

education sector reports

# The New State Achievement Gap: How Federal Waivers Could Make It Worse—Or Better

By John Chubb and Constance Clark



EDUCATIONSECTOR

[www.educationsector.org](http://www.educationsector.org)

## ACKNOWLEDGEMENTS

*The New State Achievement Gap* was funded by a generous grant from The William and Flora Hewlett Foundation. The views expressed in this paper are those of the authors alone.

## ABOUT THE AUTHORS

**JOHN CHUBB** is interim CEO of Education Sector and the author of numerous books on education reform, including, *Learning from No Child Left Behind* (2009) and *The Best Teachers in the World* (2012). He is the incoming president of the National Association of Independent Schools, where he can be reached after July 1, 2013.

[jchubb@educationsector.org](mailto:jchubb@educationsector.org)

**CONSTANCE CLARK** is a policy analyst at Education Sector. She has written about issues of race, wealth, equity and school choice. She was co-author on the 2012 Education Sector Report *High Standards Help Struggling Students: New Evidence*. [cclark@educationsector.org](mailto:cclark@educationsector.org)

## ABOUT EDUCATION SECTOR

Education Sector is an independent think tank that challenges conventional thinking in education policy. We are a nonprofit, nonpartisan organization committed to achieving measurable impact in education, both by improving existing reform initiatives and by developing new, innovative solutions to our nation's most pressing education problems.

Cover photo: iStockphoto/Thinkstock

© Copyright 2013 Education Sector

*Education Sector encourages the free use, reproduction, and distribution of our ideas, perspectives, and analyses. Our Creative Commons licensing allows for the noncommercial use of all Education Sector authored or commissioned materials. We require attribution for all use. For more information and instructions on the commercial use of our materials, please visit our website, [www.educationsector.org](http://www.educationsector.org).*

1201 Connecticut Ave., N.W., Suite 850, Washington, D.C. 20036  
202.552.2840 • [www.educationsector.org](http://www.educationsector.org)

---

“Now within the past three weeks the House of Representatives, by a vote of 263 to 153, and the Senate, by a vote of 73 to 18, have passed the most sweeping educational bill ever to come before Congress. It represents a major new commitment of the Federal Government to quality and equality in the schooling that we offer our young people. By passing this bill, we bridge the gap between helplessness and hope for five million educationally deprived children.”

—PRESIDENT LYNDON B. JOHNSON  
*upon signing the Elementary and  
Secondary Education Act,  
April 11, 1965*

---

For nearly 50 years the federal government has sought to play a critical role in our nation’s schools, a role largely resisted by state and local governments until the historic passage of the Elementary and Secondary Education Act (ESEA) as part of the nation’s Great Society reforms. The most artful of compromises, ESEA enabled Washington to channel funds to the neediest schools without usurping the power of local authorities to govern education. But the compromise never worked particularly well for the nation’s poorest students, and Washington regularly looked for ways to ensure that its dollars actually helped their intended recipients. With each reauthorization, ESEA asserted greater federal authority over the education of disadvantaged students.

In 2002, a bipartisan coalition agreed that the goal of equality required a comprehensive solution, and adopted No Child Left Behind (NCLB), in what was to become the last reauthorization of ESEA. The law has been due for reauthorization since 2007, but the coalitions that once supported a stronger federal role have splintered. Both parties seem to believe that Washington has exceeded its reach and that education equality—no less important than ever—is best promoted by the states. In 2011, the Obama Administration began granting states waivers from ESEA’s major regulations in exchange

for commitments to implement new approaches to helping disadvantaged students. The hope in Washington is that the states now can show the federal government how to help the poor—and how to reauthorize a more effective ESEA.

But hope, as the saying goes, is not a plan. On its current course, the federal government could easily allow education inequality to get worse.

History suggests as much. ESEA was passed in the first place because states lacked the resources, the commitment or both to educate those most in need. Fifty years of to and fro with Washington have been largely about encouraging lagging locales to try and do more. While states and districts have made progress, there is no evidence that freed from the strictures of NCLB they will do a better job of raising achievement for all of their students. The evidence indicates, we believe, that left to their own devices, the states will exacerbate the nation’s achievement gap between haves and have-nots. At the same time, the evidence is not that NCLB should be maintained either. While the nation has made surprising progress educating the disadvantaged during the NCLB era—surprising given the antipathy toward the law—the progress has not been nationwide. Gains have been concentrated in a handful of states. The lesson for the future is not

that it's time for the states to take the lead in promoting equality. It's that the nation should ensure that all states learn from the successes of the few.

## Waivers Are Just Words

In February 2013, the Senate Health, Education, Labor and Pensions (HELP) Committee held a hearing on NCLB waivers. Education Secretary Arne Duncan explained the logic of waivers—to maintain the ambitious goals and bright transparency of NCLB while cutting the federal red tape that has prevented states from adopting sensible school improvements. He offered examples of the good things that states are doing with their flexibility. State officials and national experts followed the Secretary with generally supportive testimony and “lessons learned” from early state experience.<sup>1</sup> But Andy Smarick<sup>2</sup> of Bellwether Education Partners, prudently—and correctly—suggested that it is too early to tell exactly how the waivers will work out. They are barely a year old. The overall feeling was that waivers offered an appropriate balance of federal guidance and relief. The federal government can pursue the noble ends of NCLB while allowing states to have far more say over the means.

The operative word is “say.” For each state waiver application is merely a set of assurances and plans. In exchange for reprieves from NCLB sanctions—for failing to make Adequate Yearly Progress (AYP) toward 100 percent student proficiency in reading and math in 2014—states now are permitted to set new end goals and offer new improvement plans, subject to federal guidelines. Under NCLB, it should be recalled, states also had to set objectives and make plans. Those commitments often went unfilled. Similarly, waivers may turn out to be just more unkept promises.

There is ample reason for skepticism. Consider the hard evidence. While NCLB waivers have been in place for only a year, providing little systematic data, NCLB itself has been in place for a decade. It is possible to get an extended look at how states have fulfilled the law's high expectations while handling its sometimes burdensome regulations. NCLB is ultimately about ensuring that all students acquire decent math and reading skills, from the early grades through high school. NCLB leaves it to the states to set math and reading standards and to administer assessments. Judging progress with 51 different

**For the last decade, it has been possible to compare the progress of states to one another and to the nation as a whole. NAEP can help tell us which states have delivered on past promises and which have not.**

state tests (50 states plus Washington, D.C.) is a challenge. An alternative gauge is provided by the National Assessment of Education Progress (NAEP). Since 1970, it has served as “the nation's report card,” offering periodic indicators of student achievement on a national basis, pegged to a set of standards overseen by a nationwide panel of experts. In the 1990s, NAEP was offered to individual states to assess their progress against national standards. With the passage of NCLB, state participation in NAEP became mandatory. For the last decade, it has been possible to compare the progress of states to one another and to the nation as a whole. NAEP can help tell us which states have delivered on past promises and which have not.

## The New State Achievement Gap

Since 2003, state level reading and math NAEP scores for students in grades four and eight have been available for all 50 states plus the District of Columbia. High school scores are available for only a dozen states, as are scores in other subjects. We confined our analysis to reading and math in grades four and eight. Since NCLB focused its assessment requirements most intensively on grades three to eight, where annual reading and math testing was required, the focus of our work is appropriate. For simplicity, we analyzed composite scores comprising the sum of the grade four and grade eight reading and math tests. We also examined each subject and grade level individually, and the patterns that emerged in the composite scores were very similar to those in the individual tests.

Student achievement improved for students of all types during the first decade of NCLB. In **Figure 1** we compare scores in 2003 to scores in 2011, the most recent assessments available. For all students, in nationwide samples, the composite scale score increased from 987 to 1007, a gain of 20 points. To put that in context, NAEP scale scores measure progress continuously from grade four through high school. From grade four through grade eight in reading and math, a student on average improved by 10 scale points per year. So, a 20 point gain across four tests, representing a 5 point gain on each of the component tests, amounts to about a half year of achievement gains per test. Is a half year of progress a large gain for eight years? It is hardly enough for the US to catch up with international leaders in achievement.<sup>3</sup> However, the gain is more than double that of the eight year period preceding NCLB, 1992–2000, when the composite gain was only 7 points, not 20.<sup>4</sup>

Of course, ESEA is not primarily about the achievement of all students. Its primary target is the disadvantaged, though NCLB brought more students within its sights. As Figure 1 also reports, traditionally low-achieving groups fared well after NCLB’s adoption. Black students gained 30 points, Hispanics 32 points and students eligible for Free or Reduced

Priced Lunch (FRL) 29 points. Those numbers represent about three-fourths of a year of improved achievement per assessment, from 2003–2011. Better gains than the national average and better gains than the years prior to NCLB.<sup>5</sup>

The role that NCLB may have played in promoting these gains is far from clear. Achievement has many causes, including family support and other powerful factors outside of schools. Perhaps certain social influences became more positive during the decade and encouraged greater learning. But, a look at how achievement improved at the state level suggests that social forces are not the explanation. In **Figure 2**, we display the gains made by all students state by state on our composite measure of NAEP scale scores. The range is quite remarkable, suggesting that the states, not national social forces, produced the observed changes in achievement. Two jurisdictions, the District of Columbia and Maryland, gained nearly 50 points—more than twice the national average. Seven states gained more than 30 points, 50 percent above the national average. At the same time, 12 states gained fewer than 10 points, less than half the national average, and two, Iowa and West Virginia, actually lost ground.

**Figure 1. Composite Scale Scores National NAEP: Grades 4 and 8, Reading and Math**

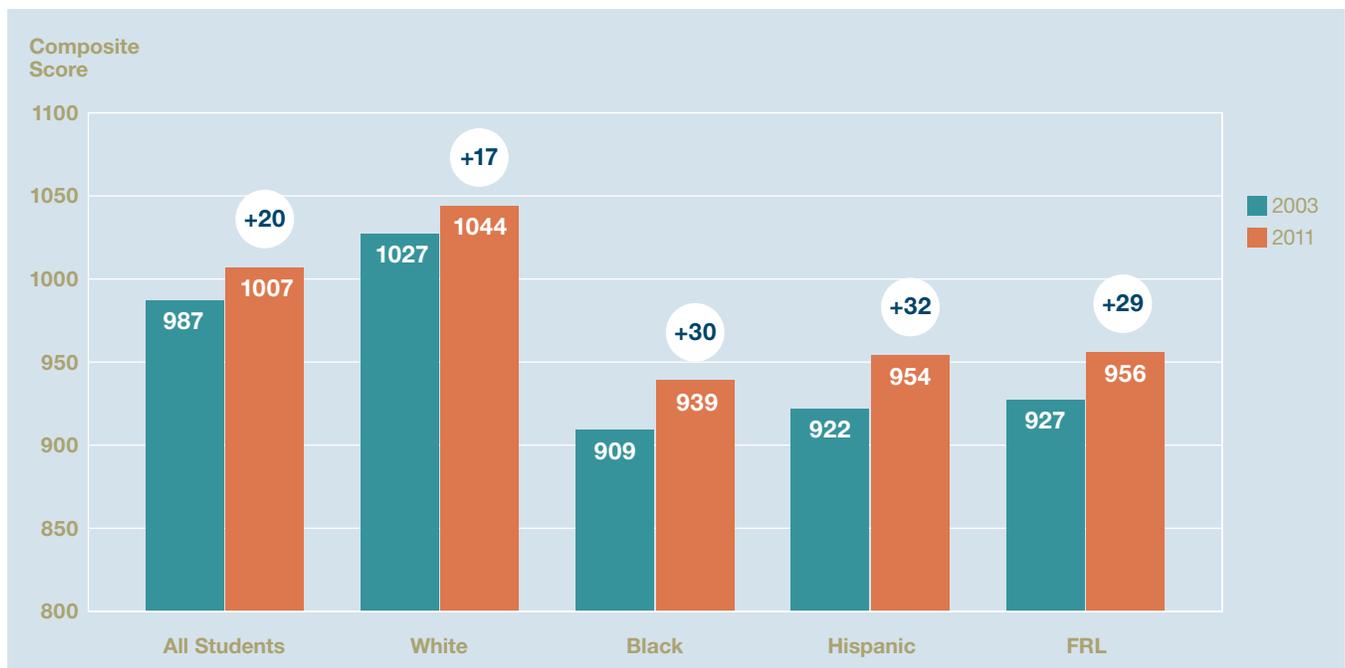
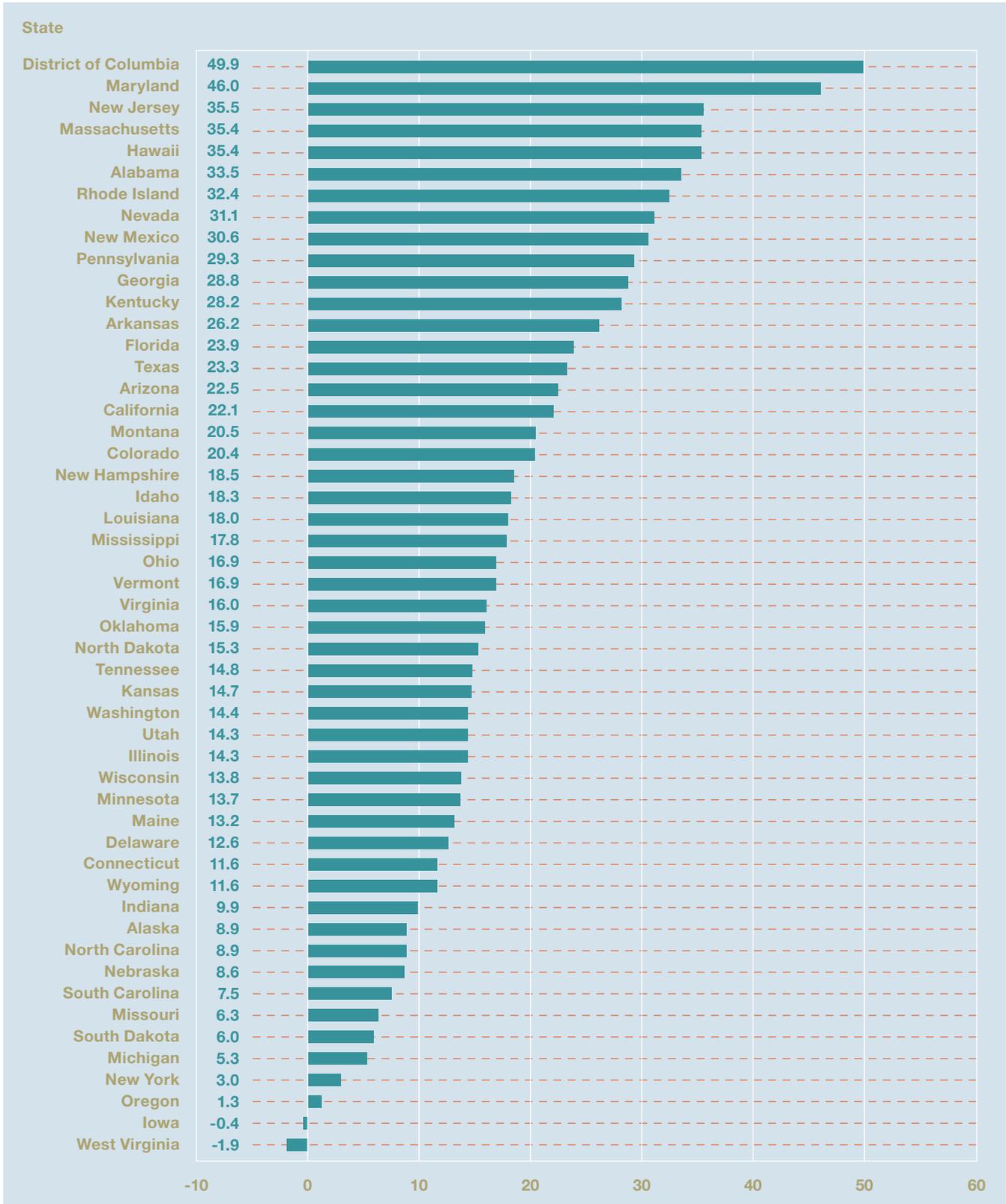
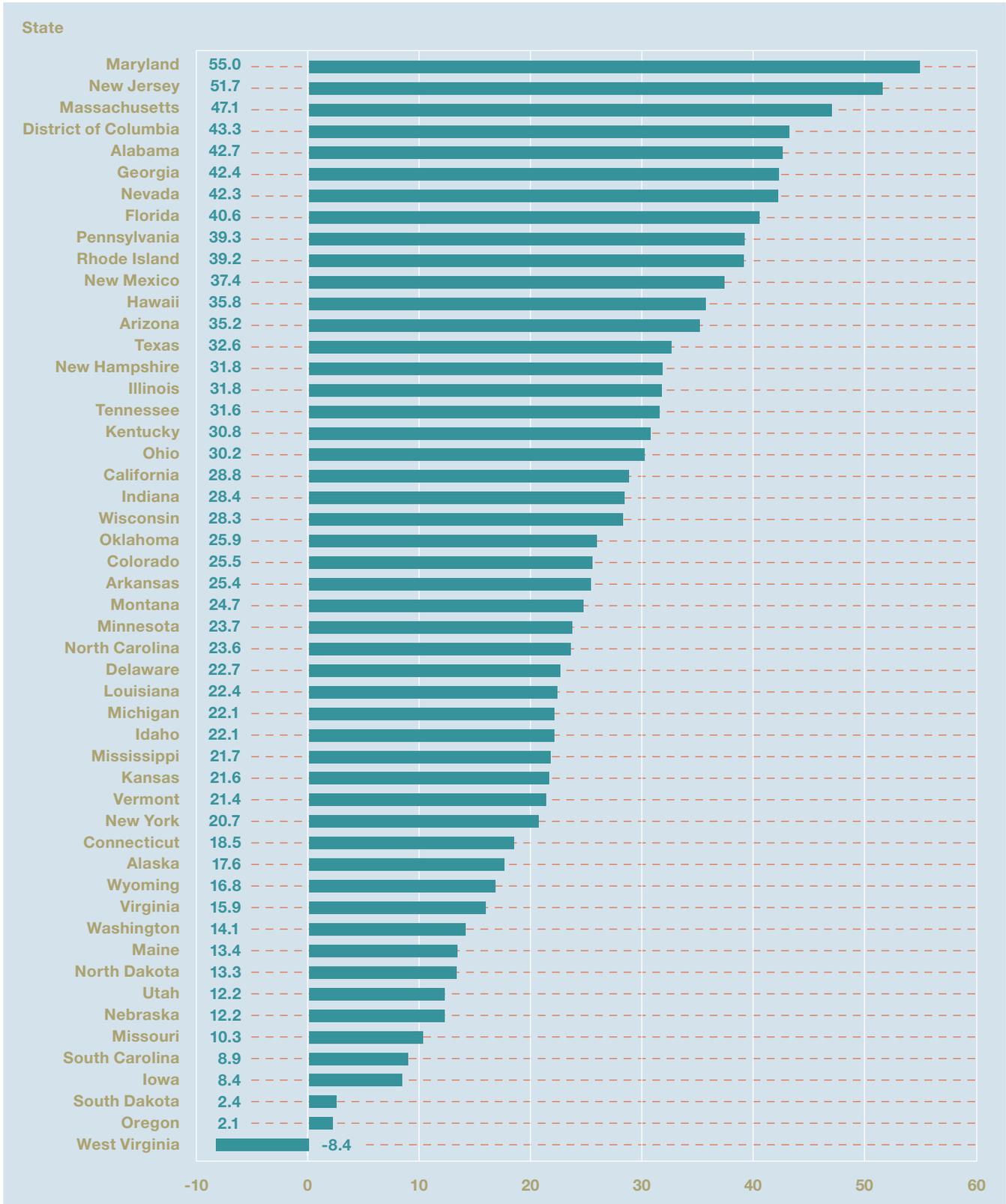


Figure 2. NAEP Composite Scale Score Gains by State, 2003–2011, Grades 4 and 8, Reading and Math, All Students



**Figure 3. NAEP Composite Scale Score Gains by State, 2003–2011, Grades 4 and 8, Reading and Math, Free-and-Reduced-Priced-Lunch Eligible Students Only**



The achievement range is wider still for NCLB's target population, the economically disadvantaged. **Figure 3** reports the gains of each state on our composite measure for FRL students only, the high poverty subgroup that NCLB singles out for special attention. At the high end, students in Maryland, Massachusetts and New Jersey gained approximately 50 points, nearly double the nationwide gain for FRL students. At the other extreme, disadvantaged students in West Virginia fell by 8 points and in 10 states student improvement was less than half the national average. The range in achievement gains is more than 63 points—a year and a half in achievement per assessment.

Let us put this new state achievement gap in perspective. The U.S. has a notorious and persistent difference in achievement between black and white students. Rooted in slavery, segregated schools and a century of social and economic discrimination, the black-white achievement gap is depressingly large. Figure 1 captures just how large. In 2011, white students scored 1044 on our composite of four NAEP assessments; black students totaled 939—a 105 point gap. Recalling that students improve about 10 points per grade level per test, a gap of 105 points translates into two-and-a-half years of achievement. Put another way, a white student midway through sixth grade achieves at about the same level as a black student at the end of eighth grade. In a nation founded on the principle of equality, that is an unacceptably large gap in achievement.

And the nation has long sought to reduce it. A decade before ESEA was passed, the Supreme Court declared separate but equal schools unconstitutional in *Brown v. Board of Education*. Research has looked hard for solutions. But a gap that took two centuries to establish has proven stubbornly resistant to improvement. Over nearly 40 years, by our composite measure, the gap has narrowed by only 25 points—about a half point per year.<sup>6</sup>

All of which makes the new gap between the states alarming. The graphic in **Figure 4** makes the point. ***In just eight years, the states have created an achievement gap that is about 60 percent of the magnitude of the racial achievement gap—that took two centuries to establish.*** In just eight years, the states have created an achievement gap that is nearly 250 percent the magnitude of hard

...A white student midway through sixth grade achieves at about the same level as a black student at the end of eighth grade.

won reductions in the racial achievement gap over 40 years.

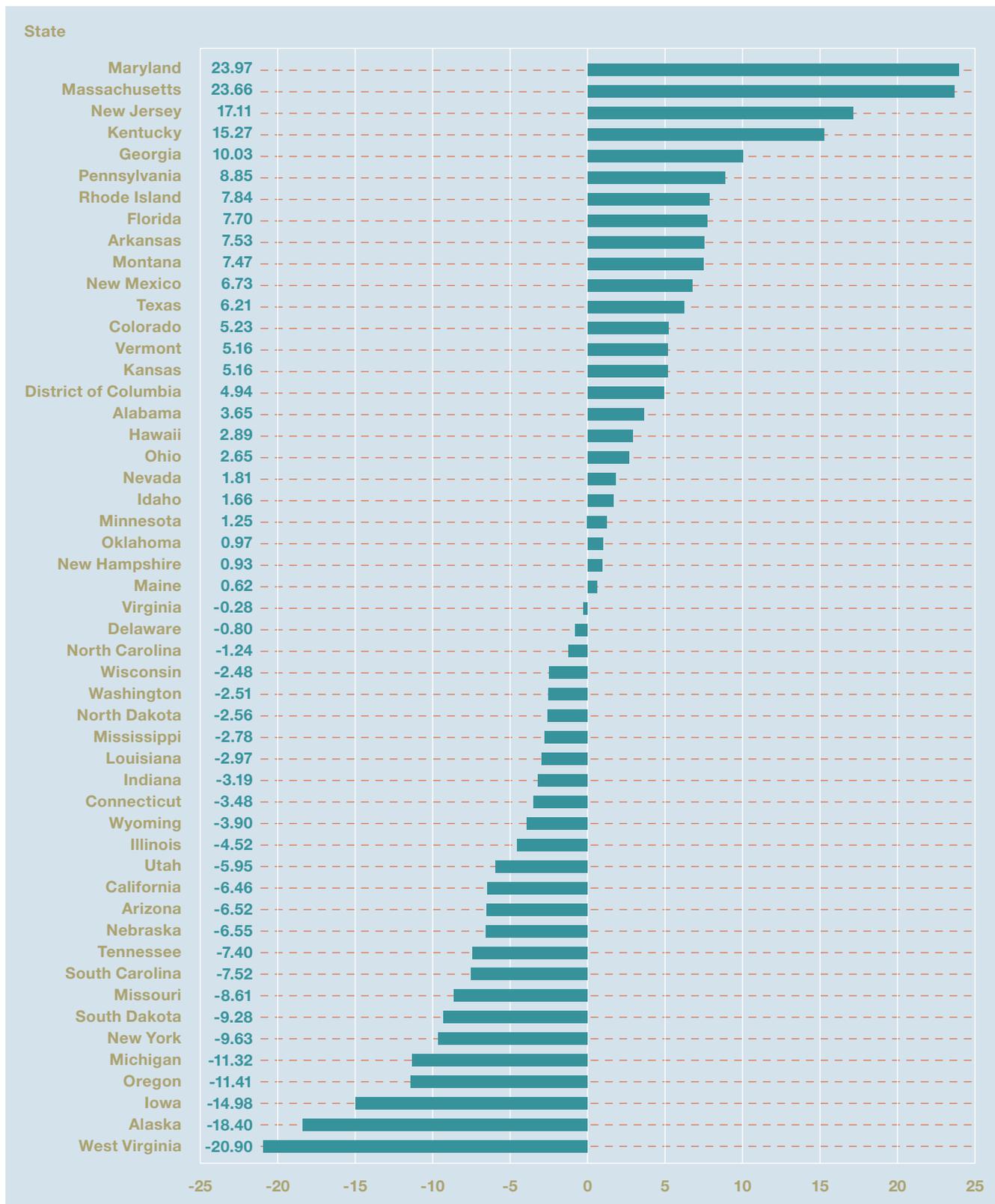
Because student achievement is so heavily influenced by family, community, and other factors beyond the reach of the schoolhouse, it is hard to find schooling making a substantial difference in student achievement. But that is precisely what the state NAEP data indicate. Achievement in some states has been soaring; achievement in other states has been lagging. And the pace of differentiation rivals, indeed exceeds, that associated with America's deepest social division.

The evidence shows that social forces are not the culprit. In **Figure 5** we recalculate achievement gains by the states, controlling for two non-school factors that might affect achievement. We look at the gains for all students by state, from 2003–2011, as in Figure 2. We take into account starting point scores in 2003 and economic disadvantage. It is customary in analyzing achievement progress to control for “regression to the mean,” the tendency for lower scores to go up over time and higher scores to come down. Random variation causes regression to the mean.<sup>7</sup> It is also customary to control for student background, which FRL rates by state accomplish. Figure 5 reports gains as deviations from predicted gains based on starting points and economic disadvantage.

**Figure 4: The New State Achievement Gap in Context**



**Figure 5. Deviations from Expected NAEP Composite Scale Score Gains, 2003-2011, Grades 4 and 8, Reading and Math, All Students, Controlling for Free-and-Reduced-Priced-Lunch Eligible**



## Student achievement is rapidly diverging at the state level and public policy is key.

The controls change the picture remarkably little. A few low achieving jurisdictions show less impressive improvements. The District of Columbia most notably falls from the ranks of the top gainers to 16th, owing to its very low initial score. To be clear, the District raised its composite score very substantially over this period—by 50 points. That is real progress and this analysis should not be taken to imply anything else. As we try to understand how much of the variation in state gains has been caused by schools and state policy, however, our analysis tells us that some of the District's improvement could be predicted from its exceptionally low scores in 2003. Overall, most of the states whose uncorrected scores showed the greatest gains retain their spots at the top of the ranking. Maryland, Massachusetts, and New Jersey were in the top four; now they are the top three. West Virginia was the lowest achiever when we looked at uncorrected scores; it remains in last place after considering its starting point and poverty.

The corrected gap between the highest and lowest gainers is 44.88 scale points. The uncorrected gap was 51.77 (see Figure 2). The similarity of the two estimates is important. Normally, observed differences in achievement gains are substantially accounted for by non-school factors such as initial scores and family background or income. Here, non-school factors account for little of the variation in gains. This means that school-based factors may be having a substantial effect on which students are learning more and which are learning less. We cannot measure those school-based factors directly. Differences in achievement gains across the states may have non-school causes that we have not captured. Random sampling error may also account for some variation. Statewide scores on individual tests are accurate to plus or minus 2.5 points. Differences on our index of less than five points may not be meaningful. But the broad range of differences surely is.

And it's not just statistics that suggest these differences may arise from the quality of schooling.

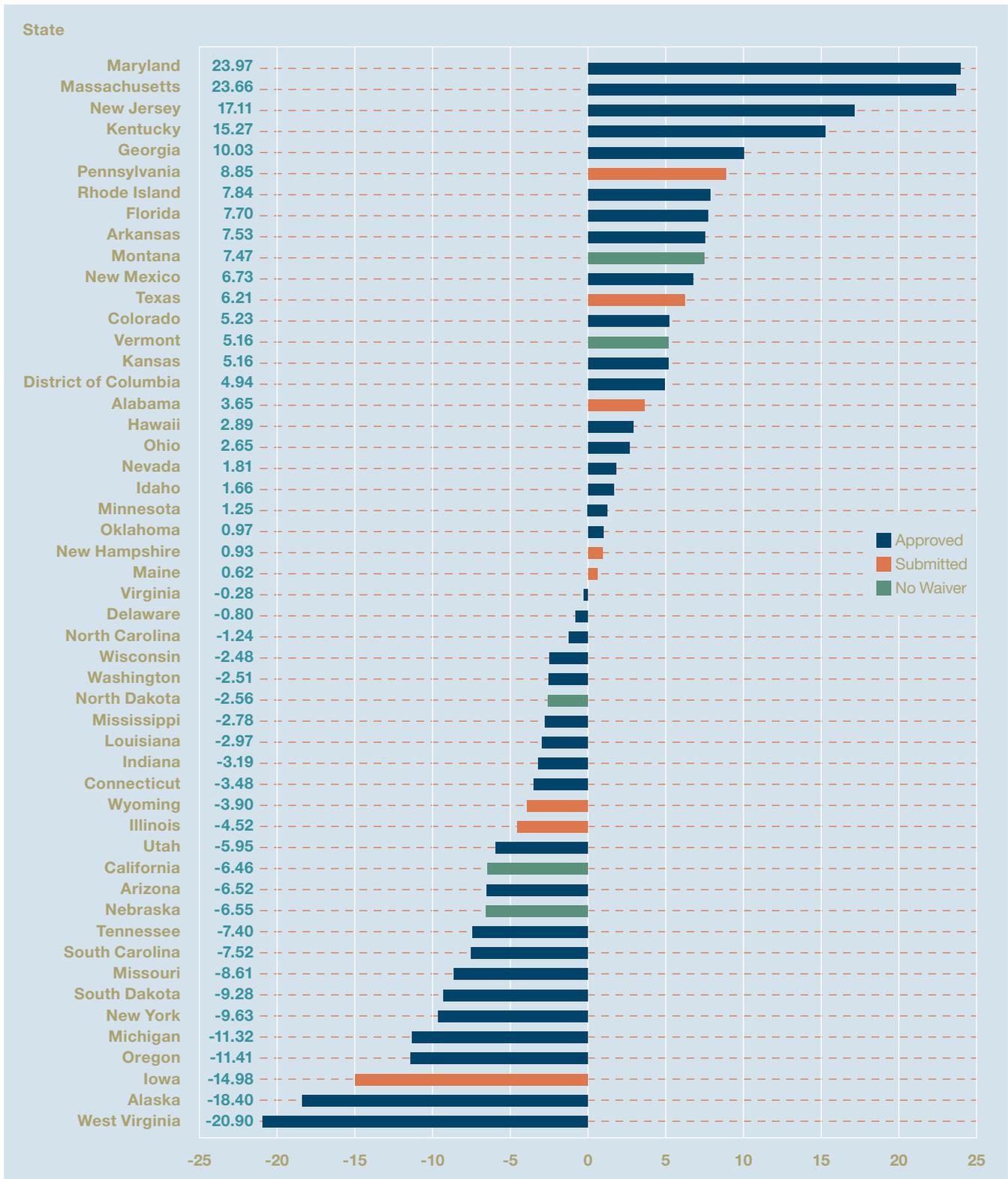
Over the last decade, states with strong corrected gains have been frequently recognized for their aggressive efforts at improvement. Massachusetts has been lauded for its high academic standards and rigorous assessments. Maryland was chosen by the US Department of Education as the nation's top achieving state.<sup>8</sup> Florida has been acknowledged for the progress of its Hispanic students. And the list goes on. With minor exceptions, the same cannot be said for states at the bottom. To be sure, the face validity of the rank order is no proof that state policy is the primary cause of the differences. But coupled with the size of the differences after taking non-education factors into account, policymakers should accept as fact that student achievement is rapidly diverging at the state level and public policy is probably playing a key role.

## Learning from Waivers

Waivers, as we said, are just words—plans and promises. Time will tell whether states have the knowledge and skill to deliver on them. History tells us to expect disappointment. Owing to differences in resources, culture, and values, the states have always produced a range of education outcomes. Southern states had traditionally been lower performers, central and northern states higher performers. Over the last decade, that pattern has begun to erode, with a number of southern states—Kentucky, Georgia, Florida, Arkansas and Texas—among the most rapidly improving, and several traditionally strong states—Iowa, New York, Michigan and Nebraska—advancing little. This new performance is not driven by deep-seeded differences—as our controlled analysis of NAEP gains suggests. It is driven by differences in state education policy. The waivers generally reflect these differences.

In **Figure 6** we report the waiver status of states by corrected gains on our NAEP composite score. One point is clear. Nearly everyone wants a waiver. All but one of the top 10 gainers sought waivers, and all of the bottom 10 sought waivers. To date, 37 states and the District of Columbia have had waiver applications approved by the Department of Education; eight states have applications pending. Only five states have elected to live with NCLB without additional flexibility.

**Figure 6. Waiver Status of States by Deviations from Expected NAEP Composite Scale Score Gains, 2003 – 2011, Grades 4 and 8, Reading and Math, All Students, Controlling for Free-and-Reduced-Priced-Lunch Eligible**



Behind the uniformity, though, lies important variation. States with stronger academic track records under NCLB have won flexibility from the department sooner than those with weaker records. Seven of the 11 states that applied for and were awarded waivers during the first round of applications in November 2011 were among the 15 most improved states. Among the 15 least improved states, two—West Virginia and Alaska, the worst performers—did not apply until round three in September 2012 and required nine months of negotiations for approval; another round-three applicant, Wyoming, is still negotiating. Three other low performers—Iowa, California and Illinois—applied in round two in February 2012 and have yet to be approved.<sup>9</sup> High performing states have tended to apply earlier and win approval more easily than low performing states. The Department does not use state achievement records to determine which states win approval. A state does not have to be a strong performer to have been granted a waiver: 19 states in the bottom half of the performance distribution have ultimately won approvals. The waiver application is highly structured, requiring states to make commitments, from a generally closed-ended menu of options. One hurdle to approval is simply agreeing to replace NCLB's requirements with alternatives specified by the Department.

A compliant state is likely to win approval, regardless of its track record.

But the application also asks states to explain how they will execute the options to which they commit. That is where lower performing states trip up. States with poor achievement records do not already have in place mechanisms to help schools succeed. Their existing apparatus is not working. Failing states must either come up with credible explanations for how they will strengthen what they are doing or come up with new plans altogether. In either case there is a glaring lack of verisimilitude. The Department sees this. The states have legitimate issues with NCLB to be sure, but you can't beat something with nothing.

In contrast, in states where students and schools have been making significant progress, applications have instant credibility. They elaborate policies and plans that states have developed and fine-tuned during the NCLB era. They show how freedom from certain NCLB requirements can make their existing programs even stronger. In many cases they have

a formal track record to document the success of their actions. For the most successful states, the waiver process has often been a way to solidify and expand programs that have been controversial locally, because they do demand change. Successful states can offer schools and districts freedom from the least popular elements of NCLB in exchange for acceptance of tough home-grown measures. In these states, applications may be less about winning freedom from NCLB—the states have succeeded with NCLB in place—and more about going above and beyond the federal law.

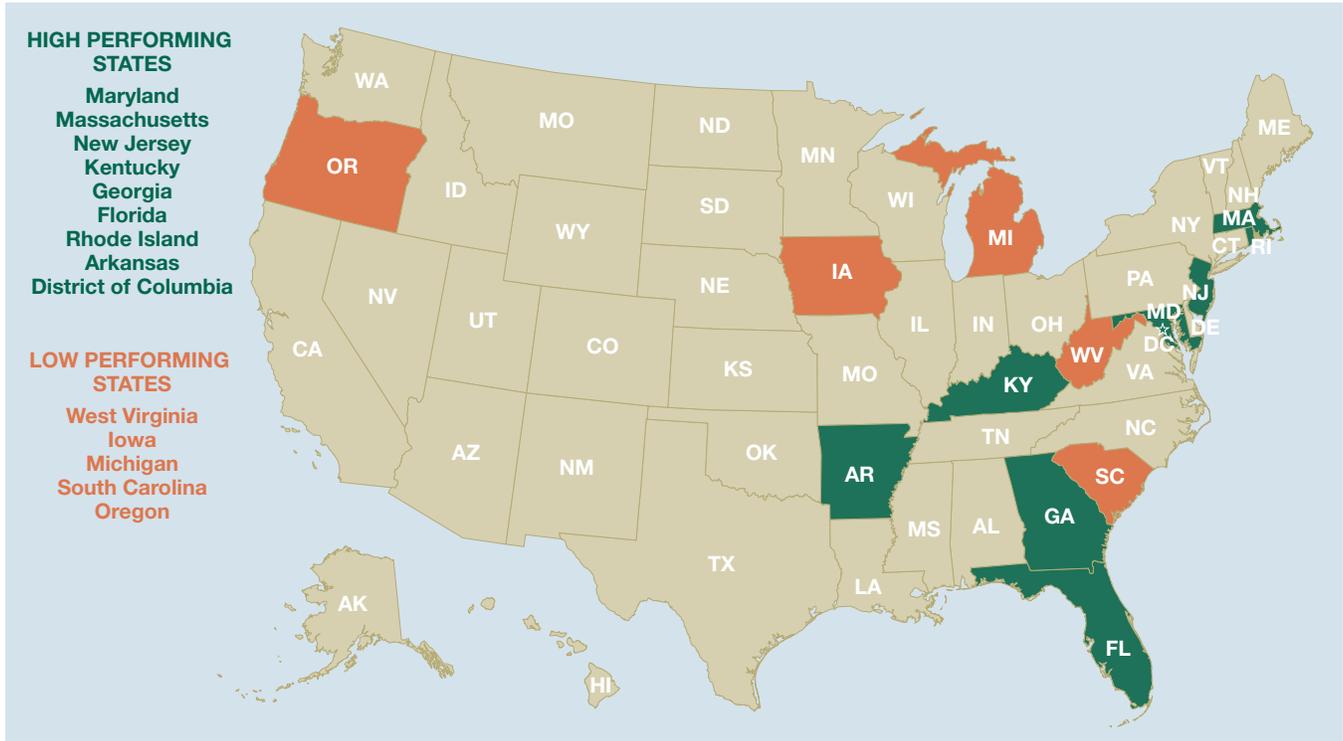
**The states have legitimate issues with NCLB to be sure, but you can't beat something with nothing.**

As part of our analysis of waivers and their implications, we reviewed all approved and pending waiver applications, as of January 2013.<sup>10</sup> The applications of the high performing states are distinguished by their authenticity. Low performing states sometimes offer similar sounding ideas but they are largely driven by the options provided by the Department. Until these states offer evidence that they can execute their plans and deliver for their students, there is little to learn from them—except perhaps a list of what these states see as attractive alternatives to NCLB. Top performing states, in contrast, have been doing much of what they have proposed in their waivers. Top performers have delivered for students, despite the inflexibility of NCLB. They have lessons to teach.

## What the Best Have to Teach

We can most easily appreciate what the top performing states are proposing by comparing their applications to those of the lowest performing states. We chose to summarize all states with approved waivers whose corrected improvement scores for all students were among the 10 best (Figure 6); this

**Figure 7. State Waiver Applications Compared**



became our top performing group. We added to this group the District of Columbia, which had the highest uncorrected rate of improvement for all students (Figure 2). For our low performing group, we chose a demographically diverse sample of states with approved or pending waivers from the bottom 10 of the corrected and uncorrected ratings. We chose to highlight fewer low performing states because, as already noted, their applications largely reflect plans and promises, not proven practices. **Figure 7** shows the states compared.

To receive a waiver from ESEA requirements, states were required by the Department of Education to detail how they would use any requested flexibility to satisfy “three principles.” The principles represent the Department’s interpretation of NCLB’s major aims. With each principle, the Department provided guidelines and options for how states could potentially satisfy it. These “strings” bound states to a range of acceptable future approaches—something critics regard as legislating a new ESEA from the Executive Branch. Close examination reveals quite a bit of variation in state plans. If there is any reason for concern, it is not that the Administration has imposed

its own version of ESEA on the states; it is that the states have been given flexibility to implement ideas with which they have little or no record of success.

***Principle 1: College and Career Ready Expectations for All Students***

The waiver application required states to explain how they would adopt and implement academic standards that ensure students are college and career ready. NCLB left unspecified exactly what state academic standards should accomplish. So, states interpreted its demand for universal “proficiency” in reading and math by 2014 very differently. Some took it as a requirement of only basic skills; others interpreted it as something much stronger. The Administration should be credited for attempting to bring clarity on this point: proficiency should be both something that is achievable for the vast majority of students and meaningful if achieved.<sup>11</sup> “College and career ready” is an eminently practical definition. It is also the definition adopted by the Common Core State Standards, already embraced in some fashion by 45 states and the District of Columbia when the waiver application

was designed. States are free to adopt other “college and career standards” and win waivers, but the Administration clearly banked on the expectation that states would double down on the Common Core, if given the added incentives of NCLB waivers.

High achieving and low achieving states have done just that. They committed to implement the Common Core standards by 2014–2015, when assessments based on them become generally available. But the similarity ends with that shared commitment. **Figure 8** highlights differences in the waiver applications of high and low achieving states. Among those differences are vastly different experiences with high standards such as those of the Common Core. Several of the top performing states already have standards and/or assessments that are the equal of the Common Core, or nearly so. Massachusetts, a long-standing high standards state, plans to adopt the new standards and assessments—but only after verifying that they meet or exceed their current ones. Maryland welcomes the performance tasks that promise to make the new assessments more demanding. Its current assessments have pioneered these tasks for nearly a decade. Kentucky demonstrated its reform bona-fides by administering an early version of the Common Core assessments in 2012, and explaining

to a shocked public why its students were not doing as well as the old tests had indicated.

High performing states are taking steps to ensure that Common Core standards and assessments really do raise the bar. Many of them are benchmarking the new assessments against other measures of college readiness: Do high scores on high school assessments correlate with high scores on the SAT and ACT or on AP and IB exams? Do scores correlate with success on technical exams required for careers or on rates of remedial course taking in college? Florida takes external validation arguably the furthest. It promises to benchmark the performance of its students nationally using NAEP and internationally using Programme for International Student Assessment (PISA) tests and Trends in International Mathematics and Science Study (TIMSS) assessments. Florida aims to perform like the five best states and 10 best nations. The aspirations are less important than the commitment to rigorous benchmarking, a practice that all top states promise. Top states also promise more attention to student success in high school, often setting explicit goals for participation in demanding courses, and more ambitious goals for high school graduation than suggested by waiver guidelines.

**FIGURE 8. State Comparisons**

<b>State Comparison Maryland vs. West Virginia</b>		
<b>Maryland</b>		<b>West Virginia</b>
+46 points	<b>NAEP Gain, All Students</b>	-2 points
+55 points	<b>NAEP Gain, FRL Students</b>	-8 points
PARCC Assessment Consortium	<b>Standards</b>	Smarter Balanced Assessment Consortium
Top ranked Advanced Placement state		Existing standards rated “D” despite being revised
Tests already use Performance Tasks like CCSS		
Waiver: Highly differentiated school performance index	<b>Accountability</b>	Waiver describes new school rating index
Waiver: Longitudinal student data system for teachers		Waiver: School supports and sanctions proposed not yet proven
Waiver: Annual reviews of all teachers and principals	<b>Teacher Evaluation</b>	Waiver: Experienced teachers self reflect, not observed
Waiver: Evaluations based 50% on student tests; 50% on formal observations		Waiver: No specific role for student tests in evaluations

State Comparison Massachusetts vs. Iowa		
Massachusetts		Iowa
+35 points	<b>NAEP Gain, All Students</b>	+0 points
+47 points	<b>NAEP Gain, FRL Students</b>	+8 points
PARCC Assessment Consortium	<b>Standards</b>	Smarter Balanced Assessment Consortium
Existing standards and test earn “A’s”		Existing proficiency standard only 41st percentile nationally
Waiver: Require proficiency on tough state tests for diplomas	<b>Accountability</b>	Waiver: New school performance index
Waiver: By 2017, reduce non-proficient students by half		Waiver: Ten-year path to statewide proficiency
Waiver: Strong supports and interventions for successful and unsuccessful schools		
Waiver: Student test-based teacher evaluations	<b>Teacher Evaluation</b>	Waiver: No student tests in teacher evaluations
State Comparison Florida vs. South Carolina		
Florida		South Carolina
+24 points	<b>NAEP Gain, All Students</b>	+8 points
+41 points	<b>NAEP Gain, FRL Students</b>	+9 points
PARCC Assessment Consortium	<b>Standards</b>	Smarter Balanced Assessment Consortium
Already benchmarking achievement against NAEP and international tests		Historically weak standards
Long-standing school letter-grading system	<b>Accountability</b>	Waiver: New letter grade school-rating system
History of improving “F” schools		Waiver: School interventions and supports not detailed
Waiver: By 2017, aim for all “A” schools to reduce non-proficient students by half		
Waiver: Detailed school interventions and rewards		
Waiver: Teacher and principal evaluations: 50% test scores + 50% formal observations	<b>Teacher Evaluation</b>	Waiver: New teacher evaluation measures being determined

Finally the top states are *all* members of the same assessment consortium, PARCC, the Partnership for Assessment of Readiness for College and Careers.<sup>12</sup> States that have adopted the Common Core have voluntarily signed on with one of two consortia of states to develop assessments consistent with the Core and to set performance levels—i.e., to define what “proficiency” really means. Guided by assessment experts and supported by testing companies, the two state

consortia will provide alternative interpretations of what more rigorous assessments should look like and what test scores or “cut points” constitute meaningful readiness for college and careers. These processes are subjective and inevitably contentious. It remains to be seen how tough the consortia will be. It is striking—and encouraging—that states that have made most progress dominate one consortium. Indeed many of the top states are leaders of the PARCC consortium. Agreement on

increased rigor will be easier for states that have a history of already biting that academic bullet.

As Figure 8 shows, lower performing states come up short in many ways. Their existing standards and assessments have been mostly of the basic skills variety—West Virginia and Iowa encouragingly acknowledge this. Weaker states sometimes hedge just how much of the Common Core they will embrace in assessments. They do not have substantial plans to benchmark new assessments against external measures. They are largely members of the other testing consortium, Smarter Balanced Assessments. It will be difficult for a consortium dominated by states with less experience and success with academic rigor to now embrace it. To be fair, the plans of the low performing states to raise standards have their virtues and have often passed muster with the Department. But compared to the plans of the top states, they are not robust nor backed by much experience.

## ***Principle 2: State-Developed Differentiated Recognition, Accountability and Support***

NCLB has been roundly criticized for its crude approach to accountability. Schools only get credit for students scoring proficient on reading and math exams—no credit below that bar or above it. By 2014 schools must get 100 percent of their students over the bar, and along the way, increasing percentages of students of every type—FRL, racial subgroups, English Learners—must succeed. If any one group fails, the whole school fails. Fail two years in a row and sanctions set in. With each successive year of failure a prescribed fate awaits, regardless of the circumstances of failure. The system surely got the schools' attention, and some students have probably benefitted, as widespread gains during the NCLB decade suggest. But far too many schools—45 percent at last count—have been branded failures, strong students have received little encouragement, and most important, schools with persistently poor performance are not being turned around at acceptable rates.<sup>13</sup> Meanwhile, 2014 fast approaches, and no state with credible academic expectations is close to universal proficiency.

The waiver process allows states to reset their goals and to adopt more nuanced accountability plans. Here, too, the top states distinguish themselves. It begins with the goals themselves. The top states, as shown in Figure 8, intend unambiguously to implement tougher assessments and higher student performance standards. They also promise to continue the march toward 100 percent proficiency. The Department offered states three options for re-setting their goals. The vast majority of the top performing states stuck with the NCLB focus on proficiency. They chose “Option A,” which requires a reduction by half in the percentage of all students not yet proficient on reading and math assessments, within six years—or by 2016–2017. The requirement applies to each subgroup in each school, a goal that demands more rapid progress by the lowest achieving students. Because these states already have made substantial progress, the goals, if achieved, will raise achievement in most schools in those states to 85 or 90 percent proficiency. This is not universal proficiency, and lagging subgroups will still lag somewhat—a point of criticism among some advocates for the disadvantaged. But based upon higher standards as well, it represents clear progress.

**The waiver process allows states to reset their goals and to adopt more nuanced accountability plans.**

Lower performing states generally eschewed the clarity of Option A—which promises straightforwardly that more students will achieve proficiency. They generally opted for Option C, which allows states to design their own goals. The states argue—correctly—that proficiency scores on reading and math exams are too narrow a gauge. Goals should reflect progress toward and beyond proficiency. Goals should include other outcomes such as high school graduation rates and scores on other assessments, most notably science assessments. States choosing Option C propose to set goals for indexes of

**Time will tell whether complex goals promote school improvement. For now we know that successful states largely have chosen simplicity.**

performance. In concept this all makes sense. And, one high performing state, Massachusetts, chose Option C as well. But there is a risk, especially for low performing states. More complex goals can cloud rather than clarify school performance. More indicators can make it more difficult for schools to focus and more difficult for parents, the public, and policymakers to appreciate progress. More indicators can compromise transparency not enhance it. Top performing states, as we shall see, address this complexity, but they have kept it out of their goal setting: they value their new proficiency standards and have the focus sharply on them. Low performing states claim that they will embrace new proficiency standards, but then they dilute them with indexed goals. Time will tell whether complex goals promote school improvement. For now we know that successful states largely have chosen simplicity.

Top performing states leave the nuance for their support and intervention. While they want schools to keep their eyes on the prize—universal college and career readiness—they recognize that schools will fall short in different ways. States that have had the most success over the last decade have already found ways to differentiate sanctions and supports, based on more than just how many students are proficient. Florida has been the pace-setter here. Going back more than a decade, Florida launched and then fine-tuned a school grading system. The state grades each school on an A–F scale, using multiple indicators, including current test scores and growth in individual level scores, each comprising 50 percent of the assessment component of the grade. Grades include progress by the lowest 25 percent of all students. High school grades demand student participation in advanced curricula and high graduation rates. The state logically includes a

measure of college and career readiness as well. Test participation requirements are high; a school cannot earn an “A” without 95 percent of all students taking state exams, a high hurdle for high schools. The grades provide understandable measures of school performance for parents, policymakers and schools themselves—and have earned generally high marks from observers.

The other top performing states have their own multi-indicator grading systems. They may use numeric scores instead of letter grades, but conceptually they align with Florida. Massachusetts rates schools from Level 1 “on track to college and career readiness” to Level 5 “chronically under-performing,” using the Composite Performance Index or CPI. Scores on the CPI include performance against proficiency levels on the widely respected state MCAS exam, plus performance against school specific annual progress goals—allowing low achieving schools to earn points for improvement. Extra credit is awarded for helping both low achieving students gain and high achieving students advance—the latter overlooked by NCLB. Graduation rates and their improvement likewise earn credit for absolute performance and progress. Maryland, Georgia, and each of the other high flyers has a similarly nuanced index. Each attempts to portray schools in a more balanced and complete way, distinguishing schools with the most serious problems from schools with more limited challenges. Each also has the virtue of a basis in state experience. These states have improved schools before; they aren’t promising to create better schools from whole cloth. Top states propose consequences for schools, based on the index—not on annual measurable objectives (AMOs). This may seem odd, to promise that students will achieve increasing levels of proficiency (Option A), but not base consequences on those promises. But that is precisely the point. The indexes offer more refined measures of what exactly may be keeping a school from meeting its AMOs. NCLB invoked only AMOs in meting out consequences, and left many schools feeling bludgeoned. Indexes enable states to classify schools for different treatment. Following Department guidelines, states must identify “priority” schools—the lowest and chronic poor performers; “focus” schools—the schools with the worst achievement gaps though not the lowest performers; and “reward” schools—the state’s top performers.

Every state seeking a waiver offers some scheme for identifying priority, focus, and reward schools. The difference between the historically successful states and the others is how they deal with the schools, once categorized. In general states have struggled to turn around the very lowest performing schools, so no state holds the key here. But the top performing states are nonetheless different. They have measures in place to make tough personnel moves, removing principals and teachers if schools fail to turn around. They have developed state capacity to intervene. New Jersey, for example, has assembled new regional teams, outside the traditional bureaucracy, to drive achievement.

Top states prescribe credible school transformation models centrally. Low performing states are more likely to defer to locally developed plans, carrying on NCLB's improvement planning tradition. Top states mandate staff training programs that recognize the failure of traditional professional development as superficial and lacking in accountability. The efficacy of professional development seems to be taken for granted in the weakest states. The best states often have specific interventions for students. Georgia developed a statewide tutoring plan, the Flexible Learning Program, to replace NCLB's Supplemental Education Services. While none of the top (or bottom) states maintained NCLB's school choice option (the right for students in schools failing to make AYP for three consecutive years to transfer to non-failing public schools), most of the top states have dynamic charter school sectors already offering choice. Finally, top states offer generous financial incentives to high achieving schools and rapidly improving ones. And, they attend to all schools, not just those in a federal category. As Florida and Massachusetts illustrate, their goal is for all schools to make an "A" or reach Level 1—to produce students who are college and career ready.

**The top performing states ...  
have measures in place to  
make tough personnel moves,  
removing principals and teachers  
if schools fail to turn around.**

Low performing states often promise some of the same measures as the best do. But, their promises lack the specificity that comes from experience in the most successful states; they lack firm grounding in the frustrating reality of school improvement. It's easy to create a school performance index, classify schools, and assign logical sounding consequences. The low performing states with approved waiver applications have done so. But their plans do not read like those of their more battle-tested counterparts. In a word, they lack evidence that in their states, with their capacity, they can actually execute.

### ***Principle 3: Supporting Effective Instruction and Leadership***

The weakest provisions of NCLB are those that require "highly qualified teachers" in every public school classroom. Most schools in the country have long since satisfied the requirements—all of their teachers are "highly qualified." But our schools continue to achieve less than they should. "Highly qualified" teachers have not materially improved school quality. The reason could have been predicted. NCLB required new and experienced teachers to demonstrate they are "highly qualified" with evidence that research has proven does not predict or guarantee any level of quality on the job: state certification, passing Praxis teaching exams, or completing Continuing Education Units.

The waiver process takes NCLB's aspiration for better teachers down a more promising path. Reinforcing incentives provided in Race to the Top grants and before that stimulus funds, waivers encourage states to look at the actual performance of teachers (and principals) and not merely at their qualifications. This makes sense. Research has yet to identify strong predictors of highly qualified teachers and principals. Research has identified pretty powerful measures of performance on the job. The waiver application asks states to adopt more valid and reliable measures of on-the-job performance, and to use those measures to improve human capital, especially in low performing schools.

Here, as seen in Figure 8, the high and low performing states differ sharply. High performing states tightly embrace the use of student assessment data to judge the work of teachers and principals.

Virtually all warrant that they will use estimates of the “value added” by educators to student achievement in their teachers’ formal personnel evaluations. They also commit to specific contributions, most requiring that school districts make test scores 50 percent of evaluations. Low performing states are noncommittal about the role of student achievement.

High performing states also endorse the use of detailed instructional frameworks for teacher evaluation. Research has shown that teachers who effectively follow certain pedagogical models produce greater student achievement gains than those who do not.<sup>14</sup> High performing states are asking districts to adopt frameworks validated with student achievement data, or to validate their own. Top states are mandating multiple formal observations of new teachers each year and annual observations of teachers with continuing employment (tenure). Top states are typically requiring districts to make these new formal observations the other 50 percent of evaluation scores.

The best states are adopting four or five point evaluation scales. Teacher evaluation is notoriously indiscriminating; teachers are either satisfactory or unsatisfactory—and 99.9 percent of tenured teachers are satisfactory. Top states are trying to avoid scales calibrated in different degrees of wonderfulness, labeling at least two categories of performance as clearly inadequate. Clarity is important if the tradition of grading all educators as successful—even as their students are failing—is to be broken. The state of Washington, a middling achiever, recently introduced a four point evaluation scale, and even in the lowest performing schools, almost no one received a critical evaluation.<sup>15</sup>

The lowest performing states are generally less committed to evaluation reform. Their plans vary, and some are definitely better than others—Michigan’s for example. But compared to the top performers, they are less specific about how student achievement will be employed. In fact, there are a number of things they are not doing. They are not promising to adopt validated instructional frameworks for evaluation. They are not prescribing evaluation formulae. They are not committing to the annual objective evaluations of veteran teachers. (In West Virginia, teachers with six or more years of experience will do self-reflections.) They also are not on board with

**Clarity is important if the tradition of grading all educators as successful—even as their students are failing—is to be broken.**

a provision that the most successful states have adopted across the board—the annual evaluation of administrators using the same academically rigorous measures as for teachers.

Experience is the final differentiator. Several of the top performing states were awarded Race to the Top grants, beginning in 2010. They have been developing new evaluation systems as part of those grants. Georgia, for example, has launched its demanding “Keys” system that includes test scores, formal observations, and student surveys. Maryland has installed a comprehensive human capital program that uses a robust and individually tailored professional development program to make use of the new evaluations. New models of evaluation may leave much to prove, but they are based solidly on research. And, once again, it is the top performing states that have the most experience with them. Succeed or struggle, they are the states that teacher quality enthusiasts should be watching.

## **A New Middle Course to Equity**

The federal government is better positioned than the states and local communities to ensure that economic circumstances do not determine educational opportunities. The federal government is unquestionably best at redistributing resources. But money is not enough, and the history of ESEA has shown consistently that money does not necessarily benefit the neediest without strong federal guidance. Or with it. And that is the rub.

Years of federal screw tightening have not improved the education of the disadvantaged—not even close to what policymakers hoped would happen. Nearly 60 years after ESEA’s historic enactment, we are

a long way from fulfilling the promise of President Johnson’s soaring rhetoric.

However, we are also closer to understanding how to fulfill that promise. The new state achievement gap, though disturbing to behold, provides encouragement. It shows that some states are reforming their education systems substantially. It shows that education policy can promote student achievement for large numbers of schools, and that gains are not all about economic circumstance. It is stunning that geography now plays a growing role in how well students learn; one might think that the Internet Age would diminish the importance of location. Left unchecked, the divergence of state education systems could easily continue, thwarting the nation’s pursuit of equal educational opportunity.

The federal government should learn from the new state achievement gap, and put the nation on a more equitable course. Two steps are paramount. First, the federal government should use the time during which waivers are being implemented to study—rigorously, scientifically—what state measures are associated with achievement gains. Second, it must hold states accountable for their implementation of ESEA by insisting on improved achievement on NAEP, or if adopted, the new Common Core assessments.

## 1. Finding What Works

The first states to watch are the recent high achievers. Their waiver applications offer a kind of informal consensus among successful state reformers about what seems to work. But, more time is needed to observe and analyze new measures and their consequences. Research is inexpensive relative to overall federal education spending; it could also improve the efficacy of spending down the road.

Waivers are only granted for two years before they must be renewed. The Department could use what it learns from research now to decide which applications have merit for renewal. The more successful reform is understood, the more future reform can be guided by proven principles. The Department should be systematically planning the research it will need to conduct or sponsor to gauge what works. Research should shape any reauthorization of ESEA. If effective practices can be identified, they could be codified in the next version of the law, and applied to all states.

The federal government should be careful, however, about concluding that it has things all figured out. There is surely an impulse after considering the state achievement gap to say, “Don’t trust the laggards; don’t give them any more leeway. Their performance hasn’t earned them the right to be trusted with flexibility.” But this would not only be a mistake; it would be hubris. States that did not improve during the NCLB decade failed while following a uniform prescription. Re-administering that medicine holds little promise. And, it is premature to say that the high performing states can safely be followed; we do not understand their successes or the applicability of their reforms to other states. Ironically, the best course for the struggling states is for the federal government to let them experiment, too. They lack experience with success and their plans appear less ambitious. As a nation, we still have too much to learn to foreclose options.

**The first states to watch are the recent high achievers. Their waiver applications offer a kind of informal consensus among successful state reformers about what seems to work.**

## 2. Demanding Accountability

This does not mean NCLB should be replaced by a free-for-all among the states. After providing flexibility and committing to studying the consequences, the federal government should demand accountability by requiring improved achievement on NAEP or the new Common Core assessments. As we have seen, NAEP is capable of registering substantial differences in a short time. NCLB authorizes benchmarking state standards with NAEP. Use this authority. As waivers are granted, the Department could make renewal contingent on NAEP progress demanding that states show that their innovations are helping students improve on independent assessments or following practices

**By searching for the one-size-fits-all solution, the federal government has alienated states that think and know they can do better.**

already proven in higher performing states. NAEP could be expanded to include high school testing and perhaps other subject testing. In time, if states adopt Common Core assessments, those could be used for accountability as well. For the next several years, NAEP is the only available measure of state progress.

As Congress looks for a workable new version of ESEA, it could consider allowing states to choose one of two paths for implementation. One path

would follow the research-based practices of the most successful states, over the last decade and the next few years. Another would allow states to chart their own courses, subject to NAEP accountability. In any case, the federal government would recognize that it will never have all the answers, and states would be free to lead the way—so long as all students achieve.

By searching for the one-size-fits-all solution, the federal government has alienated states that think and know they can do better. By shifting too much authority to the states, the federal government risks empowering states without the knowledge or wherewithal to serve those most in need. The federal government has always been right to take upon itself the role of the great equalizer. It should steer a middle course to get us there—using the states as laboratories to identify those measures that reduce inequality, but giving states the freedom to experiment—if they are willing to be held accountable for what students achieve.

## Endnotes

1. *Full Committee Hearing—No Child Left Behind: Early Lessons from State Flexibility Waivers*: Hearings before the US Senate Committee on Health, Education, Labor and Pensions. US Senate. 113th Cong. (2013) (testimony of Secretary of Education Arne Duncan). Retrieved from <http://www.help.senate.gov/imo/media/doc/Duncan2.pdf>
2. *Full Committee Hearing—No Child Left Behind: Early Lessons from State Flexibility Waivers*: Hearings before the US Senate Committee on Health, Education, Labor and Pensions. US Senate. 113th Cong. (2013) (testimony of Andy Smarick). Retrieved from <http://www.help.senate.gov/imo/media/doc/Smarick.pdf>
3. Eric A. Hanushek, Ludger Woessmann, and Paul E. Peterson, “Is the U.S. Catching Up?” *Education Next*, (Fall 2012), Vol. 12. No. 4, pp. 1–10.
4. The NAEP scores in our analysis are from the “Main NAEP,” where longitudinal comparisons must be done with some care owing to slight changes in test content over time. Because state data are only available for the Main NAEP, we had no option but to use those data. For national measurement purposes, NAEP also includes a “Long Term NAEP,” with invariant questions. It is not administered at the same time as the Main NAEP, and does not allow us to capture the pre-NCLB and NCLB eras as cleanly. Nevertheless, the Long Term NAEP corroborates the gain story from the Main NAEP. For the period 1990–99, our long-term composite measure increased 13 points; for the subsequent period, 1999–2008, the measure increased 25 points.
5. Loveless et al, Brown Center, et al., (2012) How Well Are American Students Learning? *Brown Center on Education Policy at Brookings*. Washington D.C.: The Brookings Institution.
6. Based on NAEP Long Term Trend data, the white composite scale score was 950 and the black score 806 in 1971. In 2008, the most recent trend point, white scores totaled 1060, black scores 941. The racial score gap narrowed from 144 to 119 over this nearly 40 year span.
7. So too do “floor and ceiling effects,” the difficulty of high scoring students to increase their number right on a test compared to the ease of low scoring students to do so—though NAEP has not presented substantial evidence of this problem.
8. Chandler, M.A. (2012, July 17) Maryland School Achievement Growth is Tops in the Nation, Report Finds. *The Washington Post*. Retrieved from [http://www.washingtonpost.com/blogs/maryland-schools-insider/post/md-student-achievement-growth-is-tops-in-the-nation-report-finds/2012/07/17/gJQAF4RSrW\\_blog.html](http://www.washingtonpost.com/blogs/maryland-schools-insider/post/md-student-achievement-growth-is-tops-in-the-nation-report-finds/2012/07/17/gJQAF4RSrW_blog.html).
9. California withdrew its application after the department rejected it.
10. The findings in Figure 6 are current as of May 20, 2013. The sample for in depth research was drawn in January 2013.
11. This definition is also recommended in John E. Chubb, *Learning from No Child Left Behind: How and Why the Nation’s Most Important But Controversial Education Law Should Be Renewed*, Stanford: Hoover Institution Press, 2009
12. States ranked 1–9 are all in PARCC.
13. Andy Smarick. (2012, November 19). The disappointing but completely predictable results from SIG. Retrieved from <http://www.edexcellence.net/commentary/education-gadfly-daily/flypaper/2012/the-disappointing-but-completely-predictable-results-from-SIG.html>
14. MET Project. (2013). *Ensuring Fair and Reliable Methods of Student Teaching: Culminating Findings from the MET Project’s Three-Year Study*. Bill and Melinda Gates Foundation.
15. Aldeman, Chad. (2013). The Evergreen Effect: Washington’s poor evaluation system revealed. *Education Sector*.