Why Standards Produce Weak Reform

TOM LOVELESS, BROOKINGS INSTITUTION

JANUARY 16, 2018
Standards define what students will learn and when they will learn it, establishing common goals to guide an entire educational system’s reform efforts. The attractiveness of the idea rests on its simplicity. Education policies of both the Bush and Obama administrations were predicated on a belief in standards-based reform, but the affinity for standards pre-dates these two administrations and extends beyond Washington. Standards-based reform originated in the states and has enjoyed nearly three decades of bipartisan political support.

This chapter argues that the Bush-Obama policies have exposed inherent weaknesses in standards-based reform. The weaknesses stem from the organizational and political obstacles that standards face in the transition from statements of ambitious learning goals, created in upper level policy forums, to accurate descriptions of the student learning that takes place in schools and classrooms.

How Did We Get Here? Standards before Bush and Obama

Standards have evolved incrementally, with each successive iteration correcting a supposed flaw in previous versions. To understand key elements of the Bush-Obama approach to standards, one must consider the standards that came earlier. States created curriculum frameworks in the 1980s and early 1990s, delineating what was to be taught and learned in each subject. A 1992 survey found that 35 states had adopted curriculum frameworks in English Language arts, and 38 in mathematics.¹ Most states focused on minimum competency and basic skills, but a few stood out as exceptions.² California’s State Superintendent of Public Instruction, Bill Honig, marshalled brand new curriculum frameworks (frameworks had been in existence since at least the 1960s) to overhaul K-12 subject matter, the state’s California Assessment Program, textbook adoption, and professional development programs.³ This multi-pronged approach, in which state
policies are designed to reinforce each other in the pursuit of ambitious learning objectives, came
to be known as systemic school reform or standards-based reform.

An influential 1989 article by Marshall Smith and Jennifer O’Day articulated the basic
theory of systemic reform. Smith and O’Day argued that the multi-level educational system
produces policy fragmentation, reform efforts that overlap or contradict each other. They
criticized 1980s state-level reforms for emphasizing “more of the same”—longer school days,
increased graduation requirements, more testing, and an emphasis on learning basic skills. They
criticized reforms attempting to restructure schools for being episodic and uncoordinated,
exhibiting a “project mentality” that tackled problems individually on a school to school basis.
As a 1991 policy brief from the Consortium for Policy Research in Education explained, “The
first step in developing a coherent system of instructional guidance is to work toward agreement
on a core body of challenging and engaging knowledge, skills, and problem solving capacities as
goals for all students. All state policies guiding instruction would be based on these goals,
forming a consistent, supportive policy structure for school improvement.”

Professional Groups

Professional curriculum groups assumed an important role in standards writing in the 1990s as
the U.S. Department of Education funded several organizations to write national standards. The
Department repeatedly stressed that the standards were voluntary. The National Council of
Teachers of Mathematics (NCTM) had released *Curriculum and Evaluation Standards for
School Mathematics* in 1989. Although not funded with federal money, the NCTM standards
became a model for other subjects. The Bush administration awarded grants for writing
standards in science, history, geography, foreign languages, the arts, English, and civics. The
Clinton administration followed with Goals 2000, which called for voluntary national standards. Goals 2000 created a federal agency, the National Education Standards and Improvement Council (NESIC), to certify the standards crafted by subject matter organizations and the student assessments created by individual states. In 1996, with Republicans now in control of both houses of Congress—and with party members growing suspicious that NESIC’s validation powers invited federal interference with state autonomy—legislation passed that abolished NESIC before even a single member had been appointed.7

**Political Opposition**

Political opposition to standards projects grew throughout the 1990s. The national history standards, released in 1994, spurred a backlash from conservatives who charged the document with political correctness. But the controversy transcended partisanship. A U.S. Senate resolution condemning the history standards passed by a vote of 99-1. Other standards projects also struggled. The Department of Education cut off funding for the English-language arts standards in 1994, citing delays in writing and the vague wording of the standards in early drafts. The two organizations responsible for the standards, the National Council of Teachers of English and the International Reading Association, continued the project using their own funds. When the ELA standards finally appeared in 1996, they drew harsh criticism. An editorial in the New York Times acidly criticized the jargon-filled standards, the review beginning with: “A curriculum guide for teaching English has just been released in a tongue barely recognizable as English.”8

Even support for the seemingly non-political NCTM math standards wavered. As new textbooks and materials in accord with the NCTM standards appeared in classrooms, charges
surfaced that the standards promoted “fuzzy math,” calculator use over computation skills, and a
constructivist education philosophy embracing many century-old, child-centered practices of
progressive education that had always been viewed warily by the public. In one of the earliest
examples of the internet’s potential for political organizing, two groups that included scientists
and mathematicians—many of them life-long Democrats--founded *Mathematically Correct* in
California (1997) and *HOLD* in New York City (2000) as websites for advocating resistance to
NCTM-style math reform.⁹

**Lessons for Standards in the 21st Century**

As the new century dawned, several standards projects had been battered politically, but the
basic idea that standards could improve education remained strong. The theory was right, a
common argument asserted, only the details of implementation had gone wrong. The Bush and
Obama administrations attended to the lessons learned from earlier projects as they designed
their own standards-based reforms.

It was not enough to promise “voluntary” standards, as the federal government had to be
kept far away from—or at least out of view of—the standards. The word “national” also must
be handled delicately; better to differentiate “national” from “federal” and stress the state-level
origins of standards based policies. The NCTM standards taught that political support cannot be
taken for granted, even if standards documents are greeted with widespread acceptance upon
initial release. Standards must focus on content, not pedagogy. They should define specific
expectations for learning skills and knowledge in each grade, not promote esoteric educational
principles or embrace a particular philosophy. Following Smith and O’Day’s recommendations,
the 1990s standards of professional groups had clustered learning goals into three or four year chunks. They also promoted progressive pedagogy as instructional reform.

Bush and Obama’s policymakers generally agreed that standards should focus on challenging material and seek ambitious accomplishments, not merely the acquisition of basic skills or minimum competencies. And the standards must have real consequences attached to them to move the system. Simply encouraging states to take “corrective action” with failing schools, as 1994’s Improving America School Act had done, allowed many states to avoid adopting meaningful accountability systems.

Federalism and the Bush and Obama Approach to Standards

As governor of Texas in the 1990s, George W. Bush had embraced and extended the standards-based policies of preceding governors in his state, both Democratic and Republican. When he took office as the 43rd President of the United States, Bush brought Texas school reform with him. The Bush 2000 presidential campaign used the candidate’s education record as an issue to illustrate “compassionate conservatism,” with the hope of prying away centrist voters and weakening the Democratic Party’s longstanding political advantage on domestic issues. Once in office, Bush’s rhetoric promoting the No Child Left Behind Act—that the nation must end the “soft bigotry of low expectations”—wedded ambitious educational standards to the cause of civil rights. Closing gaps became a national goal. As Patrick McGuinn documents, this argument proved important in convincing powerful liberals in the U.S. Senate, including Sen. Ted Kennedy, to relinquish opposition to testing and accountability.10

NCLB increased federal involvement in K-12 education, a matter historically—and constitutionally--left to the states. But the new law gave a nod to federalism by allowing states
considerable leeway in how they fleshed out the details. States would be free to adopt their own standards, to create their own tests, and to decide for themselves where to draw the line defining “proficient” student performance. The requirement that states enforce a goal of 100 percent student proficiency in reading and math by 2014 meant that a steadily rising number of schools fell short. By 2007, public support for the law collapsed.

The Obama administration took a different approach to federalism. Adopting common college and career ready standards, meaning the Common Core State Standards (CCSS), was one way for states to gain points in the competition for 2009’s Race to the Top funds. Starved for revenue during the Great Recession, states could be forgiven for mistaking this incentive as non-optional. As years passed without the reauthorization of NCLB, Secretary of Education Arne Duncan offered waivers to states seeking exemption from the most onerous provisions of the law – in particular, the requirement of 100 percent proficiency. In exchange for waivers, states had to sign on to a list of reforms favored by the administration, including career and college ready standards, assessments, and teacher evaluation systems that included student test scores. Many Republicans accused Duncan of using waivers to impose an unlegislated reform agenda on the states. Conservative support for Common Core sharply declined. Michael Petrilli, a Republican supporter of Common Core, urged Duncan to help the CCSS effort “by declaring that the federal government is going to stay a million miles away from the Common Core.”

The Bush and Obama administrations respected federalism rhetorically when describing standards. But the actual result was different. The federal role in K-12 education was larger in 2017, when Barack Obama left office, than in 2001, when George W. Bush entered. Standards based reform played an important part in the administrations’ expansion of influence and centralization of power.
Public Opinion

The public likes standards in education. The popularity of the general idea has remained stable over decades. A 1989 Gallup Poll, for example, found 70 percent of the public in favor of requiring local schools “to conform to national achievement standards and goals” and 77 percent supporting tests to determine if students meet the standards. Nearly three decades later, the 2017 Education Next Poll also found widespread support.

Table 1 reproduces the results of several Education Next polls on standards. Two trends are evident. The surveys presented random samples of respondents with two versions of the same question, one asking if they supported “Common Core standards” and the other simply “standards.” The widespread popularity of CCSS in 2013, with 65 percent in favor and only 13 percent opposed, fell sharply in subsequent years. By 2017, support registered 41 percent and opposition 38 percent. But support for the generic notion of “standards” remained strong, with 61 percent in favor and only 20 percent opposed.

<table>
<thead>
<tr>
<th>Position</th>
<th>Prompt</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Name</td>
<td>Support</td>
<td>65</td>
<td>53</td>
<td>49</td>
<td>42</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Oppose</td>
<td>13</td>
<td>26</td>
<td>35</td>
<td>42</td>
<td>38</td>
</tr>
<tr>
<td>Without Name</td>
<td>Support</td>
<td>NA</td>
<td>68</td>
<td>54</td>
<td>56</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Oppose</td>
<td>NA</td>
<td>16</td>
<td>30</td>
<td>28</td>
<td>20</td>
</tr>
</tbody>
</table>

Note: Question with name omitted was not asked in 2013

Although opposition to CCSS may have peaked in 2016 and now be plateauing, the Common Core brand has clearly taken a hit. By 2014, conservative opposition had led legislatures in three states—Indiana, Oklahoma, and South Carolina—to rescind earlier adoption of the standards. Dozens more states revised their standards (only superficially) or changed the name to exclude any reference to Common Core. A similar erosion of support had occurred with NCLB. Separating the tarnished brand from the potential of standards became a popular activity.
of politicians. In a 2007 speech, candidate Barack Obama tried to differentiate the virtue of standards from the less popular aspects of NCLB. He charged the law with “demoralizing teachers”; the New York Times article on the speech went on to report, “But he also said it was right to hold all children to high standards. ‘The goals of this law were the right ones,’ he said.”

**Research on the Effect of Standards**

This section summarizes research on the effect of standards on student achievement. The discussion is organized chronologically, starting with standards-based reform in the 1990s and ending with Common Core.

**Standards-Based Reform in the 1990s**

In the early days of NCLB, two studies examining NAEP data from the 1990s suggested standards-based reform could be used to boost student achievement. Importantly, both studies should be understood as evaluations of standards in conjunction with accountability systems, not of standards alone. The analyses compared states that sanctioned schools based on low test scores with states not employing such sanctions (Carnoy and Loeb 2002; Hanushek and Raymond 2004). Twelve states had school accountability systems in 1996; by 2000, a year before NCLB was enacted, the number had grown to thirty-nine. All of the accountability states had adopted standards or frameworks. The tests that identified failing schools were written to reflect the standards, but the studies only modeled the accountability regimes, not the tests or standards themselves, as potential causal factors.

Carnoy and Loeb included student and school accountability provisions in creating an index reflecting the intensity of accountability for each state. Higher NAEP scores were
associated with stronger accountability systems. The Hanushek and Raymond study was limited to school accountability. They found that states adopting consequential school accountability—that is, accountability with teeth—were more likely to make gains from fourth to eighth grade on NAEP than states without consequential accountability.¹²

NCLB

As NCLB matured, several researchers investigated the effects of NCLB by examining variation in state performance on NAEP. A study by Dee and Jacob exploited the fact that while many states had test-based accountability in the 1990s, NCLB forced the remaining states to adopt standards in reading and mathematics, to test student progress in demonstrating proficiency with the standards, and to hold schools accountable for the results. Like the pre-NCLB studies, it is the final policy element of accountability—in this case, the introduction of NCLB-style accountability—that served to differentiate the states in the analysis. Dee and Jacob found statistically significant, positive effects for NCLB on fourth grade math scores (effect size of 0.23). Gains in eighth grade math were positive and nearly significant (effect size of 0.10); no significant effects were found in reading in either fourth or eighth grade.¹³

Wong, Cook, and Steiner also used variation in NAEP scores to evaluate NCLB’s impact. They compared national public school performance to that of Catholic schools, the latter presumably outside of NCLB’s influence. In addition, they modeled the stringency of state accountability systems by comparing the number of schools declared at risk of sanctioning. The study found consistently positive effects of NCLB on fourth and eighth grade math scores but no effect in reading.¹⁴
Wong, Cook, and Steiner caution that their findings apply to No Child Left Behind as a whole, without being able to tease apart the various elements of the law. That warning underscores the problem with drawing inferences regarding the effect of standards from the NCLB research. Standards are foundational to NCLB. Despite that, evaluating the end effects of NCLB, or as the pre-NCLB studies do, the end effects in states with standards-based accountability systems, are capturing the impact of multiple policies, not just standards.

Efforts to isolate the effects of standards have been sparse. A 2009 analysis by Russ Whitehurst investigated whether the quality of a state’s standards is related to its performance on NAEP. Whitehurst used ratings of state standards produced by two organizations, the American Federation of Teachers (AFT) and the Fordham Foundation, and NAEP scores from 2000-2007 (both the longitudinal gains and the cross-sectional scores for each administration of NAEP). He found no correlation between the quality of state standards and NAEP scores. Whitehurst also disaggregated the state NAEP scores of white and black students and still found no statistically significant correlation. States with weak standards, at least in the eyes of AFT and Fordham’s evaluations, scored about the same on NAEP as states with strong standards.

Leading Up to Common Core

The 2012 Brown Center Report (BCR) presented an analysis attempting to predict the effects of Common Core. Whitehurst’s 2009 finding—that the quality of state standards bears no relationship with student achievement—was confirmed using NAEP data from different years and employing a different set of demographic controls.

One section of the BCR study focused on test score variance. The “common” part of Common Core addresses between-state differences in standards, not differences within states.
After all, in all 50 states students in 2012 had been educated under common standards since the advent of NCLB in 2002—and in many states, long before that. The assumption is that state-level learning differences are driven by differences in students’ exposure to educational factors shaped by state standards. High quality, common standards would lead to a good education for all by boosting achievement in low performing states.

The idea may be intuitively attractive, but where variance appears in test scores raises doubt. The standard deviation of NAEP scores within states is four to five times larger than the standard deviation of state means. What does this say about standards’ potential reach? Massachusetts and Mississippi report NAEP means about 25 points apart. Some believe that common standards can ameliorate such a difference. But as shown in Table 3, the average within-state standard deviation is larger than 25 points. Every state has a mini-Massachusetts-Mississippi contrast within its own borders. In some cases, mini-Massachusetts-Mississippi contrasts can be found within school districts, within schools—yes, even within classrooms. And those contrasts exist despite students receiving an education under common standards.17

<table>
<thead>
<tr>
<th>Grade/Subject</th>
<th>Between-State SD</th>
<th>Within-State SD</th>
<th>Multiple (Within/Between)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourth Grade Reading</td>
<td>6.6</td>
<td>34.7</td>
<td>5.3</td>
</tr>
<tr>
<td>Fourth Grade Math</td>
<td>6.3</td>
<td>27.8</td>
<td>4.4</td>
</tr>
<tr>
<td>Eighth Grade Reading</td>
<td>6.5</td>
<td>32.9</td>
<td>5.1</td>
</tr>
<tr>
<td>Eighth Grade Math</td>
<td>8.5</td>
<td>34.8</td>
<td>4.1</td>
</tr>
</tbody>
</table>

The BCR study concluded that the Common Core State Standards would have a negligible impact on student achievement. The quality of state standards is not correlated with state NAEP scores; the level at which states define proficiency on state tests is also largely uncorrelated with NAEP scores; and the amount of test score variance between states is dwarfed
by the variance within states, the latter four to five times larger than the former despite within-
state variance already having been “treated” by common standards.\textsuperscript{18} It is important to
emphasize that these shortcomings are not exclusive to Common Core. They apply to all
standards.

A study with an optimistic outlook for Common Core was published later in 2012. William H. Schmidt and Richard T. Houang of Michigan State University (MSU) rated states’
mathematics standards for their similarity to the CCSS math standards (CCSS-M) and compared
the ratings with states’ 2009 NAEP eighth grade math scores.\textsuperscript{19} The ratings drew on the authors’
work analyzing scores on international math assessments, which produced the theory that high
achieving nations scored at the top of international league tables because of the focus, rigor, and
coherence of their math standards.\textsuperscript{20} The MSU study employed a rubric showing that CCSS
math standards were similar to the standards of top scoring nations in focus and coherence. They
then rated state standards using the same rubric and calculated a proximity score reflecting the
congruence of state standards with CCSS-M.

The authors regressed 2009 state NAEP scores on the proximity scores and a set of
demographic controls. The initial regression produced no findings of statistical significance. An
analysis of residuals suggested dividing the states into two groups: thirty-seven wealthier, higher
scoring states (Group A) and thirteen, less wealthy, lower scoring states (Group B). Modeled
separately, the two groups produced statistically significant results, indicating that 2009 state
NAEP scores were related to the similarity of state math standards to CCSS-M. Schmidt and
Houang hypothesized that the two groups reflected differences in the implementation of
standards (Group B facing more severe resource constraints and demographic challenges) and
concluded that the findings, in terms of Common Core, offered “a vision of what can be,” a vision of what Common Core standards could accomplish if implemented appropriately.\(^{21}\)

**Effects of Common Core**

In a February 2017 article, Morgan Polikoff argued that it may be too early to tell whether Common Core has had any effect.\(^{22}\) He cites several examples—among them, studies of small high schools, vouchers, NCLB, and School Improvement Grants—in which early findings were either partially or fully reversed by later evaluations. It takes time to implement complex policies, and Common Core, seeking to tweak several downstream parts of the educational system, probably requires more patience than most policy efforts before a valid evaluation can be conducted. In addition, no one knows when standards are fully operational, in the sense of pinpointing when a fair amount of time has passed for judging whether their intended changes have actually taken effect.

Nonetheless, early research of standards has its virtues. Opportunity costs in terms of time and resources arise in the early days of implementation. The key activities that Common Core supporters are relying on for successful implementation of the standards—development of tests, creation of curriculum materials, and professional development of teachers—were well underway if not completed by 2013. Billions of dollars and hundreds of millions of hours of educators have gone into the CCSS effort. In 2017, it’s too late to get that money and time back or to redirect them towards other endeavors. When Common Core was first being debated, supporters surely knew that mid-course corrections might be necessary. A mid-course correction is impossible if the direction one is heading at mid-course is unknown.
I have conducted several studies of Common Core using NAEP scores, and they possess all of the limitations of evaluating standards in the early phases of implementation. They are by no means conclusive. Moreover, they cannot discern causal effects, only the changes in NAEP scores associated with groups of states sharing particular characteristics. Just like the studies described above, including the Schmidt and Huoang analysis used by Common Core advocates, gauging the effect of Common Core requires analysis of state variation in some aspect of the reform. I created implementation indexes to model how states differed in implementing Common Core.

The 2011 implementation index is based on responses to a 2011 survey by federal budget officials asking how states spent funds from the American Recovery and Reinvestment Act of 2009. States reported if they had engaged in four activities: 1) adopted CCSS, 2) provided, guided, or funded professional development on the CCSS, 3) provided curriculum/instructional materials for the CCSS, or 4) worked with a consortium to develop assessments aligned with CCSS. States responding affirmatively to all four activities were categorized as “strong implementers” of CCSS (nineteen states). States indicating that they had not engaged in any of the activities (five states) were designated “non-adopters.” The remaining twenty-six states were categorized as “medium” implementers; that is, they had adopted CCSS but had not engaged in all facets of implementation.23

The study found that from 2009-2013, strong implementation states gained 1.9 NAEP scale score points in eighth grade math; the medium implementers gained 1.0 points; and the non-adopters gained 0.6 points. The 1.3 scale score advantage of strong implementation states over the non-adopters is quite small (about 0.035 standard deviation). Most of the advantage appears in 2011-2013, the second half of the four-year interval, which offers a sliver of
encouragement for CCSS. It seems reasonable to assume that more CCSS states would have begun implementation after 2011 rather than before 2011. Thirty-three states and D.C. had adopted CCSS by August 2010—and another 12 states by June 2012.24

The same analytical strategy was employed in later studies, updating the initial analysis using the most recently released NAEP data and employing a second index of implementation. The second index was based on responses to a 2013 survey of chief state school officers on their state’s plans for implementing CCSS. States with plans to complete “classroom implementation” of the Common Core ELA standards by the end of the 2012-2013 school year were designated strong implementers (eleven states); states that had not adopted CCSS or had rescinded adoption were classified “non-adopters” (seven states); and the remaining states were placed in the “medium” category (thirty-two states).

The 2016 Brown Center Report (BCR) examined the relationship of state implementation to NAEP scores in fourth grade reading and eighth grade math. The results were similar to the earlier BCR study. Only trivial differences, 1.5 NAEP scale score points or less, were detected in the NAEP scores of strong implementers of CCSS and the scores of non-adoption states.

For the current chapter, I conducted a new analysis. Table 4 shows the score differences from the two models.25

Table 3. NAEP Scale Score Change, Difference between Strong CCSS Implementation States and Non-Adopters
(Average change in demographically adjusted score, 2009-2015)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourth Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>Imp11</td>
<td>-0.1</td>
<td>-1.7</td>
</tr>
<tr>
<td></td>
<td>Imp13</td>
<td>0.0</td>
<td>-2.4</td>
</tr>
<tr>
<td>Eighth Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>Imp11</td>
<td>-0.3</td>
<td>-2.3</td>
</tr>
<tr>
<td></td>
<td>Imp13</td>
<td>0.6</td>
<td>-1.1</td>
</tr>
</tbody>
</table>

Note: Positive signs favor states with strong implementation of CCSS; negative signs favor non-adoption states.
The data in Table 3 use regression-adjusted NAEP scores, calculated by Matt Chingos and Kristin Blagg of the Urban Institute, to report demographically controlled changes in state scores. Chingos and Blagg adjusted state NAEP scores for racial composition, the proportion of students identified for free and reduced lunch, special education, and English Language learner programs, and age at testing. The data compare average growth on NAEP in strong implementation states with growth in the non-adoption states. Cells with positive signs favor the strong implementers; cells with negative signs favor non-adopters. The findings are consistent with the previous analyses using unadjusted scores, detecting only tiny differences between strong implementers of CCSS and non-adopting states. The largest difference favoring strong implementation states—found in the bottom row of the column, “Entire Six Years (2009-2015)”—is only 0.6 NAEP points. All of the comparisons of NAEP score changes in the most recent interval (2013-2015) favor non-adopters of CCSS (ranging from 1.1 to 2.4 scale score points).

As mentioned above, the first analyses suggested that strong implementers of CCSS held a slight advantage in the early days of implementation. The later analyses, including the one reported in Table 3, suggest that after 2013 the trend had reversed. The advantage was now turning towards the non-adopters of CCSS. It’s important to note that for all states, regardless of whether they had warmly embraced CCSS or developed their own standards, NAEP scores were stagnant from 2009-2015. NAEP critics’ charge that Common Core is responsible for holding back U.S. achievement is not supported by this analysis.

**Lessons Learned**

In most states, standards based reform has existed for at least three decades. The Bush and Obama administrations premised their education policies on the potential of standards to improve
schools and boost student achievement. The Bush administration’s vehicle for standards based reform was NCLB. Research indicates that NCLB produced modest positive effects on student achievement—confined mainly to math and with no discernible effect in reading. The Obama administration embraced the Common Core. To date, seven years after the CCSS were released and adopted by a vast majority of states, no causal evidence exists on their effectiveness in boosting student achievement. The exploratory analyses presented in this chapter suggest minimal effects, perhaps slightly positive in the early years of implementation but then fading over time. More rigorous analyses are needed to confirm or reject this hypothesis.

The Bush-Obama record with standards is disappointing. Research suggests that standards—even clear, ambitious, and elegantly worded standards—are weak instruments for raising student achievement. Nevertheless, out of this disappointment have emerged several lessons for future policy makers and researchers.

*Lesson #1: Political support matters and supportive coalitions may not last.*

Public support for standards reaches its apogee when standards are first considered—as an aspirational idea in its most general form, lacking curricular details, and shorn of accountability. In other words, a reform without substance or costs. Once standards are written, adopted, and begin to influence educational activities downstream, support declines. Over time, Common Core became a tainted brand, just as NCLB had been. Considering that NCLB’s support plummeted after accountability systems began labeling schools as failures, support for CCSS may erode further once accountability tied to CCSS tests takes effect. The public loves their local schools and thinks they are just fine. Accountability linked to standards-based tests is for other schools.
Political coalitions that support standards during the design phase are difficult to maintain. Jay P. Greene has called standards, “just words on paper,” powerless in the abstract. It is important to appreciate that the term “standards,” to paraphrase Greene, is just a single word, but one with multiple meanings. What gets the political right excited about standards is holding the public school system accountable for students’ learning of challenging material. That’s why Checker Finn and Mike Petrilli like standards. What gets Education Trust excited is being able to promote an equity agenda for disadvantaged kids. They focus on closing gaps. What gets professional groups excited is re-orienting curriculum and instruction towards the ideals of progressive education—higher order thinking, learning how to learn over learning facts, conceptual understanding, project-based learning, heterogeneous groups, and so on. That’s why proponents of balanced literacy, 21st century skills, and constructivist math like standards. All of these groups construe the word “standards” in accord with their respective world views—world views that do not always overlap.

Lesson #2: Standards-based reform is top-down and mechanistic.

Standards advocates portray education reform as an engineering problem. Improvement will arise from getting all of the system’s pieces working in synch with each other in pursuit of ambitious goals. Setting standards is what professional educators and upper level policymakers do; the work that goes into attaining standards is done downstream by someone else—and often referred to simply as “implementation.” The political and organizational contexts of schools and school systems are regarded as potential allies in implementing goals, overlooking that they also have the power to undermine the linkages between standards and the curriculum, instruction, and assessments that standards seek to forge.
Standards are regulatory. The fundamental difference between minimal and maximal regulation is germane to a discussion of education standards. Standards that attempt to ensure that every American child is college and career ready are maximal. USDA regulations are minimal. They do not guarantee that we can all have filet mignon every night for dinner—or even a meal that tastes good. They do guarantee that the food supply will be safe and reasonably devoid of “yuck factors,” ingredients that most people would prefer not to eat but won’t kill you. Yuck factor standards are minimal. That’s why the FDA publishes a handbook setting acceptable levels for insect parts, rat hairs, mouse poop, maggots, and other disgusting “natural or unavoidable defects” showing up in our food. Setting those levels at zero would shut down the food supply. Utopianism rarely makes good public policy.

Lesson #3: Curriculum matters, but standards do not guarantee more effective curriculum.
It was pointed out above that implementation of standards involves their impact on downstream educational activities—what happens in schools and classrooms. Improving curriculum is one of those downstream events.

The effects of curriculum programs can vary significantly, as was demonstrated through a randomized control trial of first and second grade math curricula conducted by Mathematica Inc. The study today still remains one of the only experimental evaluations of K-12 textbooks. The four programs in the experiment covered essentially the same topics. Students in three of the programs (Math Expressions, Saxon, and Foresman-Addison) scored about the same, but all three outscored the fourth program (Investigations) by a statistically significant amount (effect size of about 0.22). A student at the 50th percentile who received instruction in Investigations in
first and second grade would have scored at the 59th percentile if taught from one of the other programs.

Curricular alignment to CCSS refers to whether a curriculum adheres to Common Core standards. Alignment is not synonymous with quality. Alignment evaluations, typically conducted by expert panels rating textbooks, answer the question of whether texts address the topics embodied by standards; they do not address a text’s impact on learning. Two texts may teach the same topic, with one doing it well and one doing it poorly. If that topic is addressed in the CCSS, both texts are in alignment.

A serious threat to alignment arises from a practical reality of classrooms: everything that is taught is not learned. The CCSS, for example, stipulate that addition of numbers within 1000 will be taught and learned in second grade. Not all second graders readily learn addition of two-digit numbers. Many struggle with regrouping—“carrying,” in the old-fashioned parlance and “composing and decomposing” in Common Core lingo—because they fail to grasp the underlying principles of base-ten place value that underpin all but simple computation. Slippage occurs. Addition of three-digit numbers exacerbates the problem.

Flash forward to learning subtraction. Teachers attempting to teach subtraction will notice that some kids still don’t know how to add two or three digit numbers. Since it makes no sense to teach the more difficult concept, they will try to get those kids on pace by re-teaching two-digit addition. That is what any good teacher would do, but it renders the class curriculum out of alignment with the proscribed curriculum. When these slippages occur year after year, misalignment grows.

CCSS writers are hopeful that standards can avoid slippage, or at least diminish the chances of it occurring. There are fewer topics in each grade. The pace of instruction is slowed
down to give more time to each standard. But the fact remains that grade-based standards assume that students possess the pre-requisite knowledge to tackle each grade’s new learning. CCSS supporters believe that students will be better prepared to transition from grade to grade, but that’s a belief, one not supported by empirical evidence, and by no means assured. In a 2017 poll in three states—Texas, Ohio, and Kentucky—teachers were asked to rate a list of ten challenges they face implementing college and career readiness standards. The top two responses were “a wide range of student abilities,” with 71 percent describing it as a moderate or major challenge, and “inadequate student preparation in prior grades,” with 62 percent rating that as a moderate or major challenge.28

The same potential for misalignment arises on the other end of the achievement spectrum, with high achieving students who already know the standards for their grade. Why waste time and bore students to death by teaching them something they already know? The CCSS math standards offer a single course in eighth grade that includes some algebra, but it is not an Algebra I course—that’s reserved for ninth grade. Following the adoption of CCSS math standards, enrollment of eighth graders in Algebra I began to decline, falling from a peak of 48 percent in 2013 to 43 percent in 2015. The decline reversed a steady increase in eighth grade Algebra I enrollment that had been going on since the 1990s. Many educators consider acceleration of some students (but not all) as antithetical to the “common” in Common Core. If all students can’t take Algebra I in eighth grade, this point of view reasons, no one should be allowed to take it.

The alignment studies may have exacerbated the difficulty of getting good materials for high achievers into classrooms. Singapore Math, a curriculum long admired by the gifted and talented education community for advancing students quickly through challenging material,
scored poorly for alignment (see reviews for Math in Focus: Singapore Math) in the reviews of EdReports.org. A common complaint of the evaluators was that Singapore taught content too early. Teaching content ahead of its designated time violates the CCSS notion of curricular coherence.

Lesson #4: Despite assurances that standards do not dictate how teachers should teach, standards can be used in efforts to change pedagogy.

Early research on standards recognized that external standards faced a difficult task in changing how teachers teach. Andrew Porter (1988) discussed the inherent tensions of promoting professionalism and teacher autonomy with “telling teachers what to do.” He suggested a compromise: “By involving teachers in setting standards, by restricting standard setting to student achievement and leaving pedagogical practices a matter for individual professional discretion, and by recognizing new standards require new resources (e.g., teacher knowledge, instructional materials, organizational structure), it may be possible to reap the perceived benefits of external standard setting and the perceived benefits of teacher empowerment.”

As pointed out above, Porter’s admonition to leave pedagogy to the discretion of teachers was reinforced by the failure of the 1990s standards projects, which were chock full of pedagogy. The Common Core website is emphatic: “Teachers know best about what works in the classroom. That is why these standards establish what students need to learn but do not dictate how teachers should teach. Instead, schools and teachers will decide how best to help students reach the standards.”

Jason Zimba, one of the primary authors of the CCSS math standards, commented, “Standards shouldn’t dictate curriculum or pedagogy. But there has been some criticism recently
that the implementation of CCSS may be effectively forcing a particular pedagogy on
teachers.” Indeed, press accounts documented teachers complaining about pressure to change
their teaching. Emanuel Felton in a 2014 Hechinger Report story reported, “The Common Core
wasn’t necessarily supposed to change how math is taught, but in many schools that’s exactly
what’s happening.” Later in the article, Phil Daro, one of the lead writers of the Common Core
math standards, is quoted: “The Common Core is silent about how to teach. When we wrote the
standards we were prohibited from addressing how to teach, that’s not what standards are
supposed to do.”

So what happened? In reading and mathematics, the 1990s were riven by curriculum
battles between traditionalists and progressives. Standards embody political compromises, and
the compromises manifest themselves in negotiated wording. Serving a role similar to the
communiques issued after meetings of world leaders, standards are orchestrated to emphasize
common ground. But the common ground erodes once standards are converted into curriculum
and instruction. Differences papered over with rhetoric are exposed. Dog whistles signal to local
educators that imposing pedagogical restraints on teachers may be compatible with the standards
after all.

Here’s an example. Engage New York published what it called “twelve pedagogical
shifts demanded by the Common Core State Standards,” six in math and six in English-Language
Arts. The term “instructional shifts” spread like wildfire. Teacher-blogger Barry Garelick, a
critic of Common Core and math traditionalist, observed, “[E]ven before the ink dried on the
Common Core standards, proponents of Common Core talked about ‘The Shifts’ as if they were,
[or] are, enforceable parts of the standards themselves.” Garelick has devoted several blog posts
to describing the progressive instructional practices on which he has received professional development, all in the service of implementing Common Core.\(^{30}\)

**Conclusion**

There is a wonderful story (and almost certainly a myth) about a comment made by Premier Zhou Enlai during Richard Nixon’s historic trip to China in 1972. Asked about the historical impact of the French Revolution, Zhou was reported to have responded, “It’s too early to tell.”\(^{31}\) The comment quickly circulated as affirmation of the Chinese leader’s legendary ability to take the long view on world events.

Common Core supporters have urged a long view on standards. Michael Kirst, chair of the California State School Board, named 2020 as the year when the beneficial effects of Common Core might be first apparent.\(^{32}\) That’s ten years after most states adopted the standards. The second graders of 2010 will graduate from high school in 2020. There may even be a longer wait. If Common Core is to be judged by its impact on college and career effects, such outcomes cannot be measured until several years after 2020.

Common Core sits in a public policy purgatory. We don’t know if it’s working; we don’t know when we will know if it’s working. The policy still has supporters, but its detractors are also numerous. As of 2017, the research on CCSS has been dominated by descriptive accounts of implementation (e.g., what’s going on) along with an effort to build a knowledge base that will help local educators implement CCSS (e.g., how can we do this better). That is a reasonable research agenda, but given the track record of standards-based reform, perhaps it’s time to consider an additional hypothesis: that standards simply don’t work.
Draft: Do not cite without permission from the author.

9 Mathematically Correct has ceased operation. NYC HOLD continues at http://www.nychold.com/
10 McGuinn.
12 An earlier Hanushek and Raymond (2003) analysis found no statistically significant difference between consequential accountability and simply publishing school report cards. The 2004 analysis rejected the equality of the two approaches at p<.10.
13 Thomas Dee and Brian Jacob, “Evaluating NCLB,” Education Next (Summer 2010), 54-61.
17 The 1965 Coleman Report was probably the first national study to decompose test score variance and highlight that the variance associated with between-school factors is dwarfed by within-school factors. More recently, Chingos, Whitehurst, and Gallaher’s (2013) decomposition of variance in Florida achievement data estimated 1.3% associated with district, 9% with school, 31% with teacher, and 38% with student differences (with the remainder unspecified).
18 Evidence was reported of a modest relationship between raising proficiency cut points and boosting scores in fourth grade.
21 To underscore the favorable reception Common Core advocates gave to the MSU study, the main findings were released at a Washington DC event in May, 2012 sponsored by organizations (including Achieve, Inc.) with strong ties to CCSS. Video available on Achieve website: https://www.achieve.org/videos/video-common-core-math-standards-implementation-can-lead-improved-student-achievement.
Survey data on which the implementation index is based can be found in Table H.1 “Standards and Assessment Indicators by State, 2010-2011,” State Implementation of Reforms Promoted under the Recovery Act, A. Weber et al. (2014). Minnesota did not adopt CCSS math standards but did adopt the ELA standards. For full explanation of state coding in the study, see Tom Loveless, “A Progress Report on Common Core,” in The 2014 Brown Center Report (Washington, DC: Brookings Institution), 26-34.

“Common Core Timeline,” Education Week.

Omitting demographic controls typically does not affect short-term longitudinal analyses of NAEP data because the baseline score of each state serves, in a sense, as a control. State demographic statistics take a long time to change appreciably.


CCSS Math Content.2.NBT.B.7 calls for adding and subtracting within 1000. “Understand that in adding or subtracting three-digit numbers, one adds or subtracts hundreds and hundreds, tens and tens, one and ones; and sometimes it is necessary to compose or decompose tens or hundreds.”


