FORTY-SIX STATES AND THE DISTRICT OF COLUMBIA HAVE signed on to the Common Core State Standards Initiative, a project sponsored by the Council of Chief State School Officers (CCSSO) and the National Governors Association (NGA). The Common Core spells out what students should learn in mathematics and English-language arts from kindergarten to the end of high school. The standards were written by teams of curriculum specialists and vetted by panels of academics, teachers, and other experts.[1] In 2010, the federal government funded two consortia to develop assessments aligned with the Common Core. The new tests are to be ready in 2014.

The push for common education standards argues that all American students should study a common curriculum, take comparable tests to measure their learning, and have the results interpreted on a common scale, with the scale divided into performance levels to indicate whether students are excelling, learning an adequate amount, or falling short. Past experience with standards suggests that each part of this apparatus—a common curriculum, comparable tests, and standardized performance levels—is necessary. No one or two of them can stand alone for the project to succeed.

Proponents point to the intuitive appeal of a common curriculum. "It's ludicrous," Bill Gates told the *Wall Street Journal*, "to think that multiplication in Alabama and multiplication in New York are really different."[2] In a report called *The Proficiency Illusion*, The Fordham Institute made a similar point regarding state efforts to evaluate schools using fifty different assessments and fifty different definitions of what constitutes acceptable performance.[3] How can a school in one state be labeled a failure while a school in another state and with almost exactly the same test scores can be considered a success?

The authority to operate school systems is constitutionally vested in states. But states have undermined their own credibility when it comes to measuring student learning. Accounts of dumbed-down and

poorly-written state tests, manipulation of cut scores to artificially boost the number of students in higher performance levels, and assessments on which students can get fewer than 50% of items correct and yet score "proficient" fuel the belief that states individually cannot be trusted to give the public an accurate estimate of how American education is doing.[4]

Three Theorized Effects

The Common Core State Standards are theorized to improve education in three ways. First, proponents argue that the Common Core is superior to most current state standards. In a recent study, The Fordham Institute concluded that Common Core standards are better than 37 states' standards in English-language arts and 39 states in mathematics.[5] It follows, proponents believe, that the Common Core will raise the quality of education nationally by defining a higher-quality curriculum in English-language arts and mathematics than is currently taught. Let's call this the "quality theory." Achievement will increase because students will study a better curriculum.

The second idea is that the Common Core sets higher expectations than current state standards, the assumption being that cut points on the new assessments will be set at a higher level than states currently set on their own tests. Comparisons with the National Assessment of Educational Progress (NAEP) lead many analysts to conclude that states set proficiency standards far too low. States routinely report more students attaining proficiency than NAEP indicates, often 30–40 percentage points more.[6] The *No Child Left Behind Act* left it up to the states to design their own tests and to set performance levels wherever they want, but the pattern of states reporting significantly higher percentages of proficient students preceded NCLB.[7] A new Common Core test will presumably end such discrepancies by evaluating proficiency using the same standards for every state, and these standards are to be more rigorous than those currently used. Schools and students will respond by reaching for these loftier goals. Let's call this the "rigorous performance standards" theory.

The third hypothesis is that standardization yields its own efficiencies. In the same *Wall Street Journal* interview cited above, Bill Gates referred to this idea by complaining about the time and money wasted on the many different versions of textbooks that are published to conform to individual states' curricular tastes.[8] In a reverse spin on the same argument, others argue that textbooks are bloated with redundant content as publishers attempt to incorporate numerous idiosyncratic state curricular mandates into one book.[9] The assumption of both arguments is that one, high-quality textbook—or perhaps a few that are aligned with the same content standards—used by all American students attending the same grade would be an improvement over the status quo. Other proponents point to the potential gaps in learning that occur as students move from state to state. Especially when students move mid-year, important concepts might be missed while other concepts are studied unnecessarily a second time. Teachers who move from state to state experience similar difficulties in terms of lesson planning. Let's call this the "standardization" theory.

Opposing Arguments

Some analysts question the theories behind the Common Core. Writing in *Education Week* in the summer of 2011, Andrew Porter compared the Common Core to existing state standards and international standards from other countries and concluded that

The Common Core State Standards are theorized to improve education in three ways.

...data exist that can help predict the magnitude of effects from the Common Core.

the Common Core does not represent much improvement.[10] Opponents of the Common Core, including Sandra Stotsky, James Milgram, Ze'ev Wurman, and Williamson Evers, criticize the quality of the proposed standards for English-language arts and mathematics. They conclude that the math standards, in particular, are inferior to existing standards in Massachusetts and California.[11]

Critics of the Common Core issued a "counter-manifesto" arguing that the proposed common standards would undermine the decentralized, federalist principles on which education has been governed since America's founding. Declaring that a "one-size-fits-all, centrally controlled curriculum" does not make sense, the counter-manifesto states that only weak evidence supports the push for national standards. International test data are not helpful since most countries have national standards and the few that do not, including Canada and Germany, have both impressive and non-impressive scores. Concern for interstate student mobility is overblown, the counter-manifesto claims, because very few students move between states. Most mobility is within state, which is already addressed by the *No Child Left Behind Act*'s requirement that every state establish standards. Since 2003, every state has state curriculum standards that delineate the curriculum for public schools within its borders.[12]

Can empirical evidence shed light on the main points of contention in this debate? Not entirely. Much of the argument is philosophical. Those who believe that the Common Core enumerates what schools should be teaching and students should be learning support the proposed standards. And those who believe a greater degree of standardization would produce more common educational outcomes—and that common outcomes are desirable—also support the proposed standards. Those holding to the opposite beliefs, and believing that local school governance is preferable to governance by larger entities, are critics of the standards.

Despite the philosophical disagreements, there are empirical questions on which evidence exists. The nation has had several years of experience with education standards—since the 1980s in many states and since 2003 in all states—and data exist that can help predict the magnitude of effects from the Common Core. How much does raising the quality of standards matter in boosting student achievement? Will raising the bar for attaining proficiency—in other words, increasing the rigor of performance standards—also raise achievement? And how much variance will be reduced—or how much "sameness" in achievement will be attained—by having students across the country studying a common curriculum?

Quality and Achievement

Let's start with the theory that high-quality standards promote achievement gains. In October 2009, a colleague at Brookings, Grover "Russ" Whitehurst, investigated whether quality ratings for state standards, as judged by the two most cited ratings (from the American Federation of Teachers and Fordham Foundation), are correlated with state NAEP scores. Whitehurst found that they are not. States with weak content standards score about the same on NAEP as those with strong standards. The finding of no relationship held up whether NAEP scores from 2000, 2003, 2005, 2007, or the gains from 2000–2007 were used in the analysis. And it held up for the scores of both white and black students.[13]

The current study extends that inquiry by looking at NAEP data from 2003–2009. Gain scores on NAEP reading and math tests from 2003 and 2009 are combined to form a composite gain score. The scores

are adjusted to control for demographic characteristics of each state—the percent of students qualifying for free or reduced lunch, special education, or English language learner status. More precisely, scores are adjusted to control for changes that occurred in those demographic characteristics from 2003–2009. That prevents swings in states' demographic characteristics from skewing the results. Ratings of state curricular standards conducted by the Fordham Foundation in 2000 and 2006 are used to model the quality of state standards. It is particularly apt to model the quality of state standards with the Fordham ratings considering Fordham's high opinion of the Common Core.

The results are shown in Table 1-1. Three questions are answered by the data. The first row addresses the question: Do the Fordham ratings in 2000 successfully predict the NAEP gains that states made in reading and math from 2003–2009? One could imagine, since there is undoubtedly some lag time before standards are implemented in classrooms and realized in student learning, that the curriculum standards of 2000 would influence achievement gains made three, six, or even nine years down the road. The correlation coefficient of –0.06 indicates that they do not.

The second row examines whether the ratings of 2006 are statistically related to 2003–2009 NAEP gains. In other words, was the quality of standards in the middle of the gain period related to test score gains? Again, the answer is no, with a correlation coefficient of 0.01. The final row looks at the change in ratings from 2000 and 2006. According to Fordham, some states improved their standards in 2006 while others adopted weaker standards in 2006 than they had back in 2000. Are changes in the quality of standards related to changes in achievement? Again, the answer is that they are not (correlation coefficient of 0.08).

Relationship of Fordham's Ratings of State Content Standards with State NAEP Gains (2003–2009)

Table 1-1

Standards Rating	Correlation Coefficient
Fordham 2000	–0.06
Fordham 2006	0.01
Change in Fordham 2000–2006	0.08

Relationship of State Proficiency Level with NAEP Achievement (Correlation Coefficients)

Table 1-2

	2005 NAEP	2009 NAEP	Change 2005–2009
4th Grade Reading	–0.22	–0.08	0.35*
4th Grade Math	–0.12	0.01	0.34*
8th Grade Reading	–0.11	–0.09	0.06
8th Grade Math	0.00	0.01	0.02

* $p < .05$

Rigorous Performance Standards and Achievement

The second theory of improvement is based on performance standards. A 2006 NCES report found that the difficulty of state performance standards is uncorrelated with achievement.[14] Performance levels (or "cut points") for student proficiency were mapped onto the 2005 NAEP scale. States with higher, more rigorous cut points did not have stronger NAEP scores than states with less rigorous cut points. A new NCES report was released in 2011 with updated measures using 2009 NAEP data.[15]

Table 1-2 summarizes the correlations between the rigor of state performance levels and achievement. In a replication of the earlier NCES study, we also find that the states'

A 2006 NCES report found that the difficulty of performance standards is uncorrelated with achievement.

...the absolute level of performance standards does not seem to matter but raising or lowering levels does exhibit a relationship with fourth grade changes in achievement...

2005 NAEP scores are unrelated to where the states drew the line for proficiency in 2005. Fourth-grade reading and math have slightly negative correlations (–0.22 and –0.12, respectively), as does eighth-grade reading (–0.11). The correlation coefficient for eighth-grade math is 0.00. State achievement is unrelated to the level at which states define proficiency. The same is true for 2009 NAEP scores and the level at which proficiency was placed that year (see the second column of the table).

The final column of Table 1-2 investigates whether changes in state NAEP scores from 2005–2009 are related to changes in proficiency level. Did states that raised the bar also perform better? And did states that lowered the bar perform worse? Correlation coefficients for 8th grade are near zero. Positive and statistically significant correlations were found for fourth-grade reading (0.35) and fourth-grade math (0.34). It is interesting that the absolute level of performance standards does not seem to matter but raising or lowering levels does exhibit a relationship with fourth grade changes in achievement, explaining about 12% of the variation in the change in state NAEP scores.

Whether one phenomenon is causing the other is difficult to tell. Changes in proficiency cut points are probably endogenous to trends in test scores. In other words, states with rising scores may feel emboldened to raise their proficiency cut points and those with declining scores may feel compelled to lower theirs. That is quite a different story than the raising or lowering of cut points producing changes in test scores. Unfortunately, simple correlations cannot determine the direction of causality, or if causality exists at all, only whether these two variables are statistically related. In the current analysis, change in level is related to change in fourth-grade scores.

How Common Will Achievement Become?

The third theory concerns standardization. For the Common Core movement, attaining greater standardization of educational outcomes is an important goal. If standards do not reduce variation, then even if they boost performance, simply raising average scores will still leave many states—and the districts, schools, and students within states—far behind and far below acceptable levels of performance. The two previous analyses indicate that it is unlikely that common standards will boost performance; however, it is possible for the national average on NAEP to remain stable while variation is reduced—for instance, if top states decline a little while states at the bottom rise by the same amount. Another way would be for high flying schools within states to decline a little while poorly performing schools increase their performance by a commensurate amount.

In terms of state NAEP scores, variation comes in two forms: variation between states and variation within states. We would expect common standards to reduce variation between states, so that the NAEP score difference between states at the top and bottom of the rankings would be reduced. States that currently offer vastly different curricula, assessments, and performance standards will harmonize those elements of their educational systems. One would expect test score differences to shrink. That is the essence of common standards. Within-state variation, on the other hand, remains unaffected by common standards. Every state already has standards placing all districts and schools within its borders under a common regime. And despite that, every state has tremendous within-state variation in achievement. Schools that score at the top of the world on

international assessments are within a short car trip, sometimes even within a short subway ride, from schools that score at the level of the world's lowest achieving nations.

Let's compare these two forms of variation. Table 1-3 displays data on NAEP standard deviations between and within states. Standard deviation is a measure of variation, the amount of spread in a group of data. On any particular test, about two-thirds of observations are within one standard deviation (above and below) of the average score. "Between-State SD" is the standard deviation of NAEP scores for the fifty states and the District of Columbia—how much they differ from each other. "Within-State SD" is the average of the standard deviations for the fifty states and the District of Columbia—how much the students within each state, on average, differ from each other.

The findings are clear. Most variation on NAEP occurs within states not between them. The variation within states is four to five times larger than the variation between states. Much of the similarity of state scores comes from aggregating individual student scores, which differ greatly, to the state level. The variation in student performance within states washes out to produce means that are alike across states. Consider this: fourth-grade NAEP scores in math range from Massachusetts at the top with 252 down to the District of Columbia with 219. That 33 point difference is not too much larger than the average standard deviation within states (27.8). What does that mean? Consider Massachusetts and Mississippi, a state with low scores but not at the very bottom. Their NAEP means differ by 25 points. Every state, including Massachusetts and Mississippi, has a mini-Massachusetts and Mississippi contrast within its own borders. That variation will go untouched by common state standards.

Relationship of State Proficiency Level with NAEP Achievement (Correlation Coefficients)

Table 1-3

	Average State NAEP Score	Between-State SD	Within-State SD	Multiple (Within/Between)
4th Grade Reading	220.1	6.6	34.7	5.3
4th Grade Math	239.5	6.3	27.8	4.4
8th Grade Reading	263.3	6.5	32.9	5.1
8th Grade Math	282.4	8.5	34.8	4.1

Discussion

What effect will the Common Core have on national achievement? The analysis presented here suggests very little impact. The quality of the Common Core standards is currently being hotly debated, but the quality of past curriculum standards has been unrelated to achievement. The rigor of performance standards—how high the bar is set for proficiency—has also been unrelated to achievement. Only a change in performance levels has been related to an increase in achievement, and that could just as easily be due to test score changes driving changes in policy, not the other way around. The Common Core may reduce variation in achievement between states, but as a source of achievement disparities, that is not where the action is. Within-state variation is four to five times greater.

The sources of variation in educational outcomes are not only of statistical importance but also bear on the question of how much state policy can be expected to change schools. Whatever reduction in variation between, say, Naperville and Chicago that can be ameliorated by common standards has already been accomplished by Illinois's state efforts. State standards have already had a crack at it. Other states provide even more deeply rooted historical examples. California has had state curriculum frame-

The Common Core may reduce variation in achievement between states, but as a source of achievement disparities, that is not where the action is.

Standards in education are best understood as aspirational.

works since at least 1962, statewide testing with scores for every school published publically since 1971 (except for a brief timeout in the early 1990s), state textbook adoption for K–8 since the nineteenth century, and a court-ordered equalized spending system since the late 1970s. Any effect that these laws have on reducing achievement variation within the state has already occurred. The Common Core must go beyond these efforts to reduce variation in California's achievement. That is highly unlikely.

Two lessons can be drawn from the analysis above. First, do not expect much from the Common Core. Education leaders often talk about standards as if they are a system of weights and measures—the word "benchmarks" is used promiscuously as a synonym for standards. But the term is misleading by inferring that there is a real, known standard of measurement. Standards in education are best understood as aspirational, and like a strict diet or prudent plan to save money for the future, they represent good intentions that are not often realized.

Why don't aspirational standards make much of a difference? Researchers from the International Association for the Evaluation of Educational Achievement (IEA) first sketched the concept of opportunity to learn using international test score data in the 1970s.[16] Distinctions were drawn among the intended, implemented, and achieved curriculums. The intended curriculum is embodied by standards; it is what governments want students to learn. The differences articulated by state governments in this regard are frequently trivial. Bill Gates is right that multiplication is the same in Alabama and New York, but he would have a difficult time showing how those two states—or any other two states—treat multiplication of whole numbers in significantly different ways in their standards documents.

What is crucial is the distance between the intended curriculum and the two curriculums below. The implemented curriculum is what teachers teach. Whether that differs from state to state is largely unknown; what is more telling is that it may differ dramatically from classroom to classroom in the same school.[17] Two fourth-grade teachers in classrooms next door to each other may teach multiplication in vastly different ways and with different degrees of effectiveness. State policies rarely touch such differences. The attained curriculum is what students learn. Two students in the same classroom and instructed by the same teacher may acquire completely different skills and knowledge. One student understands and moves on; another struggles and is stuck. And that even happens in classrooms with outstanding teachers.

The whole system is teeming with variation. Policies at national, state, district, and school levels sit on top of these internal differences, but they rarely succeed in ameliorating them. The Common Core will sit on top of the implemented and attained curriculums, and notwithstanding future efforts to beef up the standards' power to penetrate to the core of schooling, they will probably fail to dramatically affect what goes on in the thousands of districts and tens of thousands of schools that they seek to influence.

A final word on what to expect in the next few years as the development of assessments tied to the Common Core unfolds. The debate is sure to grow in intensity. It is about big ideas—curriculum and federalism. Heated controversies about the best approaches to teaching reading and math have sprung up repeatedly over the past century.[18] The proper role of the federal government, states, local districts, and schools in deciding key educational questions, especially in deciding what should be taught,

remains a longstanding point of dispute. In addition, as NCLB illustrates, standards with real consequences are most popular when they are first proposed. Their popularity steadily declines from there, reaching a nadir when tests are given and consequences kick in. Just as the glow of consensus surrounding NCLB faded after a few years, cracks are now appearing in the wall of support for the Common Core.

Don't let the ferocity of the oncoming debate fool you. The empirical evidence suggests that the Common Core will have little effect on American students' achievement. The nation will have to look elsewhere for ways to improve its schools.

The empirical evidence suggests that the Common Core will have little effect on American students' achievement.