By 8th grade, these percentages had slipped. Although 8 percent were performing at or above the advanced level, only 35 percent were scoring above the proficiency bar, while 73 percent were performing at or above the basic level.

In this section we first report the success of states at reducing the percentage of students performing below the basic level. If the percentage of students scoring below basic in state A is reduced from 20 percent to 10 percent, the state is identified as having reduced low performance by 50 percent. If in state B the percentage of students below basic is reduced from 50 percent to 25 percent, it, too, is identified as having reduced low performance by 50 percent.

Maryland won the gold medal by having the steepest overall growth trend. Coming close behind, Florida won the silver medal and Delaware the bronze.

states, 1992–2011



Examples and Definitions of Basic and Proficient Performance on National Assessment of Educational Progress

Basic level example question

Basic level students in 8th grade are expected to answer questions like the following:

1. A recipe requires $1\frac{1}{3}$ cups of sugar. Which of the following ways describes how the measuring cups shown can be used to measure $1\frac{1}{3}$ cups of sugar accurately?

A. Use the $\frac{1}{2}$ cup three times.
B. Use the $\frac{1}{4}$ cup three times.
C. Use the $\frac{1}{2}$ cup twice and the $\frac{1}{3}$ cup once.
D. Use the $\frac{1}{3}$ cup twice and the $\frac{1}{2}$ cup once.
E. Use the $\frac{1}{4}$ cup once, the $\frac{1}{3}$ cup once, and the $\frac{1}{2}$ cup once.

If you chose C from the list of five choices, you are in the company of the 80 percent of U.S. 8th graders from the Class of 2011 who answered correctly.

Proficient level example question

Proficient students in 8th grade are expected to answer questions like the following:

"Three tennis balls are to be stacked one on top of another in a cylindrical can. The radius of each tennis ball is 3 centimeters. To the nearest whole centimeter, what should be the minimum height of the can? Explain why you chose the height that you did. Your explanation should include a diagram."

If you chose 18 cm from the list of five choices, you are in the company of the 28 percent of U.S. 8th graders from the Class of 2011 who answered correctly.[i]

Definition of basic level of performance in math at the 8th grade

Eighth-graders performing at the basic level should complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs. They should be able to solve problems in all NAEP content areas through the appropriate selection and use of strategies and technological tools—including calculators, computers, and geometric shapes. Students at this level also should be able to use fundamental algebraic and informal geometric concepts in problem solving.

Definition of proficient level of performance in math at 8th grade

Eighth-graders performing at the proficient level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections between fractions, percents, decimals, and other mathematical topics such as algebra and functions.... Quantity and spatial relationships in problem solving and reasoning should be familiar to them, and they should be able to convey underlying reasoning skills beyond the level of arithmetic.... These students should make inferences from data and graphs, apply properties of informal geometry, and accurately use the tools of technology. Students at this level should...be able to calculate, evaluate, and communicate results within the domain of statistics and probability.[ii]

i. Questions come from NAEP's online past questions database, http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=mathematics. Accessed June 14, 2012.
ii. NAEP's definitions of the different levels of math achievement. http://nces.ed.gov/nationsreportcard/mathematics/achieveall.asp. Accessed June 14, 2012.

Secondly, we rate each state's success in lifting the percentage of nonproficient students across the proficiency bar. If the percentage of students identified as nonproficient in state C declines from 50 percent to 25 percent, state C is identified as having halved the percentage of nonproficient students. If the decrease is from 30 percent to 15 percent in state D, it too, is identified as having halved the percentage of students who were nonproficient.

It is important to understand that the NAEP definition of proficiency used here is different from the one set by each state under No Child Left Behind, the federal law passed in 2002, which asked each state to take steps to ensure that adequate progress was being made each year so that all students would be proficient by 2014. That law allowed each state to set its own proficiency standard, and as a result, state proficiency standards have varied widely.[16] In 2009 only five states—Massachusetts, Missouri, Washington, Hawaii, and New Mexico—set their proficiency standards at levels roughly equivalent to the NAEP level of proficiency.[17] Meanwhile, Tennessee, Nebraska, Alabama, and Michigan, the states with the lowest proficiency standards, set them closer to the NAEP basic level.

Since states set very different proficiency standards, it is possible they also focused their attention on different segments of the student population. Some may have concentrated on enhancing the performance of those who had not attained the NAEP basic level, while others may have focused on those close to the NAEP proficiency line.

As mentioned, we chart both the reduction in the percentage of students performing in math below NAEP's basic level and the percentage of students brought across the NAEP math proficiency bar. We examine gains in mathematics, because that is where students have made the largest advances during this period of time.[18] The percent reduction in the percentage of students below basic and the percent reduction in the percentage of nonproficient students who achieved NAEP proficiency are reported for each state in Table B.2. We also include in this table the overall trend in average scores discussed in an earlier section of this report.

The states that made the largest average gains tend to be the same states that did the most to reduce the percentage performing below the basic level. Where good things were happening on average they were also happening for the most educationally disadvantaged. A similar connection between average scores and the percentage crossing the proficiency bar exists, but as we shall see, it is not quite as strong.

We show in Figure 3 the pattern in 4th-grade mathematics. Among this group of students the correlation between trends in average scores across states

When announcing his competitiveness initiative in 2006, President George W. Bush observed that "the bedrock of America's competitiveness is a well-educated and skilled workforce."

16. The assessment language introduces some confusion. NAEP sets performance standards that it calls proficient but different (usually lower) definitions of proficiency are applied across the states to meet the requirements of NCLB. Peterson and Lastra-Anadón(2010); Bandeira de Mello, Blankenship, and McLaughlin (2009); for international equivalents of NAEP proficiency standards, see Peterson, Woessmann, Hanushek, and Lastra-Anadón (2011).

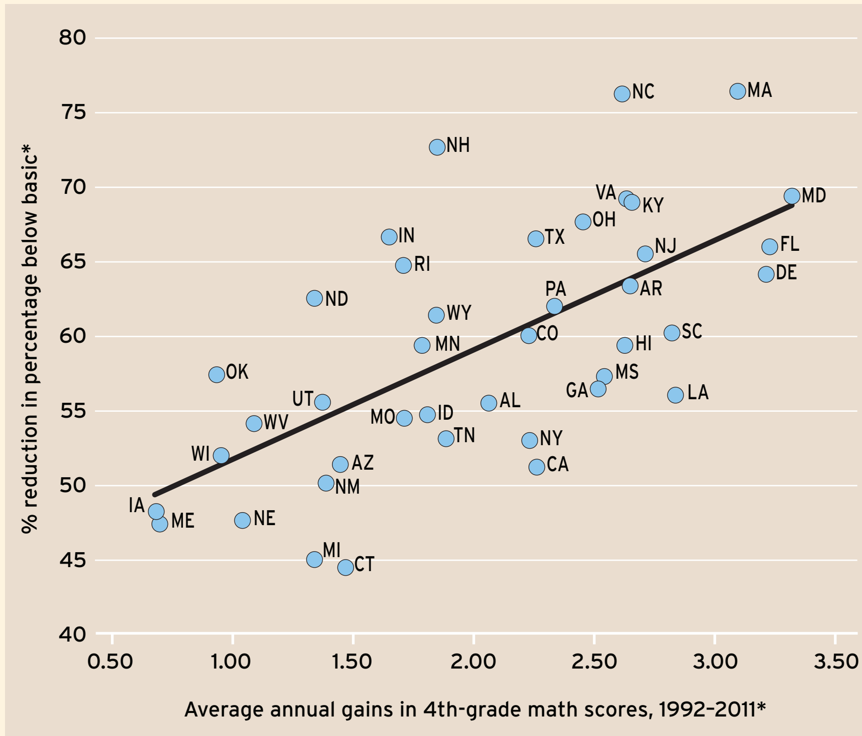
17. Peterson and Lastra-Anadón (2010).

18. Below note a relationship between initial average test scores and subsequent growth. Those who were furthest behind initially made the most progress. But we do not find a consistent pattern for percent reduction in math performance at the basic level or percent increase in proficiency on the NAEP. The pattern is weak overall and inconsistent between 4th and 8th grade.

"Human capital will determine power in the current century, and the failure to produce that capital will undermine America's security."

— Independent Task Force Report, p. 4, Condoleezza Rice, co-chair, Council on Foreign Relations

Figure 3. Relationship between gains in state average scores and percent reduction in percentage performing below basic level in math in 4th grade on NAEP

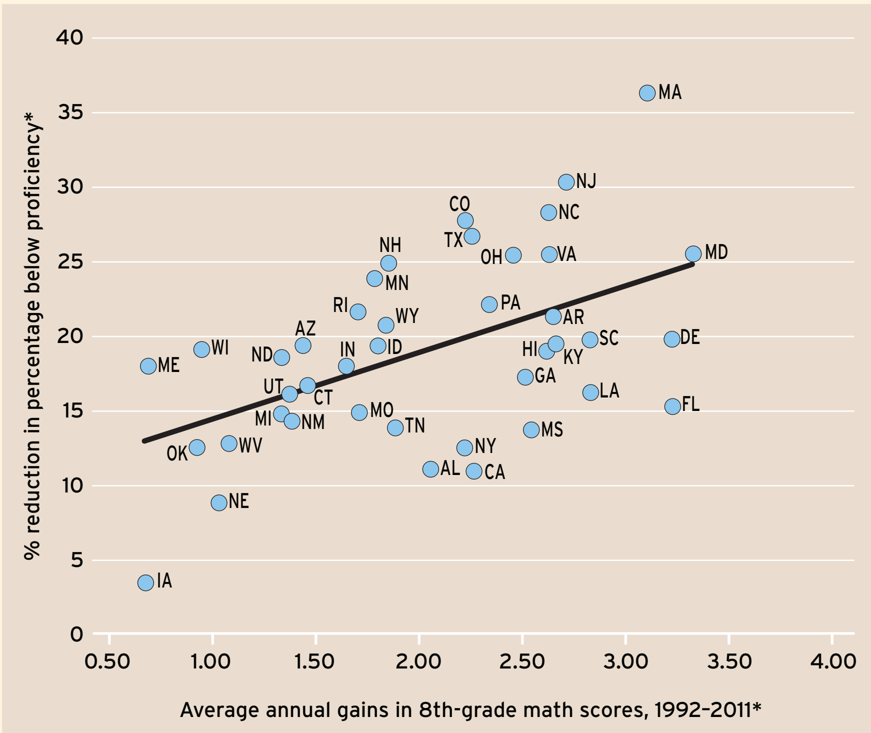


* see Table B.2 for numerical value

and trends in the percent reduction in those performing below basic in math is a solid 0.65. The steepness of the regression line in Figure 3 identifies the strength of the relationship. In those states where the positive trend in 4th-grade math performance was greater, so in many cases was the percent reduction in those performing below the basic level.

Figure 4 shows the weaker connection between average gains in a state and the percentage of nonproficient 8th-grade students moving above the math proficiency bar (correlation = 0.33). Massachusetts, New Hampshire, and North Carolina all enjoyed comparatively large increases in overall performance and a substantial shift in the percentage of nonproficient students brought up to NAEP proficiency levels. But the figure shows that there were many exceptions to this pattern, as quite a number of the observations stray a considerable distance from the regression line. Not every

Figure 4. Relationship between gains in state average scores and percent reduction below proficiency in 8th grade on NAEP

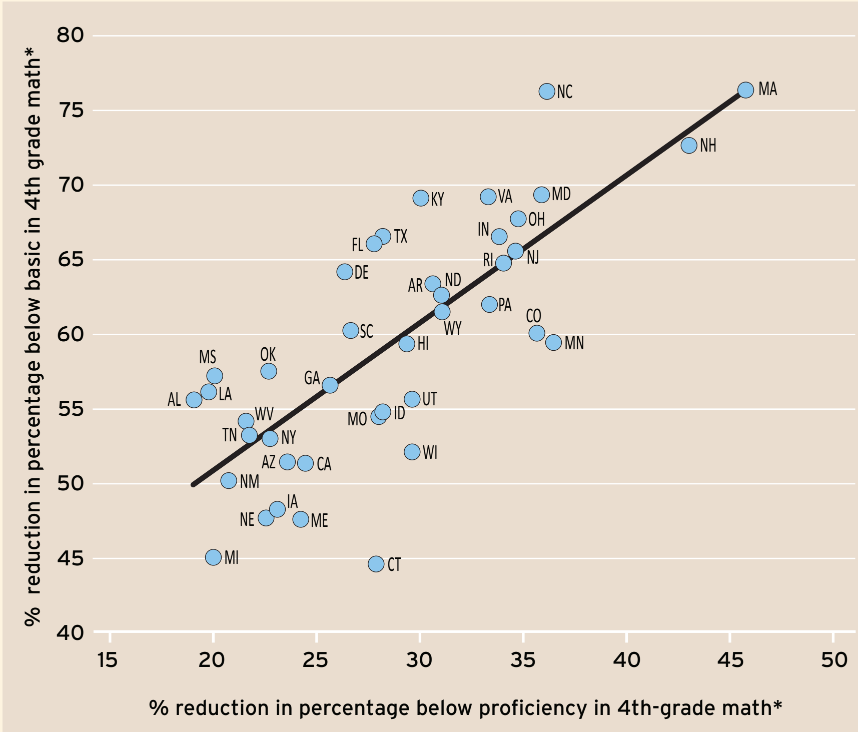


* see Table B.2 for numerical value

state's record with the average student translates into an equivalent shift across the proficiency line.

In Figure 5 the relationship between gains in 4th-grade math by low-performing and high-performing students is directly compared. The steep regression line (correlation = 0.77) shows that gains for higher-performing students do not in general come at the expense of the educationally disadvantaged. Those states that experience the greatest reduction in the number of students performing below the basic level also see the largest percentage shift across the NAEP proficiency bar. Yet some states depart from this general pattern in one direction or another. On the one side, one finds in Kentucky, Texas, Florida, and Delaware a much larger reduction (relative to other states) in the percentage of students performing below basic, given their increment in the percentages of nonproficient students attaining

Figure 5. Relationship between percent reduction in percentage of students in state performing below basic and below proficiency in math at 4th grade on NAEP

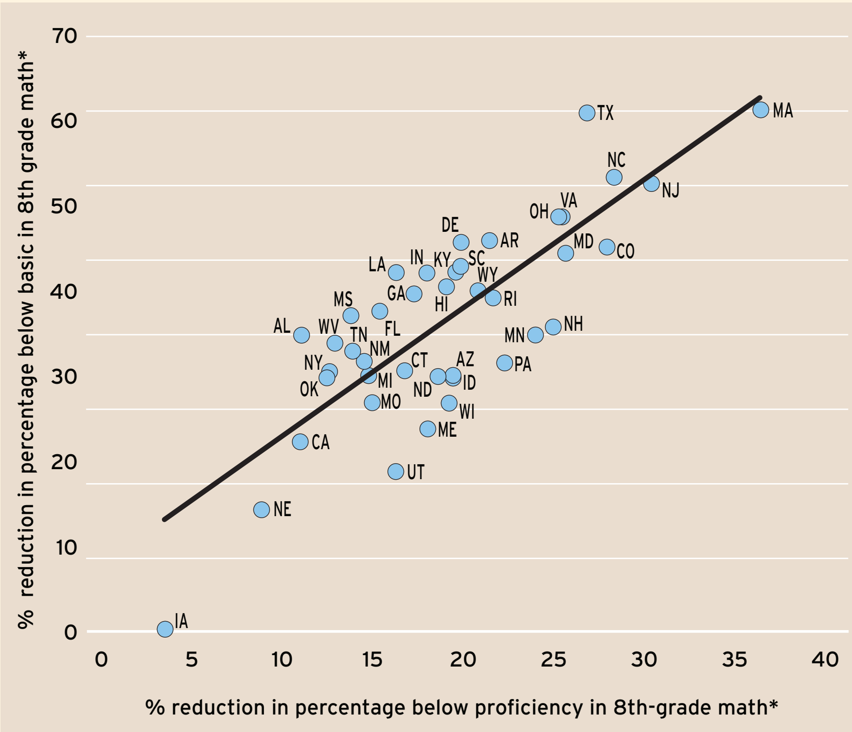


* see Table B.2 for numerical value

NAEP proficiency. On the other side, Connecticut, Maine, Wisconsin, and Minnesota witnessed a relatively large shift in the percentage of students crossing the NAEP proficiency bar, given the reduction in the percentage of students performing below basic in the state.

This same direct comparison is shown in Figure 6, this time for 8th-graders. Once again, it can be seen that states that see a relatively large percentage crossing the proficiency bar also enjoy a relatively large reduction in students performing below the basic level (correlation = 0.81). Yet a few states deviate from the general pattern. Once again, it is Texas that sees a bigger reduction (relative to other states) in the percentage of those performing below the basic level, given the percentages crossing the math proficiency line. Conversely, Utah, Nebraska, Pennsylvania, Maine, Wisconsin, and Minnesota (relative to other states) were seeing a relatively large number of students becoming proficient

Figure 6. Relationship between percent reduction in percentage of students in state performing below basic and below proficiency in math at 8th grade on NAEP



* see Table B.2 for numerical value

States can—and do—work at "leaving no child behind" and yet at the same time see an increment in the percentage of nonproficient students rising to a level of NAEP proficiency.

at the NAEP level, given the amount of reduction in below basic performance among the educationally disadvantaged.

It is this 8th-grade relationship displayed in Figure 6 that is particularly meaningful, as it shows student readiness for high school. The data demonstrate rather clearly that most states, if they make gains, do so across the board—for higher- and lower-performing students alike. States can—and do—work at "leaving no child behind" and yet at the same time see an increment in the percentage of nonproficient students rising to a level of NAEP proficiency. States in which the educationally disadvantaged are gaining the most ground are the ones where higher-performing students are doing the same, and vice versa.

In short, what is happening on average in each state is, more often than not, happening to both those who are higher performing and those who are the most challenged. For that reason we focus in the remainder of the report on the factors affecting the variation in average performance among the states.

Is the South Rising Again?

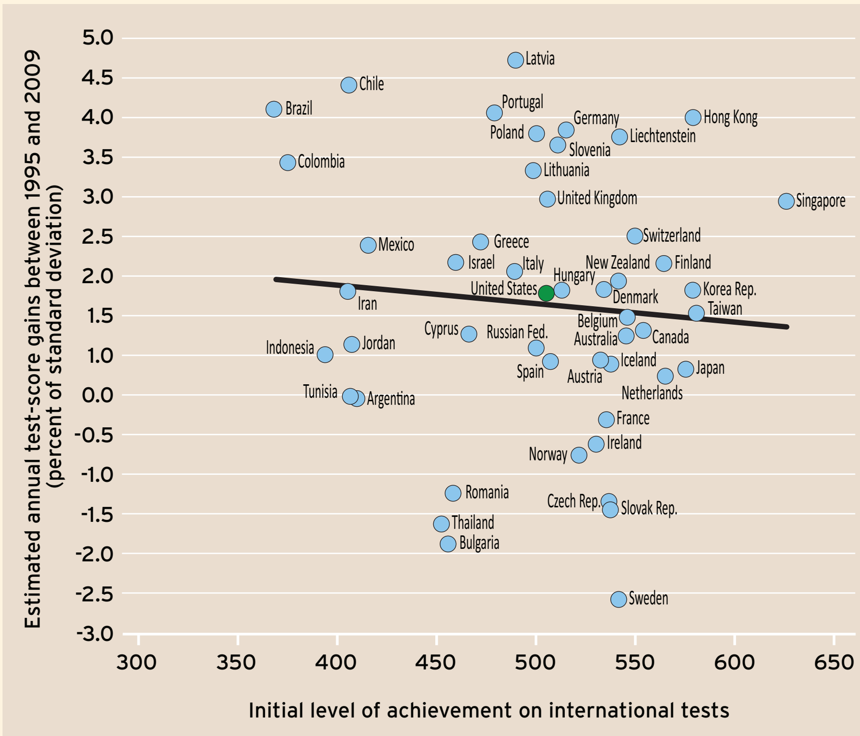
Some regional concentration is evident. Five of the top-10 states were in the South, while no southern states were among the 18 with the slowest growth. The strong showing of the South may be related to energetic political efforts to enhance school quality in that region. During the 1990s, governors of several southern states—Tennessee, North Carolina, Florida, Texas, and Arkansas—provided much of the national leadership for the school accountability effort, as there was a widespread sentiment in the wake of the civil rights movement that steps had to be taken to equalize educational opportunity across racial groups. The results of our study suggest those efforts were at least partially successful.

Meanwhile, students in Wisconsin, Michigan, Minnesota, and Indiana were among those making the smallest average gains between 1992 and 2011. Once again, the larger political climate may have affected the progress on the ground. Unlike in the South, the reform movement has made little headway within midwestern states, at least until very recently. Many of the midwestern states had proud education histories symbolized by internationally acclaimed land-grant universities, which have become the pride of East Lansing, Michigan; Madison, Wisconsin; St. Paul, Minnesota; and Lafayette, Indiana. Satisfaction with past accomplishments may have dampened interest in the school reform agenda sweeping through southern, border, and some western states.

Are Gains Simply Catch-ups?

According to a perspective that we shall label "catch-up theory," growth in student performance is easier for those political jurisdictions originally performing at a low level than for those originally performing at higher levels. Lower-performing systems may be able to copy existing approaches at lower cost than higher-performing systems can innovate. This would lead to a convergence in performance over time. An opposing perspective—which we shall label "building-on-strength theory"—posits that high-performing school systems find it relatively easy to build on their past achievements, while low-performing systems may struggle to acquire the human capital needed to improve. If that is generally the case, then the education gap among nations and among states should steadily widen over time.

Figure 7. Relationship between a country's initial level of student achievement and its growth rate, 1995–2009



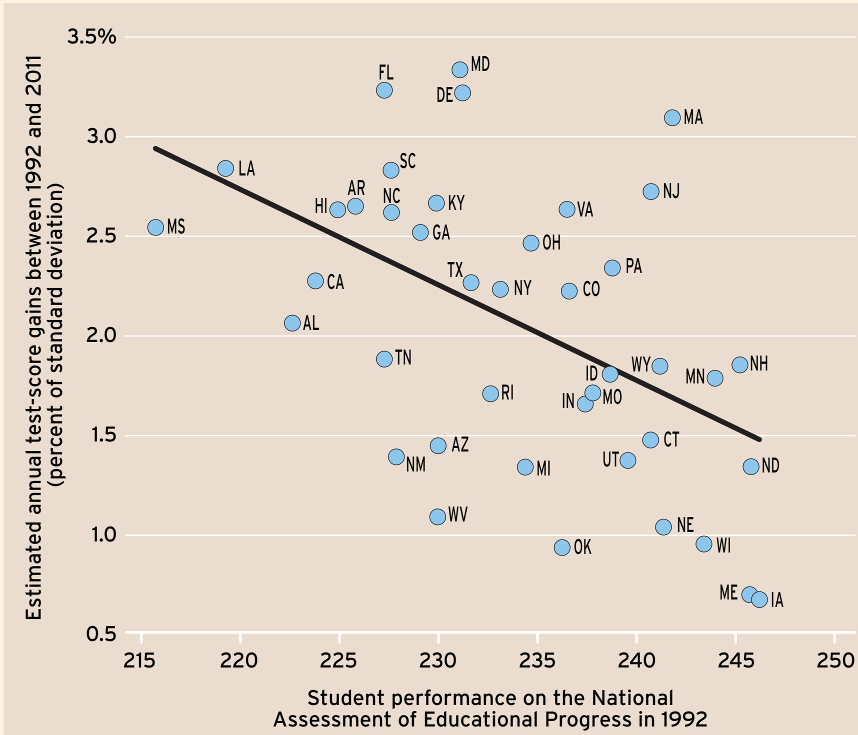
"We know what it takes to compete for the jobs and industries of our time. We need to out-innovate, out-educate, and out-build the rest of the world."

—President Barack Obama
"State of the Union Address,"
January 25, 2011

Neither theory seems able to predict the international test-score changes that we have seen, as nations with rapid gains can be identified among both countries that had high initial scores and countries that had low ones. Latvia, Chile, and Brazil, for example, were initially low-ranking countries in 1995 that made rapid gains, a pattern that supports catch-up theory. But consistent with building-on-strength theory, a number of countries that have advanced relatively rapidly were initially high-performing countries—Hong Kong and Germany, for example. Overall, there is no significant pattern between original performance and changes in performance across countries (see Figure 7).

But if neither theory accounts for differences across countries, catch-up theory may help to explain variation among the U.S. states. The correlation between initial performance and rate of growth is a negative 0.58; states starting with lower initial scores tend to have larger gains. For example, students in Mississippi and

Figure 8. Relationship between a state's initial level of student achievement and its growth rate, 1992–2011



Louisiana, originally among the lowest scoring, showed some of the most striking improvement. Meanwhile, Iowa and Maine, two of the highest-performing entities in 1992, were among the laggards in subsequent years (see Figure 8). In other words, catch-up theory partially characterizes the pattern of change within the United States, probably because the barriers to the adoption of existing technologies are much lower within a single country than across national boundaries.

Of course, catch-up theory, even if it were perfectly predictive of future growth, would not provide much in the way of policy guidance. More importantly, it explains only about one-quarter of the total state variation in achievement growth. Notice in Figure 8 that some states—for instance, Maryland, Massachusetts, Delaware, and Florida—score well above the line in the figure that displays the variation explained by catch-up theory. Note also that Iowa, Maine, Wisconsin, and Nebraska rank well below that line. Closing the interstate gap does not happen automatically.

What About Spending Increases?

According to another popular theory, additional spending on education will yield gains in test scores. To see whether expenditure theory can account for the interstate variation, we plotted test-score gains against increments in spending between 1990 and 2009. As can be seen from the scattering of states into all parts of Figure 9, the data offer precious little support for the theory. Just about as many high-spending states showed relatively small gains as showed large ones. Maryland, Massachusetts, and New Jersey enjoyed substantial gains in student performance after committing substantial new fiscal resources. But other states with large spending increments—New York, Wyoming, and West Virginia, for example—had only marginal test-score gains to show for all that additional expenditure. And many states defied the theory by showing gains even when they did not commit much in the way of additional resources. It is true that spending and achievement gains have a slight positive relationship, but the 0.12 correlation between new expenditure and test-score gain is of no statistical or substantive significance. On average, an additional $1,000 in per-pupil spending is associated with a trivial annual gain in achievement of one-tenth of 1 percent of a standard deviation.

Who Spends Incremental Funds Wisely?

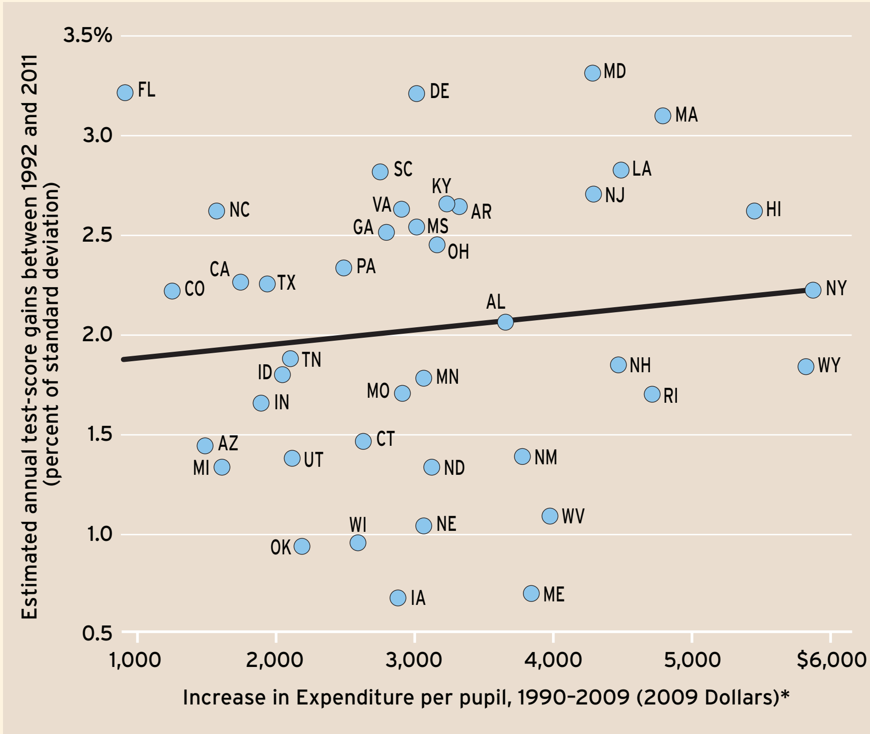
Some states received more educational bang for their additional expenditure buck than others. To ascertain which states were receiving the most from their incremental dollars, we ranked states on a "points per added dollar" basis. Michigan, Indiana, Idaho, North Carolina, Colorado, and Florida made the most achievement gains for every incremental dollar spent over the past two decades. At the other end of the spectrum are the states that received little back in terms of improved test-score performance from increments in per-pupil expenditure—Maine, Wyoming, Iowa, New York, and Nebraska. However, we do not know which kinds of expenditures prove to be the most productive or whether there are other factors that could explain variation in productivity among the states.

Causes of Change

There is some hint that those parts of the United States that took school reform the most seriously—Florida and North Carolina, for example—have

Michigan, Indiana, Idaho, North Carolina, Colorado, and Florida made the most achievement gains for every incremental dollar spent over the past two decades.

Figure 9. Relationship between increments in state expenditures per pupil and gains in student achievement, 1990–2008



*Change in expenditure per pupil adjusted for inflation; *Digest of Educational Statistics, 2011*

shown stronger rates of improvement, while states that have steadfastly resisted many school reforms (Iowa and Wisconsin, for instance) are among the nation's test-score laggards. But the connection between reforms and gains adduced thus far is only anecdotal, not definitive. Although changes among states within the United States appear to be explained in part by catch-up theory, we cannot pinpoint the specific factors that underlie this. We are also unable to find significant evidence that increased school expenditure, by itself, makes much of a difference. It is also possible that changes in test-score performance could be due to broader patterns of economic growth or varying rates of in-migration among states and countries. None of these propositions have been adequately tested, however, so any conclusions concerning the sources of educational gains must remain suggestive.

Have We Painted Too Rosy a Portrait?

Even the extent of the gains that have been made are uncertain. We have estimated gains of 1.6 percent of a std. dev. each year for the United States as a whole, or a total gain of 22 percent of a std. dev. over 14 years, a forward movement that has lifted performance by nearly a full year's worth of learning over the entire time period. A similar rate of gain is estimated for the students in all 49 participating countries. Such a rate of improvement is plausible, given the increased wealth in the industrialized world and the higher percentages of educated parents than in prior generations.

Still, this growth—normed against student performances on NAEP in 4th and 8th grades in 2000—is disproportionately affected by 4th-grade performance, possibly leading to too much optimism. When we estimate gains only from student performance in 8th grade (on the grounds that 4th-grade gains are meaningless unless they are observed for the same cohort four years later), our results show annual gains in the United States of only 1 percent of a std. dev. annually. The relative ranking of the United States remains essentially unchanged, however, as the estimated growth rates for 8th-graders in other countries is also lower than estimates that include students in 4th grade (see Appendix B, Figure B.1). Even this is above the rankings that would come from using the direct test linkages of PISA, a different approach that would place estimated annual growth rate for the United States at only one-half of 1 percent of a std. dev. (see Appendix B, Figure B.1).

An even darker picture emerges if one turns to the results for U.S. students at age 17, for which only minimal gains can be detected over the past two decades. We have not reported the results for 17-year-old students because the test administered to them does not provide information on the performance of students within individual states and no international comparisons are possible for this age group. Students themselves and the United States as a whole benefit from improved performance in the early grades only if that translates into measurably higher skills at the end of school. The fact that none of the gains observed in earlier years translate into improved high-school performance leaves one to wonder whether high schools are effectively building on the gains achieved in the earlier years of schooling. And while some scholars dismiss the results for 17-year-old students on the grounds that high-school students do not take the test seriously, others believe that the data indicate that the American high school has become a highly problematic educational institution. Amidst any uncertainties one fact remains clear, however: the measurable gains

The relative ranking of the United States remains essentially unchanged.

Education goal setting in the United States has often been utopian rather than realistic.

in achievement accomplished by more recent cohorts of students within the United States are being essentially matched by the measurable gains by students in the other 48 participating countries.

The Political Economy of Student Achievement

Few doubt that the quality of a nation's educational system is critical for its economic and political well-being. But too often the quality of an educational system is judged by the percentage of students graduating from high school or the percentages enrolled in college. While not denying the importance of attaining a high school diploma and a college degree, these credentials are meaningless unless they are accompanied by the acquisition of a set of skills that can prove useful later in life. And it turns out that those skills can be measured by the kinds of tests we have relied upon in assembling this report.

Hanushek and Woessmann demonstrate that a nation's growth rate of GDP is very closely related to the level of international test scores in math and science. The strong relationship displayed there is observed even after taking into account a variety of other factors that affect economic growth, including openness to international trade, regulations in labor and capital markets, security of property rights, and level of overall development. Significantly, educational attainment (number of years of schooling) appears to have little effect on economic growth, once the effects of educational achievement have been identified.[19]

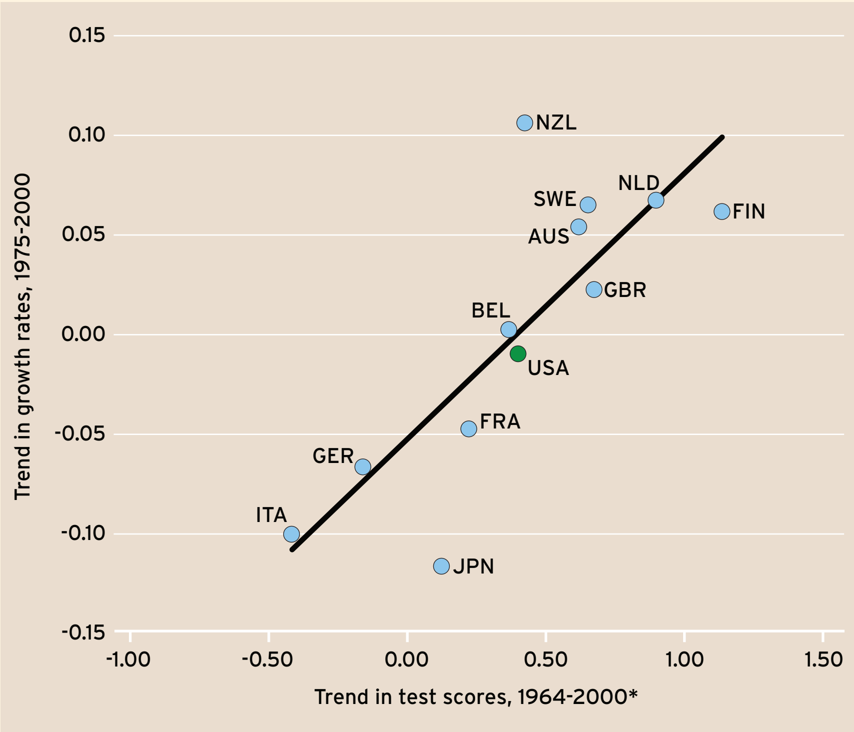
Here we focus on improvements in test scores for countries, not on the level of scores. If there is a causal relationship between test scores and growth, we should find that countries that test score *improvement* should be correlated with an *improvement in their growth rates*. In other words, *trends* in test scores should be related to *trends* in growth rates. In Figure 10, taken from a paper by Hanushek and Woessmann, is displayed the relationship between trends in a nation's test score performance and its trend in the rate of economic growth over the period 1975 to 2000.[20] It shows a very strong correlation between the two trends.

Because rates of economic growth have a huge impact of the future well-being of the nation, there is a simple message: A country ignores the quality of its schools at its economic peril.

Some would excuse the mediocre U.S. performance by claiming that it provides a more equal education to a much more diverse population than other countries do. It is claimed that test scores in the United States are lower than those in many other countries because they are not providing an education to all their students. That argument might have made some sense

19. Hanushek and Woessmann (2008, forthcming) show that, once consideration is given to the initial level of development, most of the difference in long run growth rates across countries can be explained by the very measures of educational achievement used here. (Considerations of initial developments simply reflects that countries starting behind can initially grow faster because they just have to imitate the technologies in more advanced countries rather than to develop new technologies).
20. Hanushek and Woessmann (forthcoming)

Figure 10. International trends in test scores and trends in economic growth



SOURCE: Hanushek and Woessmann (forthcoming)
*Original test score observations in 1964; scores in some countries observed later

fifty or seventy-five years ago, but it is a seriously dated view of the world. The data included in this report come from students who are between the ages of 8 and 15, and in virtually all the 49 countries participating in this study, only tiny percentages of the population within these age cohorts are not in school. And when it comes to high school completion rates, the United States, with a 72 percent graduation rate within four years of entering high school, performs no better than the *average* industrialized nation. Indeed, the countries that eclipse the United States in math and science have done it while also expanding dramatically their high school graduation and college enrollment rates, many reaching levels noticeably higher than those in the United States. Educating a broad swath of the population does not necessarily prevent high levels of performance.

The failure of the United States to close the international test-score gap, despite assiduous public assertions that every effort would be undertaken to produce that objective, raises questions about the nation's overall reform strategy. Education goal setting in the United States has often been utopian rather than realistic. In 1990, the president and the nation's governors announced the goal that all American students should graduate from high school, but two decades later only 75 percent of 9th-graders received their diploma within four years after entering high school. In 2002, Congress passed a law that declared that all students in all grades shall be proficient in math, reading, and science by 2014, but in 2012 most observers found that goal utterly beyond reach. Currently, the U.S. Department of Education has committed itself to ensuring that all students shall be college or career ready as they cross the stage on their high-school graduation day, another overly ambitious goal.

Perhaps the most unrealistic goal was that of the governors in 1990 when they called for the United States to achieve number-one ranking in the world in math and science by 2000. As this study shows, the United States is neither first nor is it catching up.

Consider a more realistic set of objectives for education policymakers, one that comes from our experience. If all U.S. states could increase their performance at the same rate as the highest-growth states—Maryland, Florida, Delaware, and Massachusetts—the U.S. improvement rate would be lifted by 1.5 percentage points of a std. dev. annually above the current trend line. Since student performance can improve at that rate in some countries and in some states, then, in principle, such gains can be made more generally. Those gains might seem small, but when viewed over two decades they accumulate to 30 percent of a std. dev., enough to bring the United States within the range of the world's leaders—unless, of course, they, too, continue to improve.

Such progress need not come at the expense of either the lowest-performing or the highest-performing students. In most states, a rising tide lifted all boats. Only in a few instances did the tide rise while leaving a disproportionate number stuck at the bottom, and most, if not all of the time, the high flyers moved ahead as well.

Appendix A: Methodology

Our international results are based on 28 administrations of comparable math, science, and reading tests between 1995 and 2009 to jurisdictionally representative samples of students in 49 countries and 4 subordinate jurisdictions. Our state-by-state results come from 36 administrations of math, reading, and science tests between 1992 and 2011 to representative samples of students in 41 of the U.S. states. These tests are part of four ongoing series: 1) National Assessment of Educational Progress (NAEP), administered by the U.S. Department of Education; 2) Programme for International Student Assessment (PISA), administered by the Organisation for Economic Co-operation and Development (OECD); 3) Trends in International Mathematics and Science Study (TIMSS), administered by the International Association for the Evaluation of Educational Achievement (IEA); and 4) Progress in International Reading Literacy Study (PIRLS), also administered by IEA.

Estimating Trends across Countries

First, we introduce the international tests and describe our sample of countries. Next, we describe the methodology used to express all international tests on a common scale that is also comparable to the state NAEP performance. Third, we discuss the methodology used to estimate each country's performance trend from the rescaled international test data.

The International Tests and the Sample of Countries. PISA was initiated in 2000 and has been conducted every three years since. Each cycle tests representative samples of 15-year-old students in mathematics, science, and reading. As a result, we can use 12 separate PISA tests: three subjects in four waves (2000, 2003, 2006, 2009; for several countries, the 2000 version of the test was administered in 2002, which we consequently use as an observation in the year 2002).[21]

TIMSS has been conducted every four years since 1995. It provides intertemporally comparable measures of 4th-grade and 8th-grade students in mathematics and science. Given its four testing waves (1995, 1999, 2003, and 2007) administered in two subjects at two grade levels (except that TIMSS did not test 4th graders in 1999), performance information is available for 14 separate TIMSS tests.[22] PIRLS was conducted in 2001 and in 2006 by the IEA. It tests the reading performance of 4th-graders, providing two tests to be used in the analysis.

In sum, 12 PISA tests, 14 TIMSS tests, and 2 PIRLS tests constitute 28 separate test results for those countries that participated in all surveys.

In most states, a rising tide lifted all boats.

21. OECD (2010). The assessment of whether the performance of an individual country on a specific test is deemed comparable over time is taken from Table A5.1 in this publication. To this, we have added the math score in 2000/2002, the science score in 2000/2002 and 2003, and remaining matched 2006/2009 scores from the corresponding publications of the respective PISA waves.

22. The TIMSS data are mostly taken from Mullis, Martin, and Foy (2008) and Martin, Mullis, and Foy (2008), which also provide the assessment of intertemporal comparability of individual country performance on a specific test. For countries not participating in TIMSS 2007 but in at least two previous cycles, we take the data from the corresponding publications on the respective previous TIMSS trends. As for the TIMSS performance of the United Kingdom, we use the population-weighted mean of England and Scotland, which participate separately in the TIMSS test. To ensure international comparability of tested ages and to avoid testing very young children, TIMSS has the rule that the average age of children in the grade tested should not be below 9.5 and 13.5 years old, respectively, in grades 4 and 8; otherwise, the next older grade will be tested in a country.

The PIRLS data, including the assessment of country-specific comparability across the two PIRLS waves, are taken from Mullis, Martin, Kennedy, and Foy (2007).

Unfortunately, only two countries (Hong Kong and Hungary) participated in all 28 tests. Twenty-seven tests are available for the United States. The U.S. did not report the results for the 2006 PISA test in reading, because problems in the administration of the test produced results that were deemed erroneous. The average number of test observations across the 49 countries covered in our analysis is 17.2 tests.

We excluded all countries for which results from fewer than nine separate tests were available and established additional rules for inclusion designed to ensure the trend analyses are based upon an adequate number of observations. First, a country's performance on any given test cycle (PISA, 4th-grade TIMSS, 8th-grade TIMSS, and PIRLS) is only considered if the country participated at least twice in the respective cycle, because otherwise no trend information would be contained in that cycle. Second, to ensure that any trend estimate is based on an adequate period of time, a country is excluded if the time span between its first and its last participation in international testing is less than seven years. Third, we do not consider a country if it did not participate after 2006 so as to ensure that all trend estimates extend to recent observations. Together, these rules exclude countries that did not participate in international testing prior to 2003.[23]

Deriving a Common Scale for All Tests. The international tests are measured on scales that are not directly comparable across the testing cycles. To transform the different international tests to a common scale, we follow procedures similar to those used in prior studies by Hanushek and Woessmann (2008, 2009, 2011). The following paragraphs describe the way in which these procedures were applied to the current analysis.

For the estimations reported in Figure 1, trends over time are expressed in terms of the 2000 wave of the NAEP testing cycle. Because the scores on the different subjects and grade levels of the NAEP are not directly comparable to one another, we first have to propose a method for making the trends on the NAEP subtests comparable. To do so, we express each testing cycle (of grade by subject) in terms of standard deviations of the U.S. population on the 2000 wave of each testing cycle. That is, within each testing cycle (which is comparable over time), the new scale is such that the U.S. performance has a standard deviation of 100 and a mean of 500 (the latter is arbitrary and without substance for the analysis of trends over time). This is a simple linear transformation of the NAEP scale on each testing cycle.

For example, the U.S. performance on the original NAEP score in mathematics in 8th grade is 273.1 (with a std. dev. of 38.1) in 2000 and 282.9

23. In fact, with the exception of Argentina, all countries in our analysis have at least 10 individual test observations.

(std. dev. 36.4) in 2009, i.e., the 2009 performance is 9.8 points or 25.8 percent of a 2000 std. dev. above the 2000 performance. By definition, the performance in 2000 on the transformed scale is 500 (std. dev. 100). The performance in 2009 on the transformed scale is 528.5, i.e., again 25.8 percent of a 2000 std. dev. above the 2000 performance, now expressed on the transformed scale. Similarly, we can put the 2009 std. dev. on the transformed scale, which is 95.6 (simply the original 2009 std. dev. expressed relative to the original 2000 std. dev.).

We express each international test on this transformed NAEP scale by performing a simple linear transformation of each international test based on the U.S. performance on the respective test. Specifically, we adjust both the mean and the std. dev. of each international test so that the U.S. performance on the tests is the same as the U.S. NAEP performance, expressed on the transformed NAEP scale. Specifically, the following steps are taken: First, from the international test expressed on the original international scale, subtract the U.S. mean on that scale. Second, divide by the U.S. std. dev. on that scale. Third, multiply by the U.S. standard deviation on the respective transformed NAEP scale for that year, subject, and grade (interpolated linearly within two available years if year is not a NAEP year).[24] Fourth, add the U.S. mean on the respective transformed NAEP scale for that year, subject, and grade.[25] Once these steps have been taken, all international tests are expressed on the transformed NAEP scale, where the U.S. population on the international test now has the performance (mean and std. dev.) that it has on the transformed NAEP scale, and all other countries are expressed relative to this U.S. performance on the respective international test. This allows us to estimate trends on the international tests on a common scale, whose property is that in the year 2000 it has a mean of 500 and a standard deviation of 100 for the United States.

Estimating Trends in Performance. The aim of our analysis is to estimate how each country's performance has changed over time. For that, we use all data points that a country has on the international tests, expressed on the transformed scale. Since a country may have specific strengths or weaknesses in specific subjects, at specific grade levels, or on specific international testing series, our trend estimation holds such differences constant by regressing, for each country, the test scores on a year variable, indicators for the international testing series (PISA, TIMSS, PIRLS), a grade indicator (4th vs. 8th grade), and subject indicators (mathematics, reading, science). This way, only the trends within each of these domains are used to estimate the overall trend of the country. This trend is indicated by the estimated coefficient on the year variable. It represents the

24. To rescale the TIMSS 1995 tests, we use the 1996 U.S. NAEP performance (1998 in reading), which is the earliest available intertemporally comparable NAEP score. For science performance beyond 2005, we use the 2005 U.S. NAEP science performance, which is the latest available intertemporally comparable NAEP science score. For the rescaling of the PISA tests, we use NAEP tests for 8th-graders.

25. The data on the U.S. means and std. dev. on the different NAEP tests are taken from http://nces.ed.gov/nationsreportcard/naepdata/ (accessed January 23, 2012). The data on the U.S. std. dev. on the different international tests are taken from the respective publication of each international testing cycle. The U.S. std. dev. on the rescaled 1995 TIMSS performance (which was subsequently expressed on a different, intertemporally comparable scale, without the rescaled U.S. std. dev. being published) was kindly provided by Michael Martin and Pierre Foy from the TIMSS & PIRLS International Study Center at Boston College.

annualized change in a country's test performance, expressed as a percentage of the standard deviation of the performance of the U.S. population in 2000. [26]

To see whether the results reported in Figure 1 are affected by the decision to norm all scales on NAEP 2000, we also compared the performance of countries on an alternate scale that is fully independent of NAEP information. We used the TIMSS and PISA tests (and ignored the two PIRLS observations), both of which have been performed in 2003, and used the U.S. performance (mean and standard deviation) on both tests in 2003 in order to splice the two series together. PISA scores are left just as they are. Then, we rescale the TIMSS 2003 tests so that the U.S. has the U.S. mean and standard deviation on the PISA 2003 test (in the respective subject). Then, we rescale the other TIMSS waves so that the U.S. performance (mean and standard deviation) on them is such that its difference to TIMSS 2003 is simply rescaled according to the rescaled TIMSS 2003 scale. What this ultimately provides is a series in which the TIMSS tests are rescaled in a way that the U.S. performance in 2003 is the same as in PISA, and where the TIMSS trends are the original trends, only that their size is expressed according to the U.S. std. dev. in PISA 2003. The rankings of the countries remain essentially the same as those reported in the main analysis. On this scale, the U.S. ranks number 26 among the 49 countries. However, the annual gain of the United States is only 0.46 percent of a std. dev., substantially less than 1.53 percent of a std. dev., as estimated in the main analysis. Gains for other countries are also substantially reduced in size. In other words, the most reliable information that we report are the gains made relative to those of other jurisdictions, not the absolute size of the gains, which vary depending on the scale that is used. Full results are reported in Appendix B, Figure B.2.

We also performed the analysis separately for each subject, for each testing series, for each grade level, and for mathematics and reading (dropping the science observations). Results are qualitatively similar. Results that exclude gains for 4th-graders and nine-year olds are reported in Appendix B, Figure B.2. The following procedure was used to estimate the statistical significance of trend lines. Step 1: Calculate the difference between the point estimates of the trends of two countries. Step 2: Calculate the square root of the sum of the variance of the two trend estimates (i.e., the standard error of this difference is given by the square root of the sum of the squared standard errors of the two estimates). The result from step 1 divided by the result from step 2 yields the t-statistic for the significance of the difference.

26. Note that, by construction, the international trend estimate for the United States is effectively a weighted average of the U.S. trend in NAEP performance (in the same regression, controlling for grade and subject), where the weights are the international test occurrences (which are: 3.6% 4m96 [4th-grade mathematics test in 1996], 3.6% 4m03, 3.6% 4m07, 3.6% 4s96, 2.1% 4s00, 5.0% 4s05, 1.8% 4r00, 1.8% 4r02, 1.8% 4r05, 1.8% 4r07, 4.5% 8m96, 6.3% 8m00, 7.1% 8m03, 1.8% 8m05, 5.4% 8m07, 3.6% 8m09, 4.5% 8s96, 10.5% 8s00, 13.6% 8s05, 1.8% 8r98, 1.8% 8r02, 3.6% 8r03, 1.8% 8r05, 1.8% 8r07, 3.6% 8r09).

Estimating Trends across U.S. States

For the analysis of U.S. states, observations are available for only 41 states. The remaining states did not participate in NAEP tests until 2002. Annual gains for states are calculated for a 19-year period (1992 to 2011), the longest interval that could be observed for the 41 states.

Trends for each U.S. state are estimated by using procedures similar to those used to estimate country trends. That is, the NAEP data are first transformed to the common scale that has a mean of 500 and a standard deviation of 100 for the United States population in the year 2000. Then, for each U.S. state, the transformed test scores are regressed on a year variable, a grade indicator (4th vs. 8th grade), and subject indicators (mathematics, reading, science). The overall trend of the state is indicated by the estimated coefficient on the year variable.

International comparisons are for a 14-year period (1995 to 2009), the longest time span that could be observed with an adequate number of international tests. To facilitate a comparison between the United States as a whole and other nations, the aggregate U.S. trend is estimated from that same 14-year period and each U.S. test is weighted to take into account the specific years that international tests were administered. Because of the difference in length and because international tests are not administered in exactly the same years as the NAEP tests, the results for each state are not perfectly calibrated to the international tests, and each state appears to be doing slightly better internationally than would be the case if the calibration were exact. The differences are marginal, however, and the comparative ranking of states is not affected by this discrepancy. ■

Appendix B:
Alternative Estimations of Trends in Student Math Achievement

Table B.1 Annual growth in test scores in 49 countries

State	Annual test score change as % of std. dev.*	Std. err.	t**	Number of observations	Time span
Latvia†	4.70%	0.77	6.07	26	1995–2009
Chile	4.37	2.21	1.98	13	1999–2009
Brazil	4.05	1.47	2.75	12	2000–2009
Portugal†	3.99	0.93	4.31	12	2000–2009
Hong Kong†	3.93	0.90	4.36	28	1995–2009
Germany†	3.77	0.85	4.45	14	2000–2009
Poland†	3.72	1.18	3.16	12	2000–2009
Liechtenstein	3.67	1.74	2.10	12	2000–2009
Slovenia†	3.58	0.99	3.61	20	1995–2009
Colombia	3.33	1.34	2.48	10	1995–2009
Lithuania	3.21	1.02	3.13	20	1995–2009
United Kingdom	2.84	1.04	2.72	22	1995–2009
Singapore	2.80	1.06	2.63	16	1995–2007
Switzerland	2.33	0.84	2.77	12	2000–2009
Greece	2.25	1.34	1.68	12	2000–2009
Mexico	2.21	1.43	1.54	12	2000–2009
Israel	1.98	1.43	1.38	17	1999–2009
Finland	1.97	1.15	1.71	12	2000–2009
Italy	1.83	0.79	2.33	24	1999–2009
New Zealand	1.73	0.73	2.36	26	1995–2009
Denmark	1.62	0.81	2.00	12	2000–2009
Korea, Rep.	1.61	0.88	1.82	20	1995–2009
Hungary	1.61	0.57	2.82	28	1995–2009
Iran	1.59	1.21	1.31	16	1995–2007
United States	1.57	0.39	4.04	27	1995–2009
Taiwan	1.30	1.14	1.14	16	1999–2009
Belgium	1.24	0.98	1.26	12	2000–2009
Canada	1.07	0.72	1.47	16	1995–2009
Cyprus	1.02	1.73	0.59	12	1995–2007
Australia	0.99	0.58	1.70	24	1995–2009
Jordan	0.88	1.20	0.73	12	1999–2009
Russian Fed.	0.83	0.88	0.94	26	1995–2009
Indonesia	0.73	1.43	0.51	18	1999–2009
Austria	0.67	0.67	1.00	13	1995–2007
Spain	0.65	1.24	0.52	12	2000–2009
Iceland	0.62	0.79	0.79	14	2000–2009
Japan	0.55	0.62	0.89	26	1995–2009
Netherlands	0.45	0.65	0.69	23	1995–2009
Tunisia	0.18	1.78	0.10	19	1999–2009
Argentina	0.14	2.16	0.06	9	2002–2009
France†	–0.13	0.69	–0.19	14	2000–2009
Ireland†	–0.47	1.03	–0.46	12	2000–2009
Norway†	–0.61	0.74	–0.82	26	1995–2009
Romania†	–1.12	0.84	–1.33	19	1995–2009
Czech Rep.†	–1.25	0.71	–1.76	22	1995–2009
Slovak Rep.†	–1.33	1.08	–1.24	17	1995–2009
Thailand†	–1.54	0.78	–1.96	16	1999–2009
Bulgaria†	–1.81	1.65	–1.10	18	1995–2009
Sweden†	–2.55	0.76	–3.35	20	1995–2009

*Annual test score change as % of std. dev.
**significantly different from zero
† significantly different from United States

Table B.2 Average gain and percent change in percentage below basic and percentage proficient, 1992–2011, by state

State	Average gain as % of std. dev.*	4th Grade: % reduction in percentage below basic	4th Grade: % reduction in percentage below proficiency	8th Grade: % reduction in percentage below basic	8th Grade: % reduction in percentage below proficiency
Maryland	3.3%	69%	36%	44%	26%
Florida	3.2	66	28	37	15
Delaware	3.2	64	26	46	20
Massachusetts	3.1	76	46	61	36
Louisiana	2.8	56	20	42	16
South Carolina	2.8	60	27	43	20
New Jersey	2.7	65	35	53	30
Kentucky	2.7	69	30	42	20
Arkansas	2.6	63	31	46	21
Virginia	2.6	69	33	48	25
Hawaii	2.6	59	29	40	19
North Carolina	2.6	76	36	53	28
Mississippi	2.5	57	20	37	14
Georgia	2.5	56	26	40	17
Ohio	2.5	68	35	49	25
Pennsylvania	2.3	62	33	31	22
California	2.3	51	24	22	11
Texas	2.3	67	28	61	27
New York	2.2	53	23	30	12
Colorado	2.2	60	36	45	28
Alabama	2.1	56	19	35	11
Tennessee	1.9	53	22	33	14
New Hampshire	1.8	73	43	36	25
Wyoming	1.8	61	31	40	21
Idaho	1.8	55	28	30	19
Minnesota	1.8	59	36	35	24
Missouri	1.7	54	28	27	15
Rhode Island	1.7	65	34	39	22
Indiana	1.7	67	34	42	18
Connecticut	1.5	45	28	31	17
Arizona	1.4	51	24	30	19
New Mexico	1.4	50	21	32	14
Utah	1.4	56	30	19	16
North Dakota	1.3	63	31	30	19
Michigan	1.3	45	20	30	15
West Virginia	1.1	54	22	34	13
Nebraska	1.0	48	23	14	9
Wisconsin	1.0	52	30	27	19
Oklahoma	0.9	57	23	30	13
Maine	0.7	47	24	24	18
Iowa	0.7	48	23	0	3

Note: Data unavailable for nine states
* Annual test score change as % of std. dev. for math, reading, and science (see methodology section for calculations.)

Figure B.1 Overall annual rate of growth in student achievement in math, reading, and science,

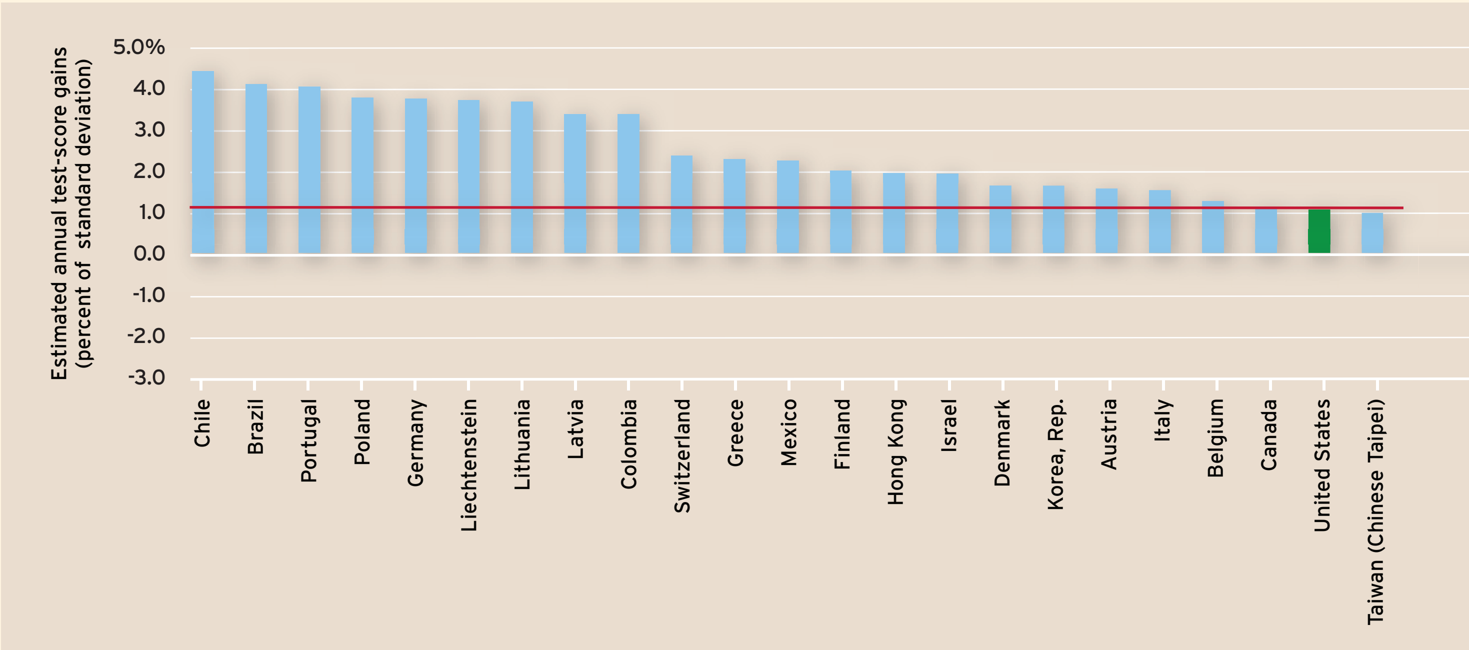
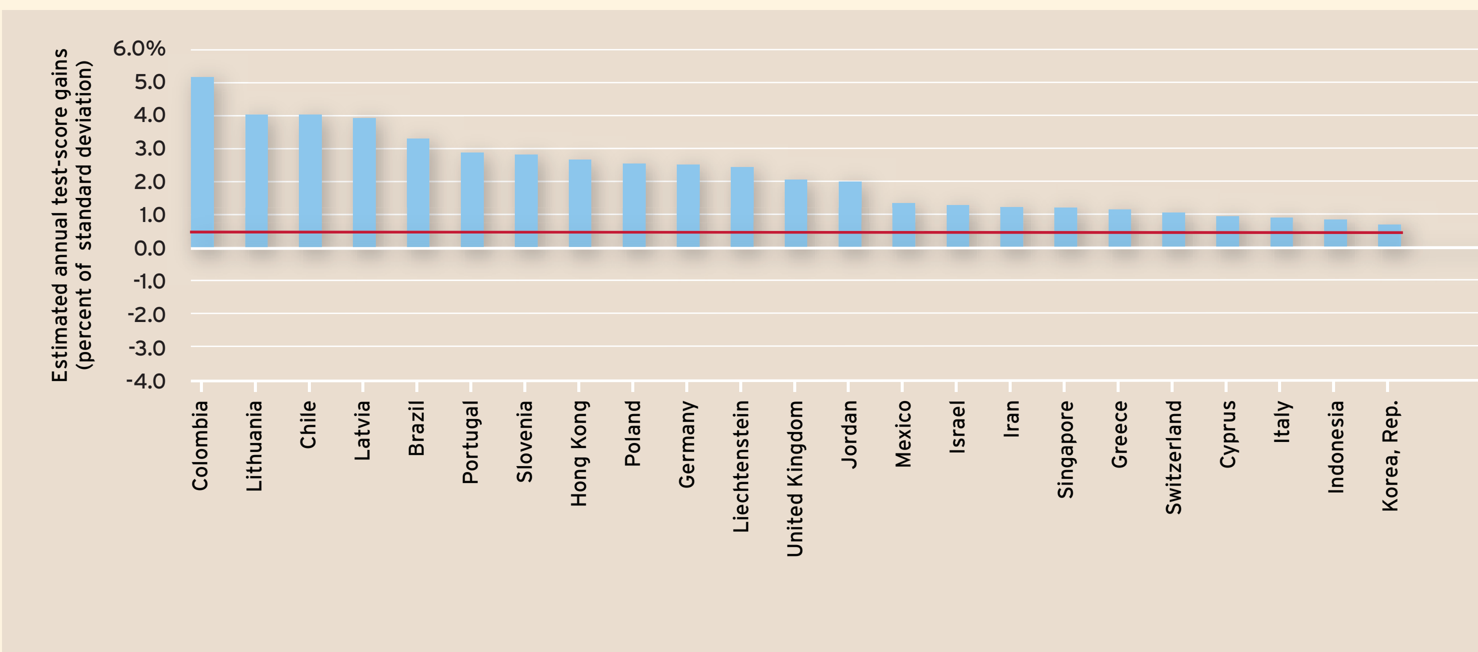
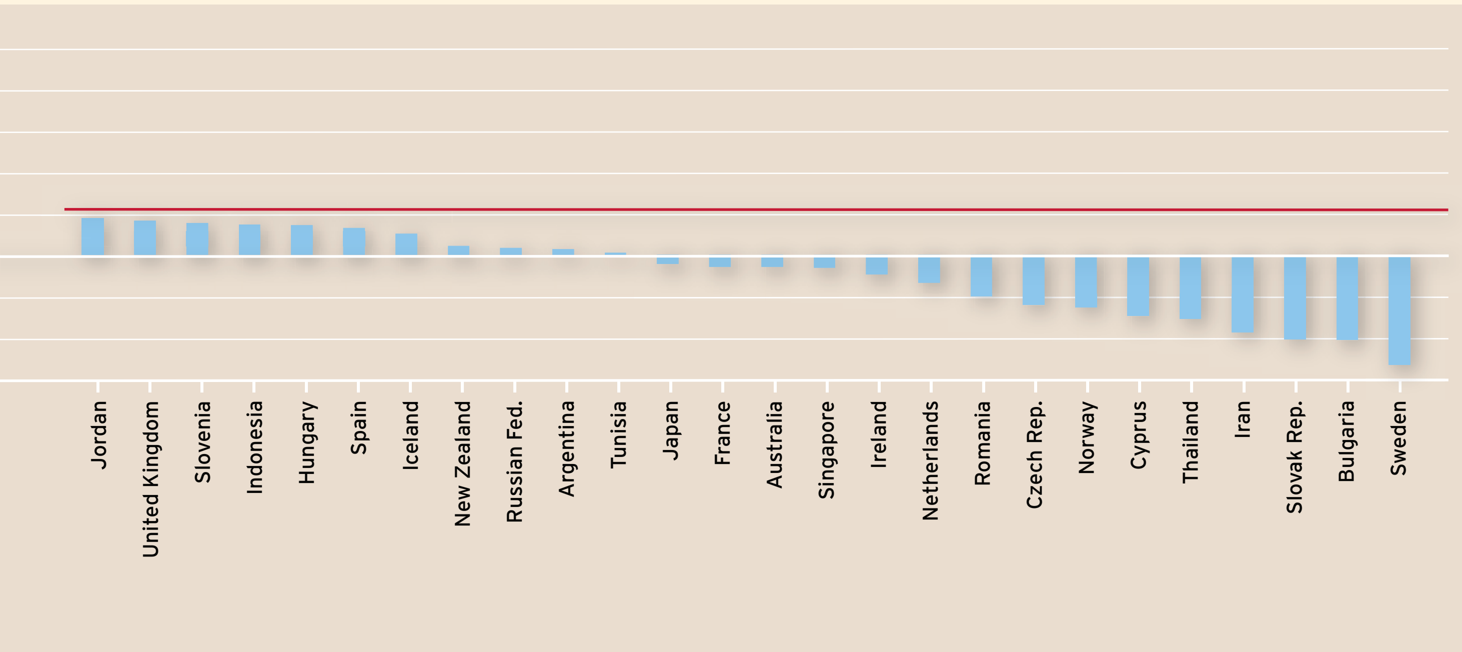


Figure B.2 Overall annual rate of growth in student achievement in math, reading and science in



*See Appendix A, pp. 23–25 for details on this specific methodology

1995–2009 (excluding 4th grade and 9-year-old performances)



49 countries, 1995–2009 (normed on performances on PISA 2003)

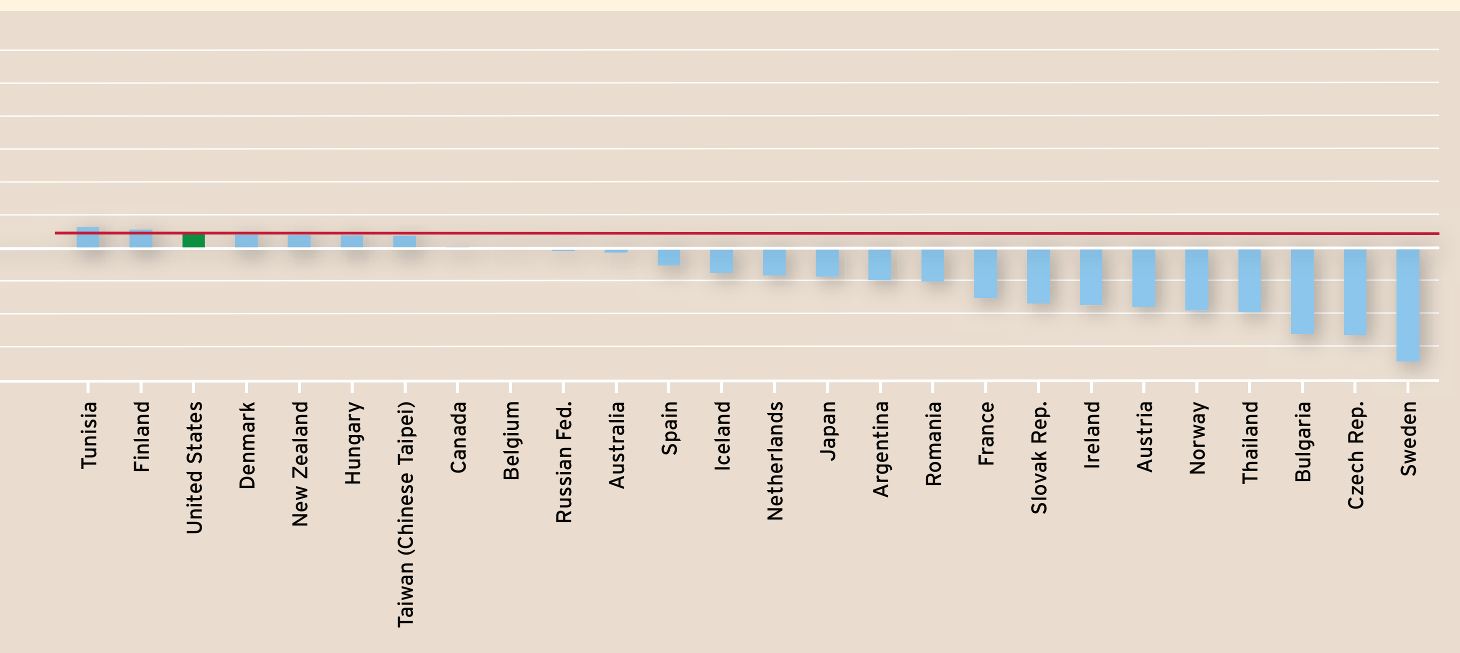


Figure B.3 Overall annual rate of growth in student achievement in math and reading in 41 U.S.

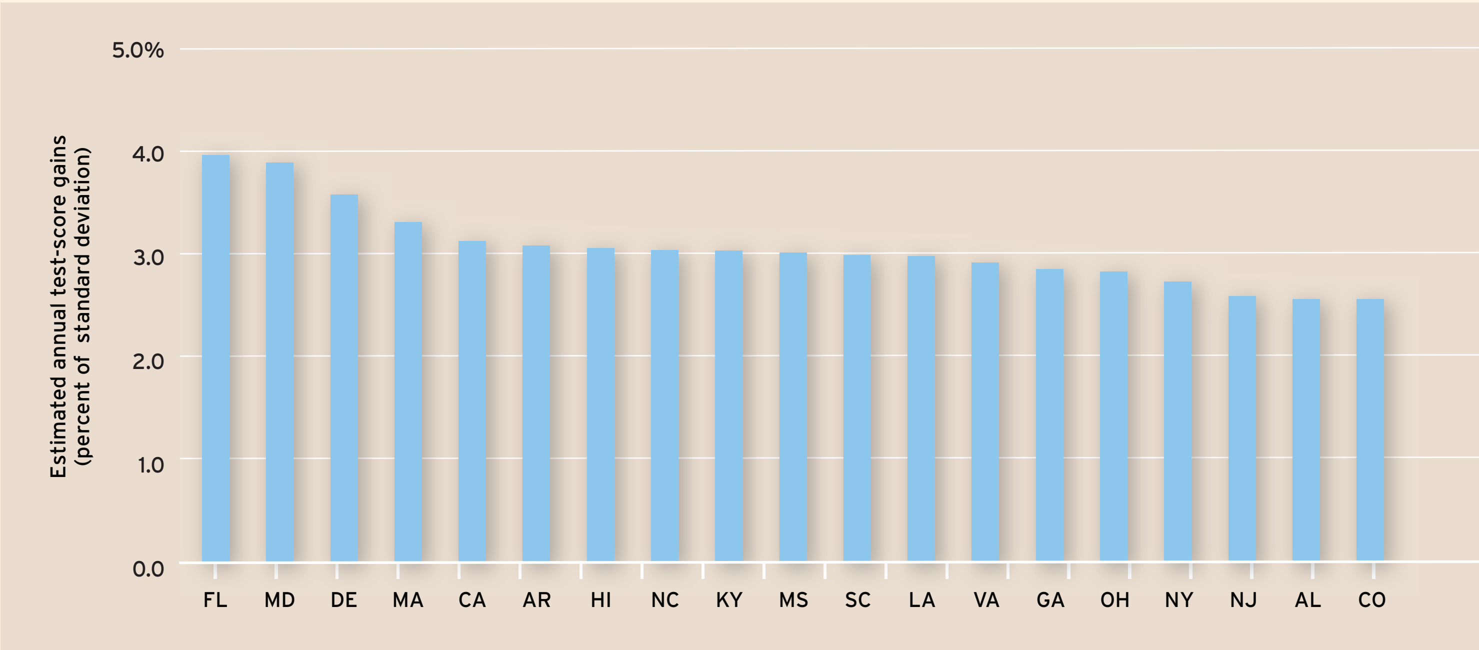
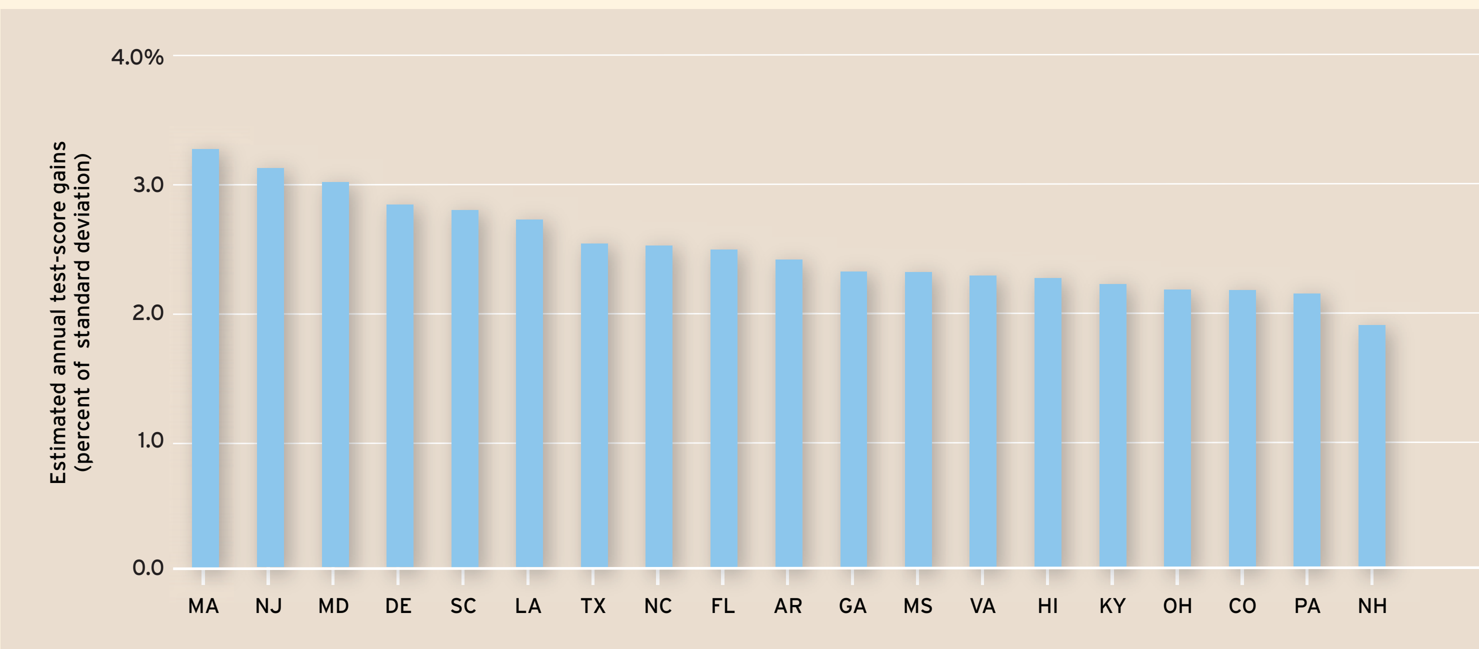
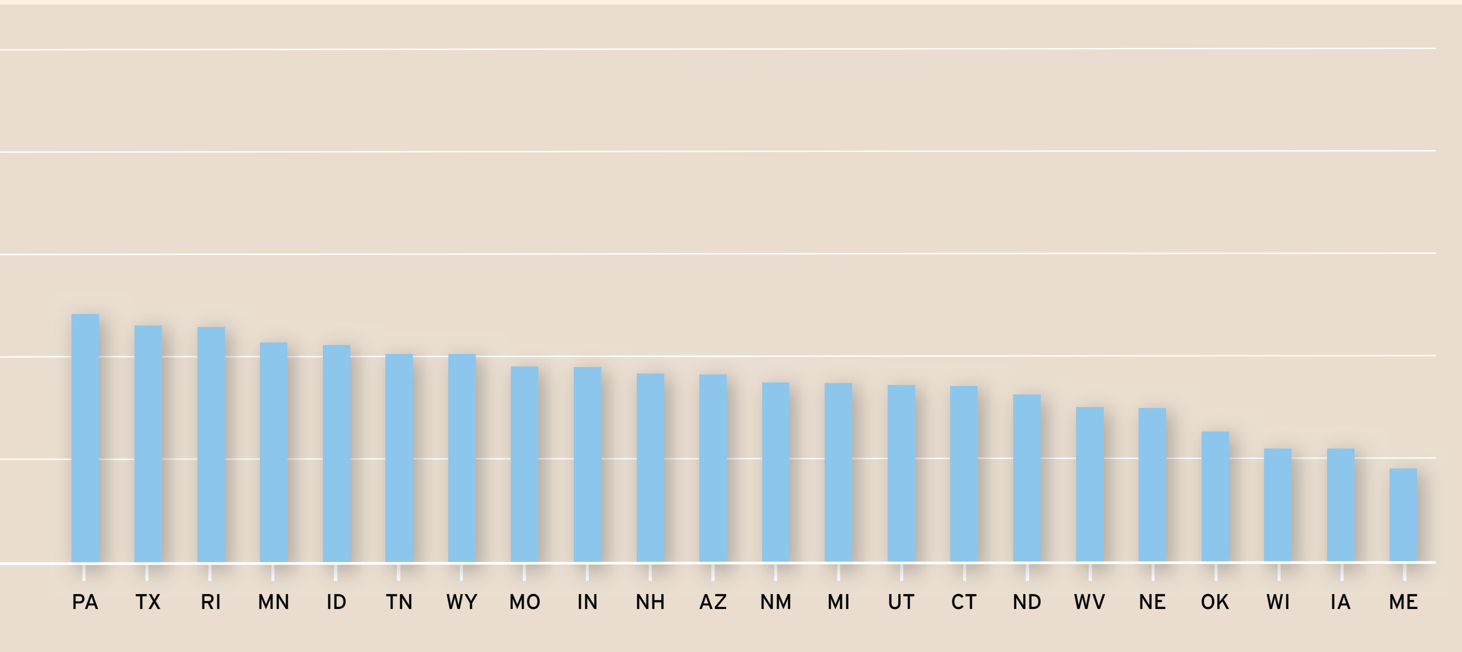


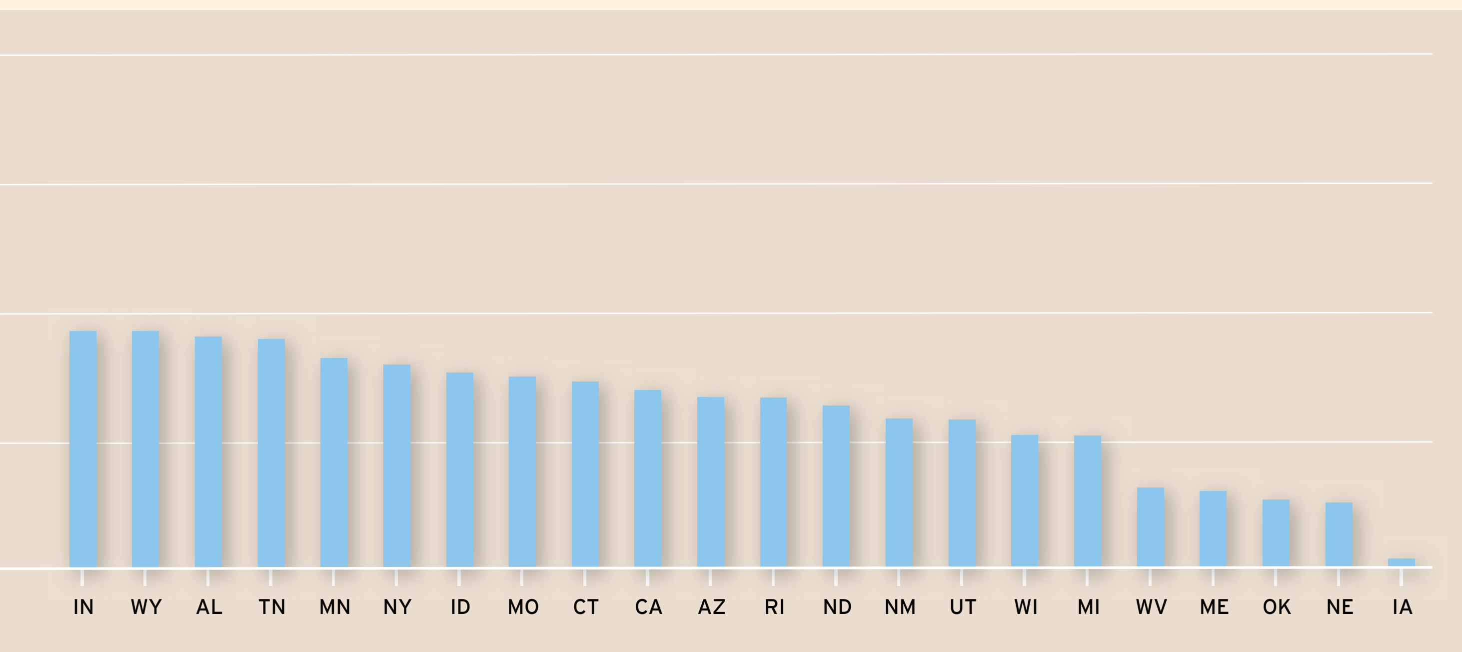
Figure B.4 Overall annual rate of growth in student achievement in math and reading in 41 U.S.



states, 4th grade



states, 8th grade



References

Bandeira de Mello, Victor, Charles Blankenship, and Don McLaughlin (2009): *Mapping state proficiency standards onto the NAEP scales: 2005–2007*. Washington, DC: National Center for Education Statistics.

Flattau, Pamela Ebert, Jerome Bracken, Richard Van Atta, Ayeh Bandeh-Ahmadi, Rodolfo de la Cruz, and Kay Sullivan (2006): *The National Defense Education Act of 1958: Selected Outcomes*. Washington, DC: Institute for Defense Analyses (March).

Hanushek, Eric A., Paul E. Peterson, and Ludger Woessmann (2010): *U.S. Math Performance in Global Perspective: How well does each state do at producing high-achieving students?* Harvard's Program on Education Policy and Governance, Report No.: 10–19.

Hanushek, Eric A. and Ludger Woessmann (2008): "The role of cognitive skills in economic development," *Journal of Economic Literature* 46, no. 3 (September): 607–668.

Hanushek, Eric A. and Ludger Woessmann (forthcoming): "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation," *Journal of Economic Growth*.

Hanushek, Eric A. and Ludger Woessmann (2011): "The economics of international differences in educational achievement." In *Handbook of the Economics of Education, Vol. 3*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland: 89–200.

Howell, William G., Paul E. Peterson, and Martin R. West (2009): "The Persuadable Public," *Education Next* Vol. 9, No. 4: 20–29.

Independent Task Force on U.S. Education Reform and National Security, Council on Foreign Relations (2012): *U.S. Education Reform and National Security*, Report No. 68 (March).

Martin, Michael O., Ina V. S. Mullis, and Pierre Foy (2008): *TIMSS 2007 international science report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Mourshed, Mona, Chinezi Chijioke, and Michael Barber (2010): *How the world's most improved school systems keep getting better*: McKinsey and Company.

Mullis, Ina V.S., Michael O. Martin, Ann M. Kennedy, and Pierre Foy (2007): *PIRLS 2006 international report: IEA's Progress in International Reading Literacy Study in primary schools in 40 countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College: 43–46.

Mullis, Ina V. S., Michael O. Martin, and Pierre Foy (2008): *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College: 43–56.

National Commission on Excellence in Education. (1983): *A nation at risk: the imperative for educational reform*. Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics (2011): *Digest of Education Statistics 2010*. Washington, DC: NCES.

OECD (2010): *PISA 2009 Results, Volume 5 Learning Trends—Changes in Student Performance since 2000 (Volume V)*. Paris: OECD.

OECD (2011): *Strong performers and successful reformers in education: Lessons from PISA for the United States*. Paris: OECD.

Peterson, Paul E. (2010): *Saving Schools: From Horace Mann to Virtual Learning*. Cambridge, MA: Belknap Press of Harvard University Press.

Peterson, Paul E. and Carlos X. Lastra-Anadón (2010): "State Standards Rise in Reading, Fall in Math," *Education Next* Vol. 10, No. 4: 12–16.

Peterson, Paul E., Ludger Woessmann, Eric A. Hanushek, and Carlos X. Lastra-Anadón (2011): *Globally Challenged: Are U.S. Students Ready to Compete?* Harvard's Program on Education Policy and Governance, Report No.: 11–03.

This is the third in a series of reports on the state of American education in international perspective. The two prior reports are:

Peterson, Paul E., Ludger Woessmann, Eric A. Hanushek, and Carlos X. Lastra-Anadón (2011): *Globally Challenged: Are U.S. Students Ready to Compete?* Harvard's Program on Education Policy and Governance, Report No.: 11–03.
http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG11-03_GloballyChallenged.pdf

Hanushek, Eric A., Paul E. Peterson and Ludger Woessmann (2010): *U.S. Math Performance in Global Perspective: How well does each state do at producing high-achieving students?* Harvard's Program on Education Policy and Governance Report, No.: 10–19.
http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG10-19_HanushekPetersonWoessmann.pdf

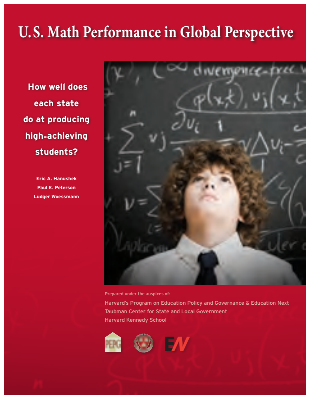
Biographical Sketches

Eric A. Hanushek is the Paul and Jean Hanna Senior Fellow at the Hoover Institution of Stanford University. He is chairman of the Executive Committee for the Texas Schools Project at the University of Texas at Dallas, a research associate of the National Bureau of Economic Research, a member of the Koret Task Force on K–12 Education, and the area coordinator for Economics of Education of the CESifo Research Network. He currently serves as a commissioner on the Equity and Excellence Commission of the U.S. Department of Education and was the past chair of the Board of Directors of the National Board for Education Sciences. His most recent book, *Schoolhouses, Courthouses, and Statehouses*, describes how improved school-finance policies can be used to meet our achievement goals.
Website: hanushek.net

Paul E. Peterson is the director of the Program on Education Policy and Governance in the Taubman Center for State and Local Government at Harvard's Kennedy School of Government. He is the Henry Lee Shattuck Professor of Government in the Faculty of Arts and Sciences at Harvard University and a senior fellow at the Hoover Institution at Stanford University. He is the editor-in-chief of *Education Next: A Journal of Opinion and Research*. His book *Saving Schools: From Horace Mann to Virtual Learning* (Harvard University Press, 2010) describes the rise, decline, and potential transformation of the American school.
Websites: hks.harvard.edu/pepg • educationnext.org • savingschools.com

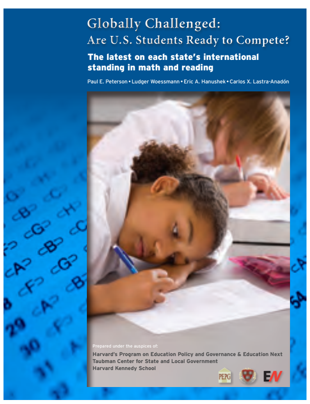
Ludger Woessmann is professor of economics at the University of Munich, head of the Department of Human Capital and Innovation at Ifo Institute for Economic Research, and coordinator of the European Expert Network on the Economics of Education (EENEE). His most recent book, *School Accountability, Autonomy and Choice around the World*, considers sources of international differences in student achievement.
Website: cesifo.de/woessmann

The Harvard PEPG & *Education Next* series of reports on the state of American education in international perspective:



Eric A. Hanushek, Paul E. Peterson and Ludger Woessmann (2010):

U.S. Math Performance in Global Perspective: How well does each state do at producing high-achieving students?
Harvard's Program on Education Policy and Governance Report, No.: 10–19.
http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG10-19_HanushekPetersonWoessmann.pdf

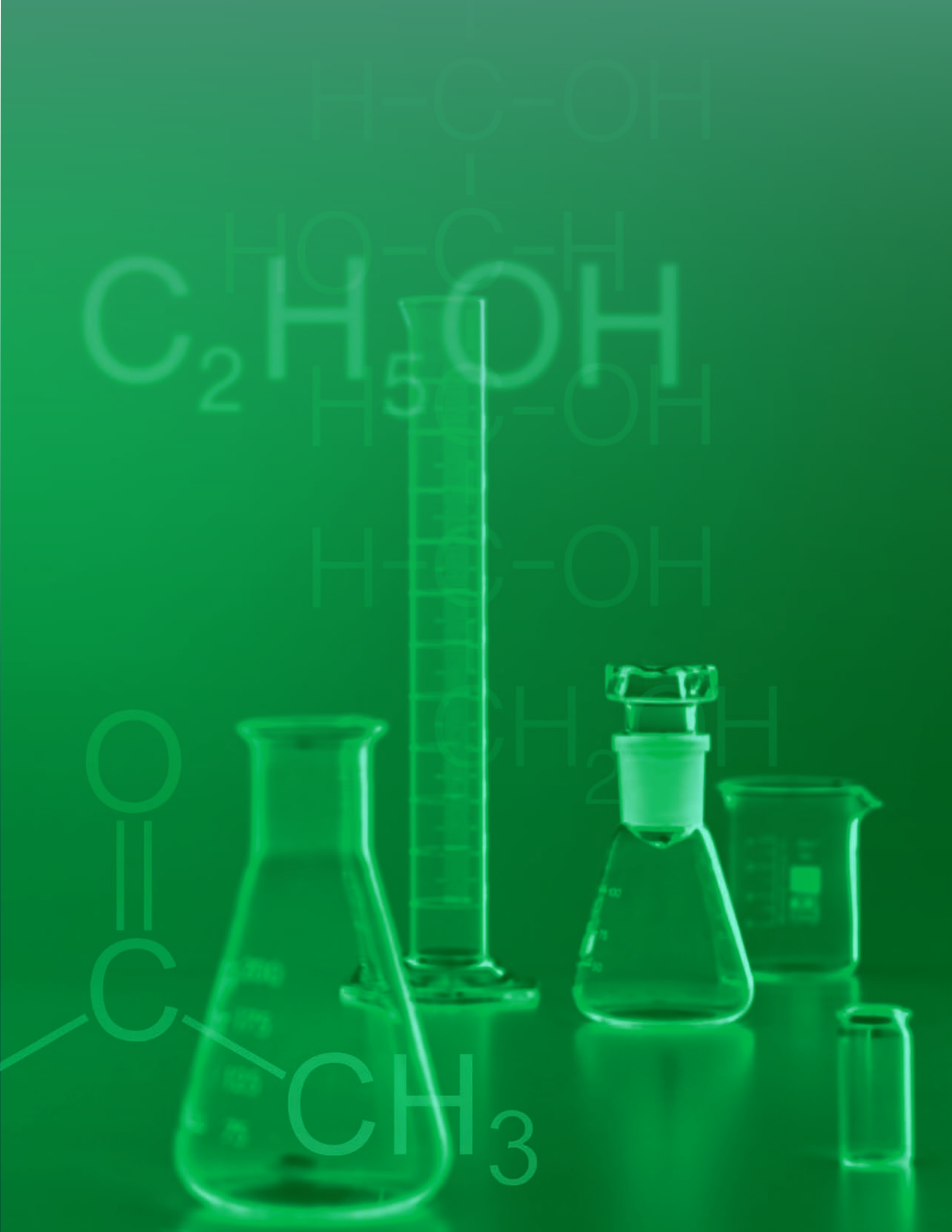


Paul E. Peterson, Ludger Woessmann, Eric A. Hanushek, and Carlos X. Lastra-Anadón (2011):

Globally Challenged: Are U.S. Students Ready to Compete?
Harvard's Program on Education Policy and Governance,
Report No.: 11–03.
http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG11-03_GloballyChallenged.pdf

Eric A. Hanushek, Peterson, Paul E., and Ludger Woessmann (2012):

Achievement Growth: International and U.S. State Trends in Student Performance.
Harvard's Program on Education Policy and Governance & *Education Next*
Report No.: 12–03.
http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG12-03_CatchingUp.pdf



C_2H_5OH
CH_3

Harvard's Program on Education Policy & Governance (PEPG)
Harvard Kennedy School
79 John F. Kennedy Street, Taubman 304
Cambridge, MA 02138
Phone: (617) 495-7976 • Fax: (617) 496-4428
Email: pepg_administrator@hks.harvard.edu
Websites: hks.harvard.edu/pepg • educationnext.org